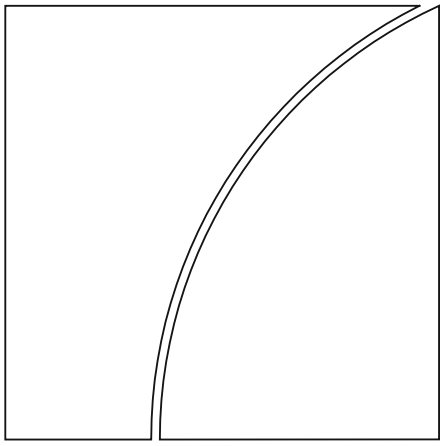
Irving Fisher Committee on Central Bank Statistics



IFC Bulletin

No 61

Granular data: new horizons and challenges

July 2024

Irving Fisher Committee on Central Bank Statistics

BIS

Contributions in this volume were prepared for the proceedings of the IFC Satellite Seminar at the ISI 64th World Statistics Congress on *"Granular data: new horizons and challenges for central banks"*, co-organised with the Bank of Canada, in Ottawa, Canada, on 15 July 2023.

The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the Bank of Canada, the IFC and its members and the other institutions represented at the meeting.

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2024. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1991-7511 (online)

ISBN 978-92-9259-740-5 (online)

Granular data: new horizons and challenges

IFC Bulletin No 61
July 2024

Proceedings of the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada, Ottawa, Canada, on 15 July 2023

Overview

Granular data: new horizons and challenges for central banks

Joshua Brault, Senior Data Scientist, Data Science, Analytics and Insights Department, Bank of Canada

Maryam Haghighi, Director, Data Science, Analytics and Insights Department, Bank of Canada

Bruno Tissot, Head of Statistics and Research Support, BIS, and Head of IFC Secretariat

Introductory remarks

Pablo García Silva, IFC Chair

1. Micro data: general benefits and new insights

The secret ingredient of being relevant? Microdata. Making the move from ready-made to fresh statistics

Luís Teles Dias and Lígia Maria Nunes, Banco de Portugal

Granular statistical data – Are we charmed with a blessing or a curse?

Jasminka Dodeva Kovaceva, National Bank of the Republic of North Macedonia

On the need for granular, forward-looking physical risk data to support policy analysis

Jens Mehrhoff, International Monetary Fund

Payment habits during Covid-19: Evidence from high-frequency transaction data

Tatjana Dahlhaus and Angelika Welte, Bank of Canada

Classifying Job Postings into NAICS Codes

Tatjana Dahlhaus, Reinhard Ellwanger, Gabriela Galassi and Pierre-Yves Yanni, Bank of Canada

2. Prices and monetary policy

Narrative monetary policy uncertainty

Charles Martineau, Zissis Poulos, Yuntao Wu, Cameron Thompson, Jun Yuan and John Hull, Joseph L. Rotman School of Management, University of Toronto, and Maryam Haghighi, Bank of Canada

Determinants of the price of bitcoin: An analysis with machine learning and interpretability techniques

José Manuel Carbó Martinez and Sergio Gorjón Rivas, Bank of Spain

Property classification with administrative data: The case of the Chilean household price index
Juan José Balsa and Javiera Vasquez, Central Bank of Chile

Estimating effective inflation using scanner data
Younghwan Lee, Bank of Korea

Financing imbalances and heterogeneous effects of monetary policy
Jorge Antonio Pozo Sanchez and Youel Rojas Zea, Central Reserve Bank of Peru

The Eurosystem's granular data on securities, loans and entities – How we use such information today and what we may develop differently next time
Frank Mayerlen, European Central Bank

3. Macroeconomic perspectives

The use of payment transaction data for economic forecasts
Guerino Ardizzi, Giuseppe Bruno and Juri Marcucci, Bank of Italy; Roberto Iannaccone, Filippo Moauro, Alessandra Righi and Davide Zurlo; Italian National Institute of Statistics (ISTAT)

Retail credit estimates in Federal Reserve Board's G.19 "Consumer Credit" Statistical Release
Michael Chernousov, Federal Reserve Board of Governors

Estimating the effects of economic indicators on the language used in the business outlook survey
Dave Campbell, Charita Koya and Naveen Rai, Bank of Canada

Unit labour cost, wages, and labour productivity in Canada: A bottom-up analysis
Joshua Brault, Maryam Haghighi and Masoud Nasari, Bank of Canada, and Hashmat Khan, Carleton University

Sectoral uncertainty
Efrem Castelnuovo, University of Padova, Kerem Tuzcuoglu and Luis Uzeda, Bank of Canada

Territorial activity in Chile, by macrozone and economic activity
Pablo García Silva, Central Bank of Chile

4. International dimensions

Classification of participants in the foreign exchange market
Amir Khatib, Bank of Israel

Access to CSI for research purposes
Matthias Rumpf, European Central Bank

Connecting the dots of the international debate on the standardization and granularity of regulatory data

Massimo Casa, Bank of Italy

International trade in creative economy: an analysis with granular data
Andrea Carboni, Marco Di Bella and Flavia Ciccioli, Bank of Italy

5. Data governance and quality issues

Statistical Disclosure Control (SDC) for results derived from combined confidential microdata
Jannick Blaschke, Christian Hirsch and Robin Kollmann, Deutsche Bundesbank

Missing values imputation for Central Balance Sheets Data Office (CSBO) accounting data
Iker González Crespo and Pablo Jiménez Segovia, Bank of Spain

Utility and confidentiality assessment of synthetic company data
Eugenia Koblents Lapteva, Bank of Spain

Finding a needle in a haystack: a machine learning framework for anomaly detection in payment systems
Ajit Desai and Jacob Sharples, Bank of Canada, and Anneke Kosse, BIS

Data and analytics risk management framework for central banking
Maryam Haghighi and Vivianne Taillefer, Bank of Canada

Irving Fisher Committee on Central Bank Statistics

BIS

Granular data: new horizons and challenges for central banks

Joshua Brault, Maryam Haghighi, Bruno Tissot[1]

Executive summary

Central banks, as both producers and users of statistical information, have been increasingly interested in **making the most of the wealth of data available in modern societies**. But they are also fully aware of the need to address the important challenges associated with dealing with granular data sets. These data sets typically comprise micro-level records (such as at the individual account, instrument or transaction level) as well as disaggregated data (that are below the level of aggregated statistics and have a higher likelihood of identifying reporting units). The challenges they pose include not only the large size of these data sets and their often limited quality, but also their complexity due to the high level of details they entail and the sheer variety of their formats.

Despite these challenges, **granular data are appealing for a variety of reasons**. The most obvious is that they can offer a level of precision and comprehension into economic phenomena that aggregation may typically mask or hide. Further, granular data provide sufficient flexibility and adaptability to tackle a wide scope of issues. The insights gained can enable central banks to have a better understanding of emerging behaviours and lead to more targeted and effective policy prescriptions. Additionally, since many micro data sources are a by-product of the increasingly digitalised world, they provide information that can be timelier and more resilient to disruption, such as during financial crises. Moreover, granular data can support the production of novel economic indicators (for example on economic confidence or uncertainty) as well as experimenting with alternative ones. Finally, they have proved to usefully complement traditional macroeconomic statistics, for instance by helping to assess real-time accuracy and address gaps or missing data.

While granular data offer a level of detail and insight not previously possible with aggregate sources, **central banks must confront two important issues**. First, these data inherently lack the thorough production processes and quality assessments of typical macroeconomic statistics. Second, to provide value they must be situated within a coherent framework and contextualised so that they can ultimately inform policy decisions.

[1] Senior Data Scientist, Data Science Innovations Division, Bank of Canada (JBrault@bank-banque-canada.ca); Director of Enterprise Data Science and Insights, Bank of Canada (MHaghighi@bank-banque-canada.ca); and Head of Statistics and Research Support, BIS & Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC) (Bruno.Tissot@bis.org).

The views expressed here are those of the authors and do not necessarily reflect those of the BIS, the Bank of Canada, the IFC or any of the institutions represented at the event.

We thank Douglas Araujo, Gloria Peña and Olivier Sirello for supporting comments and helpful suggestions.

Fortunately, central banks' experience shows that **data science and its large set of new tools and techniques can be very effective at tackling these issues**. Four aspects deserve careful consideration:

- **Granular data in their raw form are often unstructured, unverified and dispersed**.[2] Consequently, lengthy pipelines are typically required to receive, transform and validate the information collected before it can be useful. Due to the vastness of the data, these operations can hardly be labour-driven and require a high degree of automation. Fortunately, recent advances in artificial intelligence (AI) and machine learning (ML) offer great opportunities to harness the insights from granular data, by performing a wide variety of tasks such as anomaly detection, real-time monitoring, pattern recognition and predictive analytics.
- The defining feature of granular data is the possibility of identifying individual reporting units. While **there are serious considerations related to privacy**, central banks in isolation might not be able to process the vast quantity of granular data and benefit fully from the insights they offer. Finding the means to collaborate with other authorities, the industry and/or academia without jeopardising privacy or confidentiality requires new and innovative ways to work with and share granular data.
- A third important feature relates to the governance and standardisation of granular data. Unlike aggregate statistics, for which common frameworks are usually in place to ensure consistency and comparability, **harmonising granular data across countries is more challenging** for a variety of reasons. These include varying levels of technological infrastructure, idiosyncratic regulatory and legal frameworks, as well as different degrees of data availability and access. Given the interconnected nature of today's globalised world, developing universally adopted frameworks and norms related to granular data appears essential.
- A majority of key policy models and analyses at central banks have been built over the years based on macroeconomic statistics. Important challenges need to be addressed in order to update or provide **new models that leverage granular data**, and to reconcile granular data with their aggregate counterparts.

Looking forward, making the most of the opportunities provided by granular data calls for **developing appropriate mitigation measures** to safeguard their security, address their quality problems and ensure the usefulness of the information provided in a transparent way, especially for supporting policy decisions.

[2] Fortunately, the quality of granular data sets used by central banks (eg those derived from supervisory reporting systems, accounting data) is typically much better compared with the "average". Moreover, important projects have been undertaken since the Great Financial Crisis to make further progress, for instance in Europe with the development of the Banks' Integrated Reporting Dictionary (BIRD) and of the Analytical Credit Database (AnaCredit); see Israël and Tissot (2021). Yet dealing with micro data sets nevertheless requires important validation work, a telling example being the data collected by trade repositories (IFC (2018)). Moreover, one important challenge is to integrate/reconcile information coming from different reporting exercises, especially for statistical, prudential and resolution purposes. In Europe, and as argued by Casa, authorities have already launched strategic initiatives to rationalise, standardise and integrate existing reporting frameworks – cf the ongoing Eurosystem Integrated Reporting Framework (IReF) initiative (ECB (2022)).

1. Introduction

The defining feature of granular data is their level of detail, broadly encompassing both disaggregated data and micro data. On the one side, disaggregated data are below the level of aggregates and are more likely to identify economic agents. On the other side, micro data refer to individual reporting units (persons or entities) which can be identified (IAG (2017)).[3] Examples of granular data include social media interactions, firm-level financial statements, credit and bank transaction records, news articles, sector-specific output and productivity metrics, online job postings, and geospatial climate data.

Historically, the quantity and scope of granular data have been limited, primarily due to the labour-intensive and costly nature of their collection, processing and storage. Another reason was the primary focus on macroeconomic fluctuations and the related subdued interest devoted to micro-level developments. **Central banks have thus traditionally relied on aggregate data sources** to analyse economic conditions and make data-informed policy decisions. These sources, which are typically accurate, reliable and methodologically consistent, include statistics on national accounts, the labour market and prices. Yet aggregate data have several drawbacks, not least limited frequency, timeliness and level of details. Further, as highlighted during the Covid-19 pandemic, major disruptions to statistical production processes may hamper the supply of traditional indicators, leaving central banks in the uncomfortable spot of having to conduct policy without their usual information inputs. Expanding the data offering, particularly with alternative and more granular sources, can be essential to mitigate such risks and foster statistical resiliency (Jahangir-Abdoelrahman and Tissot (2023)).

The situation is now rapidly changing. Macroeconomic analyses have increasingly recognised the merit of incorporating micro-level insights, especially since the Great Financial Crisis of 2007–09 (GFC).[4] Moreover, **an unprecedented quantity of real-time and near real-time granular data are now available** with the ongoing data revolution, reflecting both the increasingly digitalised world and the proliferation of the Internet of Things (IoT) associated with the development of connected devices. This wealth of data presents central banks with a plethora of opportunities to foster innovation with respect to their work, as argued by Pablo García Silva (Irving Fisher Committee on Central Bank Statistics (IFC)). However, using granular data is also accompanied by new risks that central banks must now confront.

In this context, the IFC and the Bank of Canada co-organised "Granular data: new horizons and challenges for central banks" as a satellite seminar to the 64th World Statistics Congress (WSC) of the International Statistical Institute (ISI) in 2023.[5] This event was an opportunity to take a comprehensive view of the uses of granular data by central banks, highlighting their analytical benefit, the required tools and

[3] It is worth noting that while granular data and big data are related notions, they are conceptually different. Granular data are defined by the level of detail, whereas big data are characterised by particular features such as their volume, velocity and variety (IFC (2017)). Hence, granular data can form part of big data, but not all big data are granular.

[4] See for example Gabaix (2011) for the interest of using firm-level data to analyse aggregate shocks.

[5] Proceedings of ISI WSCs are available at Proceedings & Abstracts | ISI (isi-web.org).

approaches needed to unlock their full value, and the challenges associated with using granular data. The various contributions, referred to in this overview and included in this *IFC Bulletin*, discussed the general benefits provided by granular data (section 2), the opportunities offered by technical innovation (section 3) as well as the key risks faced when using granular data and the associated mitigation measures in central banks (section 4).

2. General benefits – novel insights from granular data

Granular data offer several benefits and novel insights to central banks, whose experience underscores four key areas of interest:

i. Granular data provide deeper and more flexible insights across a broad spectrum of topics.

ii. The wealth of granular data originating from the increasingly digitalised world facilitates faster collection and utilisation of statistics, with a timelier and more resilient production process.

iii. The richness of granular data along with improvements in technological capacity can yield entirely new types of economic indicators, while also allowing statisticians to refine and improve methodologies for the existing ones.

iv. Granular data can also act as a valuable complement to available statistical aggregates, helping to address issues such as real-time measurement error and the imputation of missing data.

Deeper and more flexible insights

The wealth of granular data, both in scope and detail, can support a more complete understanding of economic phenomena, with deeper and more flexible insights into emerging patterns, trends and relationships.

The first way which granular data sets can aid central banks is **the precision and nuance of the information** they offer. With traditional data, aggregation leads to a loss of information and limits the scope of questions that can be adequately answered. A major reason for this pertains to heterogeneity, especially across different sectors, geographic regions or demographic groups. For instance, a recent Bank of Canada study shows that the degree of economic uncertainty may vary significantly across sectors and can be driven by idiosyncratic developments in specific areas. Moreover, combining different individual positions into "normal" macro aggregates raises the risks of overlooking tail risks, as evidenced during the GFC. In contrast, granular data enable a much more comprehensive view of economic developments and the impacts of policy decisions, not only on macroeconomic aggregates but also on the distributions of those aggregates (Israël and Tissot (2021)). This sentiment was echoed by the former President of the European Central Bank a few years ago (Draghi (2016)):

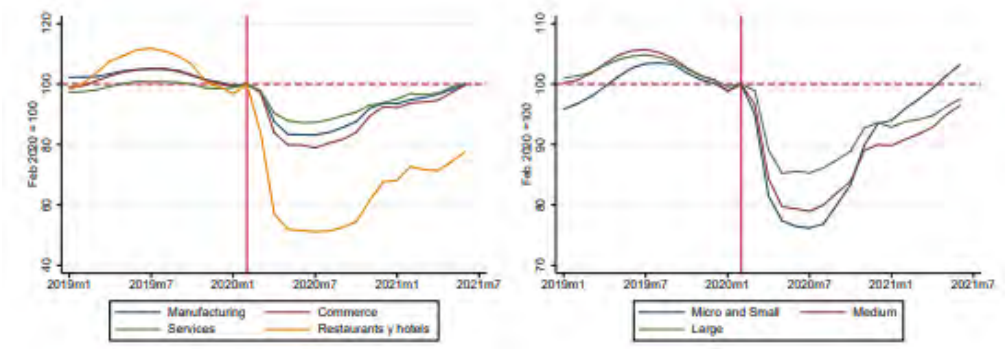
"*Disaggregated data are indeed necessary to identify and analyse the heterogeneity that characterises the real world. For central banks this is particularly*

important: to implement policy in the most effective way, we need to know how our policy actions affect all sectors of the economy."

This type of nuance was also **especially helpful during the Covid-19 pandemic** in 2020–22 when the situation was shifting rapidly, making it difficult to assess the current state and future trajectory of the economy.[6] These challenges were further complicated by the fact that both the spread of the virus and associated containment measures had disproportionate impacts on different regions, sectors and demographic groups.

One telling example of how granular data could help to inform policy in this context was the **use of firm-level administrative registers** in Chile (Albagli et al (2023)). This rich data set provides precise insights on firm-level production, firm-to-firm transactions, firm-to-bank credit transactions, employer-employee interactions, and access to credit and employment support programs. This information proved key to understanding how firms' margins of adjustments (in terms of changes in output, capital, labour, intermediate inputs or productivity) responded to the pandemic as well as to better gauging firm-level heterogeneity. For instance, while employment fell in all industries during the pandemic, the decline was particularly strong for firms in food and hospitality; additionally, employment losses occurred primarily in micro, small and medium-sized enterprises (Graph 1).

Employment in Chile during the Covid-19 pandemic: evolution across sectors and types of firm — Graph 1



(a) Sector (b) Size

Note: Seasonally adjusted employment, excluding workers enrolled in the employment protection program, normalised at 100 in February 2020.

Source: Albagli et al (2023).

Such detailed insights can in turn allow policymakers to **undertake and calibrate more targeted policy responses**. Indeed, the Central Bank of Chile, for example, took a number of exceptional measures during the pandemic to specifically address the

[6] For instance, Chetty et al (2024) were able to build a publicly available database that tracks US economic activity at a granular level in real time using anonymised data from private companies.

challenges faced by credit borrowers.[7] Of note was the opening of the Conditional Financing Facility for Increased Loans (FCIC), a credit line for commercial banks with the aim of increasing lending to consumers and smaller businesses. Turning to the case of Peru, the use of entity-level data has proved helpful to analyse the monetary policy transmission channel, depending on the level of leverage of financial institutions.

Another example is the widespread use of **granular data sets covering credit and bank transactions**. These can provide high-frequency glimpses into household behaviour and spending patterns, allowing policymakers to gauge economic developments and formulate appropriate policy responses. For example, one study conducted in Italy used transaction data from the retail settlement system to produce real-time estimates and forecasts of final consumption and its components. It shows that the addition of transaction data significantly improved both nowcasting and forecasting performance, confirming other research on gross domestic product (GDP) forecasts (Aprigliano et al (2019)) and on the use of debit card transactions in Norway (Aastveit et al (2020)). Similarly, Buda et al (2022) show that transaction-level data from Spanish retail accounts can be combined according to national accounting principles to produce highly accurate real-time measures of aggregate consumption.

The **second way in which granular data can provide greater insights is the adaptability that they offer**. Such flexibility is particularly important for conducting evidence-based policies. For instance, since the impact of monetary policy decisions can have long and variable lags, it is paramount that central banks are cognisant of not just the current state of the economy, but how the economy will evolve over the subsequent months and years.

This type of adaptability was particularly useful to policymakers in Canada during the Covid-19 pandemic, when questions arose about the actual level of inflation facing households. The consumer price index (CPI) is computed using a fixed basket of goods and services, with expenditure weights typically updated every two years; however, consumption patterns shifted dramatically during the pandemic. This raised concerns that reported figures could be quite different than the inflation households were experiencing. To address this issue, the Bank of Canada and Statistics Canada partnered to construct an adjusted CPI measure based on a real-time basket of goods and services, drawing on a variety of information sources, including consumer credit card purchases, survey data and transaction data from grocery retailers. Interestingly, this alternative measure showed that, after adjusting for changes in consumption patterns, the inflation faced by consumers was only slightly higher than what was reported in the official numbers (Huynh et al (2020)).

Yet, making the most of the flexibility provided by granular data calls for **having a coherent statistical framework to ensure the correct integration of micro-level observations into macroeconomic aggregates**. It is essential to support top-down approaches that allow "zooming in" on observed macro developments. Conversely, it helps to "zoom out" and compute idiosyncratic aggregates addressing specific policy needs, concerns or research questions (IFC (2021a)). One example of this two-way micro-macro interaction is the recent project developed in Korea to use granular

[7] A full description of measures taken can be found at www.bcentral.cl/en/web/banco-central/exceptional-measures.

scanner data – that is, information collected by barcode scanners at supermarkets – to reconstruct a CPI index that better captures changes in aggregate consumer behaviour. The new index appears to provide leading insights compared with official figures and is also available at a higher (weekly) frequency. Similarly, a recent bottom-up analysis conducted at the Bank of Canada generated more detailed insights into the behaviour of Canada's unit labour cost, through the development of a coherent framework reconciling granular data with aggregate information. A last example is the growing interest central banks have in developing granular security-by-security databases that can be used to derive meaningful macro aggregates (Dilip and Tissot (2024)).

More timely and resilient sources of information

In addition to offering a level of detail and insight otherwise unattainable, **granular data allow for faster production and utilisation of statistical information**. A key reason is that many micro-level data sources now originate as a by-product of technology innovation, digitalisation and the IoT. Examples of these "financial big data" include digital footprint indicators such as e-commerce sales, Google search trends, social media interactions and mobile phone usage; signals from electronic connected "smart" devices such as mobile phones, thermostats, physical sensors and grids; financial operations such as credit card transactions, bank transfers and payments; and supply chain logistics data such as inventory levels, shipping activities and transportation networks (IFC (2021b)). Certainly, and reflecting the large differences in how these various types of data are produced and collected, the information obtained has several shortcomings compared with traditional aggregate statistics compiled under well established methodologies. Yet, this information can be both more timely and more resilient in specific circumstances, such as during the Covid-19 pandemic (Jahangir-Abdoelrahman and Tissot (2023)).

An **appealing aspect of granular data relates to timeliness**. In contrast to aggregate statistics – such as national accounts data which are computed at regular intervals (eg quarterly or annually) – micro data can be collected and compiled at a much higher frequency, in some cases on a practically continuous basis. Further, the period between collecting and processing can be nearly instantaneous, meaning that one can observe changing conditions or emerging trends in almost real time and respond to situations with greater speed and accuracy. As illustrated in the case of Banco de Portugal, this represents a key opportunity to move from traditional, "ready-made" statistical reports to dynamic, micro data-driven insights.

Another **key feature of granular data relates to resiliency**. Since in many cases the information originates as a by-product of activities, it can be more resilient to disruptions than traditional economic statistics which specifically require adequate compilation exercises (eg censuses, surveys). For example, the US federal government shutdown observed in December 2018–January 2019 because of the inability to reach a political agreement on funding coincided with a period of heightened uncertainty and financial market turbulence. It also led to delayed publication of many official statistics. These statistics were not available until around mid-February 2019, leaving little information to assess the state of the economy. This was the case, in particular, for retail sales reports, one of the most used real-time indicators of consumer spending. In contrast, a retail spending index could be constructed from transaction

data (Aladangady et al (2021)). This alternative measure was unaffected by the shutdown and provided policymakers with a lens into US consumer spending during this period, while official statistics did not. Another example of granular data resiliency was during Covid-19, as the pandemic disrupted the production of official statistics across many countries. Fortunately, statisticians could address these information gaps and related methodological and data quality issues by using a variety of "alternative" micro-level sources, including transaction data, web and social media data, and other digital footprint indicators, to name a few (de Beer and Tissot (2021)).

However, while granular data can be more resilient to disruptions in the flow of information in certain cases, one needs to be cautious about potential changes in their underlying production process and the implications for analysing them. As an illustration, large shifts were observed in payment choices at the onset Covid-19, as consumers opted to use cards much more frequently than cash. This led to a bias in household spending estimates based on card transaction data, with the need to carefully account for the shift in the card-to-cash ratio.

New economic indicators

There has been a **substantial increase in the types of economic indicators available to policymakers**, reflecting the wealth of new data generated by digitalisation as well as the greater use of "traditional" information sets (eg administrative registers, textual content) thanks to improvements in technology and methodological innovation.

Most notably, **central banks now place a large emphasis on compiling and analysing *soft indicators***, for instance agents' perceptions and expectations about the economy (Shapiro and Wilson (2017)). While variants of such sentiment, confidence and uncertainty indicators had been available prior to the recent proliferation of granular data, they were typically survey-based. This can be resource-intensive and may be of limited help in times of abrupt change. In contrast, new approaches – such as the rapidly developing field of natural language processing (NLP) tools – can facilitate the compilation of soft indicators directly from news and media. For example, the Federal Reserve Bank of San Francisco now publishes a high-frequency measure of economic sentiment, constructed by comparing a lexicon (ie a list of pre-defined words) with economics-related text collected from leading newspapers. This approach has proved quite effective in providing timely warning signals. Notably, the index exhibited a sharp decline in January 2020 at the outbreak of the Covid-19 pandemic, nearly two months earlier than survey-based sentiment measures available at that time (Buckman et al (2020)). However, lexicon-based approaches can be limited in their ability to capture context and nuance, and attention is increasingly put on using more sophisticated techniques based on ML to quantify textual information and get more subtle indications and context-specific signals (see Box A).

A second important development has been related to the **new types of indicators originating from so-called *hard granular data***. One example is the Global Supply Chain Pressure Index published by the Federal Reserve Bank of New York, which is derived from a variety of micro data sources, including information on global transportation costs, air freight cost indices and surveys related to supply chains (Benigno et al (2022)). New economic indicators can also be derived from

granular, high-resolution satellite images; for example, the Global Agriculture Monitoring project uses satellite imagery to measure crop conditions, agricultural land use and estimates of future crop yields.

Box A

Harnessing breakthroughs in language processing for compiling new economic indicators

In recent years **the field of natural language processing (NLP) has seen significant advancements**, most notably with the development of large language models (LLMs), for which a key architecture is the Bidirectional Encoder Representations from Transformers (BERT) (see Araujo et al (2023)). Using deep learning, these state-of-the-art models pre-train on a large amount of textual data with the objective of predicting missing words in sentences based on the surrounding context (for instance the words to both the left and right of the missing ones). To successfully accomplish this task, these models must learn the basic structure, rules and organisation of the language used. A major contributor to their widespread usage is that, once pre-trained, they can be fine-tuned for other purposes with a relatively small amount of task-specific data, a process known as *transfer learning*.

BERT models represent a **promising avenue for constructing soft information measures that can be useful for central banks**. One example is the recent initiative to build a quantitative measure of narrative-based monetary policy uncertainty (NMPU) by drawing on US newspaper articles published over several decades. The approach involves, first, assigning each individual article a score based on the fraction of its content expressing uncertainty and, second, compiling an aggregate measure by averaging scores at desired intervals (eg daily, weekly, monthly). This NMPU index has tended to rise in the days before announcements made at Federal Open Market Committee (FOMC) meetings and to fall afterwards – although it has remained elevated in the days following announcements that were particularly "hawkish". Moreover, the index responded to large changes in macroeconomic fundamentals such as inflation, unemployment and housing prices.

Compared with existing narrative-based uncertainty measures constructed from lexical approaches (cf Baker et al (2016)), which effectively check for the presence of the word "uncertainty" or synonyms or derivatives of it, **one advantage of NMPU-type indicators is that they incorporate the context in which "uncertainty" is expressed**, such as through questions, tentative statements or expressions of doubt. This aspect is particularly salient, as a challenge with static dictionaries is the fact that words expressing uncertainty often vary across time and publication outlets. Additionally, NMPU measures can be computed from relatively few articles and reported at a high frequency (eg daily), in contrast to lexicon-based measures that can require many articles to be aggregated meaningfully and are typically restricted to monthly frequencies.

Lastly, **an interesting aspect of the new approaches is their flexibility**. In addition to the measurement of monetary policy uncertainty, similar soft information measures could be computed in a relatively straightforward way to cover other various areas of interest to central banks, such as energy, housing and inflation.

Third, a wealth of untapped granular data sets can still be used to expand, sometimes on a very experimental basis, **the scope and variety of economic indicators that can address new information demands** depending on circumstances. As evidenced by the Covid-19 pandemic, it is not always *a priori* obvious what data will be useful in rare events such as crises. By allowing a wide range of granular data sets to be integrated into existing analytical frameworks, central banks can have more flexibility to address potential information needs that are hard to predict at the current juncture. For instance, one identified data gap in supporting climate policy analysis relates to geospatial forward-looking physical risk indicators

that are increasingly needed to assess the development of environmental hazards and their potential financial stability implications (Aurouet et al (2023)).

Complementing the offering of macroeconomic statistics

In addition to offering a novel and timely source of information, granular data can be used to **complement official macroeconomic statistics**. For one, initial releases of aggregate economic data can contain some form of measurement error. In the case of variables like GDP, for instance, preliminary releases often reflect incomplete information because of reporting lags for the various data inputs. As these inputs become available, values are corrected to reflect this updated information. But measurement error can also occur for other reasons, such as sampling errors, seasonal variation and changes in methodology, or because the construction of a variable relies on assumptions or estimation methods which are imperfect themselves.[8]

Dealing with these issues can have **important policy implications**. Measurement errors are not simply statistical anomalies which dissipate with time; they can potentially influence more lasting decisions taken based on the initial information released. A well-known example comes from transcripts of FOMC meetings in late 1992, which showed that policymakers were quite concerned about economic weakness. These worries were based on incoming data (eg on industrial production) suggesting little or no GDP growth. But these initial statistical releases contained a significant amount of measurement error, and revised data subsequently showed that the economy in that period was actually quite strong (Croushore and Stark (2001)).

A second way that granular data can be used to complement macroeconomic indicators is **by filling in information gaps**. Such gaps in aggregate official statistics can arise for a variety of reasons, such as when methodological refinements incorporate data that are not available for the entire time series history or because of disruptions in production processes. These gaps may be temporary, as was the case with the US federal government shutdown described above, or more permanent. In this second case, missing data can not only impact contemporary policy decisions but also limit the understanding of historical episodes. This clearly puts a premium on trying to fill information gaps by using the micro data sources available. One example is the recent discontinuation of the collection of accounts receivable data in the US Census Bureau's Annual Retail Trade Survey, which led to gaps in macroeconomic consumer credit indicators published by the Federal Reserve Board. To address this problem, the solution was to use another granular data set available – ie the Consumer Credit Panel data set managed by the Federal Reserve Bank of New York Center for Microeconomic Data. A second example of the use of micro data sources to fill permanent gaps in aggregated indicators relates to the various measures taken to address disruptions in official statistics caused by the Covid-19 pandemic (see Box B).

[8] The importance of these shortcomings can explain why central banks are increasingly dedicating a significant number of resources to nowcasting exercises as a complement to official statistical data releases.

3. Unlocking the value of granular data using data science

Granular data can offer fast and detailed information about economic developments, but they inherently lack the thorough production processes and quality assessments of typical aggregated statistics. Further, for granular data to provide value to central banks they must be situated within a coherent framework and contextualised so that they can ultimately inform policy decisions.

Data science can be essential to addressing these two issues.[9] First, it offers a large set of new tools and techniques (including traditional statistical analysis, various data visualisation and optimisation methods as well as AI/ML techniques) that can operationalise granular data in a systematic way, despite the challenges posed by their quantity, complexity and variety of formats. One example is the recent project using complex language models to understand the impact of economic conditions on the qualitative sentiment of industry leaders, as collected in the Bank of Canada's business outlook surveys.

Box B

Using granular data to address gaps in aggregate statistics compiled during the Covid-19 pandemic

In the course of 2020 many central banks and statistical agencies encountered disruptions to the production of official statistics, as the spread of the virus and associated containment measures made it difficult to collect data through usual channels, such as surveys and interviews. Additionally, pandemic-induced shifts in consumption behaviour led to heightened uncertainty about the reliability of the information that agencies could compile. This left authorities in the uncomfortable position of having to conduct policy absent many of their usual information inputs.

For instance, the Job Vacancy and Wage Survey (JVWS) in Canada was temporarily suspended during the second and third quarters of 2020. This survey, conducted by Statistics Canada, provides detailed information about employment conditions (eg job vacancies, wage rates and job characteristics across different industries and regions). Due to the suspension, important data gaps arose at a time when the labour market was undergoing significant changes. It was thus **essential to find complementary sources to mitigate these gaps**.

Regarding for instance information on job vacancies, one can utilise granular data from Indeed.com, a popular online job market platform in Canada that allows individuals to search for employment and employers to post job listings. Data from this platform are available in near real time at a high frequency. But one difficulty in using this source to estimate missing JVWS data is that Indeed uses its own sectoral classification standard which differs from the North American Industry Classification System (NAICS) used in the JVWS.

To overcome this issue, the approach followed by Bank of Canada researchers started with matching Indeed data with another granular data set from Advan, a company that produces geospatial data of businesses, and which also happens to contain corresponding NAICS codes. This matching exercise was achieved using an algorithm for identifying similarity between the names of the companies when they belong in the same area. The resulting data set

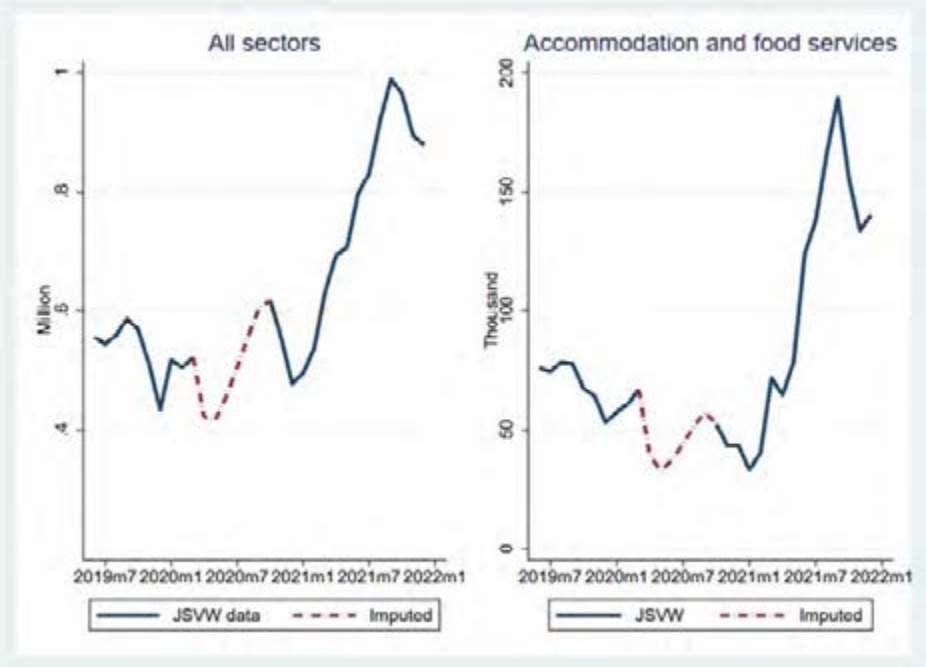
[9] The IFC has been organising recurrent data science workshops with the Bank of Italy to review developments in the big data ecosystem and the ongoing adoption of data analytics. These have provided the opportunity to review the contribution of ML applications to a wide range of use cases in central banks (IFC (2022)) and the development of related data science applications and tools, including NLP techniques and LLMs (IFC (2023a)).

yields a highly detailed measure of job vacancies at the city, provincial and national levels, and with a sectoral classification consistent with NAICS.

As a validation exercise, the job vacancies estimated from the new data set were compared with reported figures in the JVWS for periods for which data from both sources are available. The two measures appeared highly correlated, allowing for estimating a stable relationship between them and imputing missing values for job vacancies in the JVWS during the period when the survey was suspended (Graph B1).

Addressing permanent statistical gaps: imputation of missing data on job vacancies in the Canadian Job Vacancy and Wage Survey during Covid-19

Graph B1



Source: T Dahlhaus, R Ellwanger, G Galassi and P-Y Yanni (2024): "Classifying job postings into NAICS Codes", *IFC Bulletin*, no 61.

Second, data science provides new and alternative ways to derive insights that are needed to conduct evidence-based policies. From this perspective, the most recent advances in AI techniques (eg generative AI), combined with ever-increasing computing power, are widening the range of possibilities to support dealing with granular data (IFC (2020)). Since these methods are being increasingly adopted across industries and offer the potential for significant shifts in many sectors, it is important for central banks to remain agile and aware of ongoing innovations.

Data collection, management and operation

Data science provides tools and techniques which can operationalise dealing with granular data and in particular help to address the challenges they pose. These challenges include not only the large size of micro data sets, usually available in great

quantities with unchecked quality, but also their complexity due to disparate sources and a high level of details. A further problem is the variety of their formats, from large, structured databases formatted with standardised criteria to unstructured big data sets (Schubert (2021)). Examples of the former include detailed administrative databases, such as credit registers or security-by-security databases, while the latter include web and social media data, sensor data, and logs generated from software, applications and devices. And an additional issue is the difficult inferring from micro-level observations developments that are pertinent at the aggregate level.[10]

These challenges illustrate that, before granular data can be utilised, they **typically require a number of pre-processing steps**. These steps include making transformations of the raw data, correcting typos, dropping redundant information, removing duplicates and so forth (ie "cleaning" the data). Additionally, in many cases a granular data set is only one piece of a larger integrated information set, and the data need to be merged with other sources of information or aggregated to a desired level.[11] Historically, these pre-processing steps could be managed manually, but with the diversity, quantity and frequency of granular data, it has become essential to be able to automate them. Needless to say, this effort is best suited for data scientists, because of their expertise in working with large information sets across a variety of formats.

The next phase is to have an efficient pipeline to **extract from the raw data actionable knowledge that can be used to inform policy**. The goal is to establish a systematic approach, in a well documented way, to scale with the ever-growing quantity of granular data and also to be able to constantly evaluate and make improvements to production pipelines. In the event of shortcomings or problems related to the data, having such a framework allows for conducting ex post analysis of errors and taking corrective measures. Here also, data science expertise can help to develop processes in a modular way and efficiently manage a variety of granular data sources.

Finally, to provide policy insights in real time there **must be essentially zero delay** between obtaining the raw granular data and the point where they are transformed and ready for analysis. Data science expertise in code and hardware optimisation, parallel processing and other techniques promoting efficiency and speed can be critical here.[12]

Quality assessment

In addition to making granular data and their insights available to policymakers in a timely fashion, data science tools and techniques can be instrumental in **assessing**

[10] See Chodorow-Reich (2020) for an example of the difficulties faced when inferring from regional developments the aggregate impact of a macroeconomic shock, due in particular to "micro" spillovers.

[11] The need for such multi-purpose data integration is illustrated by the Eurosystem's handling of granular data on securities, loans and entities. Similarly, a recent Bank of Italy study underscores the merits and "the urgency" of integrating granular trade data and business statistics.

[12] See Cui et al (2023) for an example of the benefits of setting up a big data lake and using cloud computing tools for processing large data sets at the Bank of Canada.

the quality of the data and addressing any shortcomings. The fact that granular data are often an "organic" by-product of specific activities implies that important issues can arise with data completeness, consistency and integrity. Further, as public-facing institutions, central banks must rigorously evaluate the data and be aware of their drawbacks or limitations before integrating them as a part of their decision-making processes.

At first, data science tools and techniques have been especially useful in addressing **issues around data completeness** at central banks. For example, the use of ML at the Bank of Israel has supported the efficient classification of daily foreign exchange market transactions. The data collected can omit important details such as whether non-residents are financial or non-financial entities, or whether business sector transactions are from importers or exporters. Supervised classification algorithms were used to impute this missing information, with a reported high degree of accuracy. A similar project conducted at the Central Bank of Chile helped to address gaps in the data collected for compiling a house price index. The underlying source was the administrative data set of the tax office, which does not always distinguish between houses and apartments. After training on historical data, the tool was able to correctly classify residences with very high accuracy. A further example of the use of ML techniques relates to the Central Balance Sheet Data Office accounting data compiled by the Bank of Spain (see Box C).

In addition to data completeness, data science tools and techniques can also be used to **address problems of consistency and integrity**. Examples of such issues are selection, response or survivorship bias, measurement error, data anomalies, etc. For example, a recent project used an isolation forest ML algorithm to detect anomalies in transaction-level settlement data from Lynx, Canada's high-value payment system. These anomalies can hint at cyber events, market stress, systemic stress, fraud or operational issues, all of which are of key interest to central banks in pursuing their mandates. Consequently, detecting them as early as possible is critical. Yet one problem is that there are a large number of transactions, most of which are normal, implying that finding anomalies is tantamount to finding needles in a haystack. A layered approach was adopted to separate typical payments and analyse them for anomaly detection, with a detection rate well above existing, commonly used models.

Box C

Using machine learning to address missing data: dealing with firms' balance sheet information at the Bank of Spain

Each year the Bank of Spain receives nearly one million accounting questionnaires filled by non-financial firms. These reports contain information on firms' balance sheets, income statements, employment and wages, and other relevant financial statistics. The collected data are published in the annual Central Balance Sheet Data Office (CBSO) results. During the production process staff need to manually debug reported information, and ultimately some questionnaires have to be discarded due to missing information.

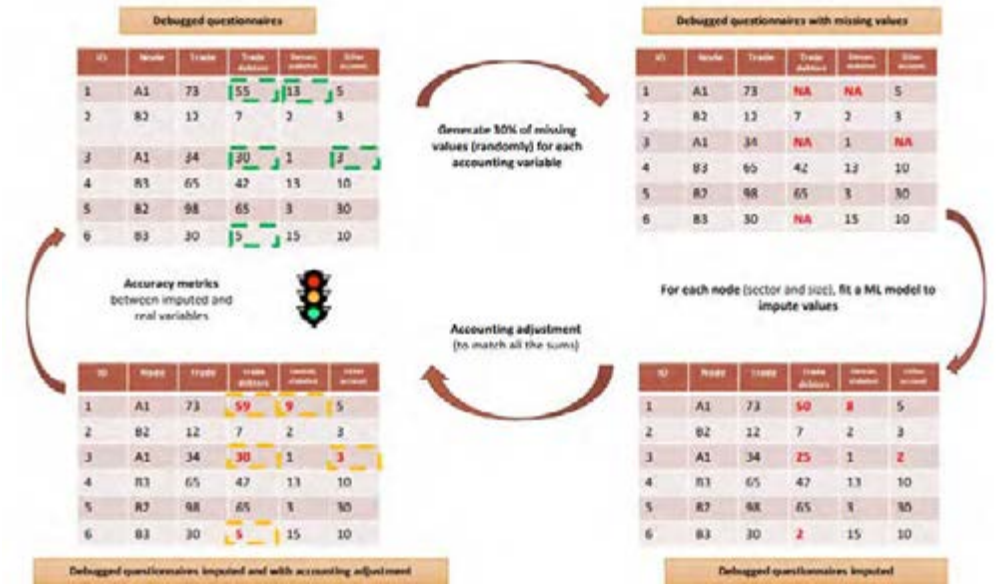
In this context, a specific project was launched **to investigate whether ML can be used to impute missing information**, allowing the Bank to enrich the CBSO database and publish more accurate statistics. The emphasis was

on sub-categories of balance sheet variables – total assets, liabilities and equity – for small and medium-sized enterprises, which are reported in questionnaires but not always with finer disaggregation, such as how total assets are distributed across inventories, cash, financial investments, etc. To impute missing information, the Multiple Imputation by Chained Equations (MICE) algorithm, a technique based on groups of random forest algorithms, was trained on received questionnaires and used to predict data from 2020. Yet one important challenge when dealing with accounting data is to ensure that subcategories can be summed to match coarser aggregates. To address this issue and ensure that the accounting decomposition holds, a scaling factor was computed based on the ratio of non-imputed to imputed values (Graph C1).

As a validation exercise, aggregates were computed from the questionnaires after imputation and compared with the initial reported data (this was done in a variety of settings, such as across different variables, firm sizes and sectors). The comparison suggested that the imputations made were highly accurate, underscoring the merit of using ML techniques to address missing information in granular data sets.

ML-based process for imputing missing values in the Bank of Spain database covering firms' financial statements

Graph C1



Source: I Crespo (2024): "Missing values imputation for Central Balance Sheet Data Office (CBSO) accounting data", *IFC Bulletin*, no 61.

A second example is the use of data science tools and techniques to **identify data tampering**. It is now well-known that a significant portion of social media activity can be driven by pre-programmed bots who post, like or re-post with the aim of spreading or amplifying misinformation or manipulating public opinion.[13] This type of activity can lead to inflated engagement metrics, making it difficult to accurately gauge public opinion. Consequently, the quality of economic indicators such as those

[13] Even in the absence of data tampering, a key issue with social media data is their representativeness, since their activities may not represent a random sample from the population of interest (Wibisono et al (2019)). Another issue relates to the (limited) availability of transparent information (eg metadata, process documentation).

measuring sentiment, confidence, expectations or uncertainty derived from social media could be adversely affected. Data science methods have proved useful in filtering out probable bot-generated behaviour, by examining a variety of profile features such as post timing, repetition, frequency and similarity with other posts, along with interaction patterns and content – a task essentially impossible for traditional data analysis. A notable example is the Botometer project from the Observatory on Social Media (OSoMe) at Indiana University. A classification system based on a random forest ML algorithm with more than 1,000 features from available metadata and information extracted from interaction patterns and content was used to identify probable bots.

New approaches and insights

Data science offers many **opportunities to central banks interested in developing a data-driven approach** to extract deeper insights from the data at hand. The main reason is that, relative to traditional techniques, data science is often better suited to incorporate a great wealth of information – by scaling with large amounts of data; handling a diverse set of formats such as text, image, voice etc; and extracting specific insights.

A main benefit is that these methods offer **tremendous ability to synthesise, summarise and present economic facts** – a key issue when dealing with the vast amounts of observations and variables contained in granular data sets. In addition, the possibility to identify complex patterns, trends and relationships is particularly salient for central banks as well as financial and prudential supervisors interested in getting insights at the level of only a small fraction of the units covered by the data – for instance, on those institutions, transactions or instruments deemed of systemic importance (Israël and Tissot (2021)). Moreover, many data science tools are tailored towards predictive analysis, which can be used to forecast key economic variables, such as GDP, inflation and unemployment (all important ingredients for conducting central bank policies). Furthermore, a notable point is that models within the realm of data science are quite heterogeneous, both in the inputs they use and the way they produce outputs. As a result, central banks can employ a variety of different models to inform policy, each of which may have their own respective benefits and drawbacks.[14] These models can also be used together, for instance for producing a single "combined" forecast instead of a series of alternative forecasts.

4. Looking ahead: addressing risks and setting up adequate mitigation measures

While granular data undoubtedly offer central banks key insights into economic and financial developments, **a number of risks and challenges** are also associated with their use. Central banks' experience underlines in particular four important aspects that deserve careful consideration and appropriate mitigation measures looking ahead, namely (i) data security and confidentiality issues, (ii) quality aspects, (iii) data-

[14] To facilitate the sharing of experience and in particular the use of ML models in economic and finance use cases, the free open-source library gingado is available on the BIS website (Araujo (2023)).

sharing and (iv) the usefulness of the information provided, especially for policy purposes.

Data security and confidentiality

A key area is security and the need to protect sensitive information (especially in terms of individual behaviours, preferences and potentially identities) from unauthorised access, corruption or even theft. Dealing with micro data clearly reinforces the risks faced in this area, reflecting four main reasons.

First, granular data sets are attractive targets for **cyber attacks and unauthorised access**. The more detailed the information, the more valuable it becomes to malicious actors seeking to exploit it for various purposes, including identity theft or financial fraud. Robust security measures must thus be implemented by organisations like central banks that collect, manage and disseminate data, not least to support their reputation and public trust and in turn the reliability of the information being reported. Hopefully, several approaches can aid in mitigating these risks, including access management, anonymisation and synthetic data and privacy enhancement techniques. The ultimate objective is to prevent unauthorised access, misuse and breaches throughout the entire data life cycle.

Despite these risks, central banks have been taking active steps to facilitate access to confidential statistical information, especially for research purposes, as documented by the ECB in the European context. One telling example relates to the Bank of Spain's project, in collaboration with the European Commission and a private software provider, to **facilitate access to synthetic versions of non-public data sets**. The approach aims to preserve the characteristics of the original data, including in terms of statistical distributions and relationships among variables, without compromising privacy. But the pilot conducted underscored important challenges, such as the effect of outliers' suppression and the limited ability for synthetic data to preserve certain linear and non-linear relationships. Another initiative is the Research Data and Service Centre (RDSC) of the Deutsche Bundesbank, which offers access to sensitive micro data for non-commercial research. A key lesson is that the set-up of robust control methods can help to reduce the risk of unauthorised disclosure, particularly when merging micro data sets. The project also underlined two important practical rules: (i) results intended for release must be based on at least five different observation units; and (ii) the combined share of the two largest observation units should not exceed 85% of the total value under analysis.

Second, while innovation can be part of the solution, it can also lead to **new, unexpected challenges, especially regarding re-identification risks** posed to granular data even when anonymised. The development of more advanced techniques and algorithms, especially the greater ability of AI models to incorporate a wide range of information sets (in particular, from additional and sometimes unexpected sources) as well as higher calculation capacity (eg quantum computing) may be a game changer from this perspective. They could for instance facilitate the re-identification of individuals by correlating seemingly anonymised data with other available sources. This can undermine existing privacy protection measures and expose individuals to unexpected scrutiny or harm.

A third, related point is that **security issues should not be solely confined to data-compiling institutions**, as they involve a wide range of stakeholders. In particular, the fact that highly sensitive information might be revealed underscores the importance of obtaining informed consent from the individuals and firms reporting their own data. They should be adequately made aware of the risks involved and also have the ability to control their records. Yet, this is particularly challenging when dealing with granular data, since it is often difficult to fully understand the implications of such detailed information – in particular regarding how it may be used by new, AI-based algorithms. Addressing these issues requires a multifaceted approach, involving state-of-the-art technical solutions and strong data governance frameworks to secure overall trust in data management processes (UNECE (2024a)). Central banks are also focusing on enhancing data literacy in the population of interest, for instance with public awareness campaigns and the promotion of responsible data practices and AI approaches.

Lastly, an essential element is to **ensure compliance with existing regulatory frameworks and ethical guidelines**. Yet, with the need to deal appropriately with granular data, privacy regulations such as the EU General Data Protection Regulation or the California Consumer Privacy Act have become increasingly complex. Organisations such as central banks must navigate a maze of legal requirements concerning data collection, processing, storage and sharing, all while ensuring that their practices align with privacy principles and rights. This has particularly important implications in terms of internal resources and skills.

Data quality

A common issue faced when working with granular information sources is their limited quality, with the term "quality" covering the various characteristics sought for in official statistics, including their accuracy and trustworthiness, integrity and security, and the need to be properly documented and easy to find and access (Križman and Tissot (2022)). In fact, granular data sets may contain errors, inconsistencies or biases that could lead to incorrect policy decisions or economic assessments. Ensuring their accuracy, reliability and robustness is therefore crucial for maintaining the effectiveness and credibility of central bank operations.

More fundamentally, the limited quality of granular sets can lead to **important inconsistencies with macro aggregates**. Discrepancies may arise for a variety of reasons, including double-counting issues or inaccurate reference data, for instance on counterparties.[15] Such inconsistencies may severely hamper the ability to use micro and macro data as interchangeable sources. A typical example is the use of security-by-security data to compile portfolio investment in the balance of payments and international investment position. Despite considerable efforts and progress, the available micro and macro sources can still exhibit some important discrepancies (Dilip and Tissot (2024)).

[15] These micro-level challenges reinforce the discrepancies already existing in macro statistics that reflect various data vintages, revision effects and different data as well as differences in interpretation and practical implementation of methodologies (ECB and Eurostat (2024)).

Mitigating these challenges calls for **setting up strong data management processes**. This is needed to achieve sound data acquisition and, more importantly, data integration processes, especially to ensure the effective linking of information sets across domains and sources – for instance by matching geospatial information with business registers and financial statements (UNECE (2024b)). They also entail the set up of plausibility checks, both internally – to check consistency based on the granular information collected – and externally – for instance to benchmark against other granular or aggregate data sources. Further, on a practical level, robust production pipelines are needed if central banks want to use granular data effectively in their operations. A key aspect relates to the automation and simplification of workflows and deployments across the statistical production chain, for instance in software development ("DevOps") and ML production pipelines ("MLOps"). Fortunately, modern engineering procedures – eg continuous integration and continuous deployment practices – can be effectively mobilised to better integrate granular sources into data management operations.

Lastly, the strengthening of granular data collection processes should be favourable for **reporting agents**, by lowering their reporting burden with more stable and resilient requirements, reducing the need for them to compile aggregated data, and in turn improving overall data quality. Indeed, the experience of the National Bank of the Republic of North Macedonia is that a key element for success is to avoid collecting the same information twice and to ensure good communication and collaboration with the reporting agents.

Data-sharing

One critical aspect of granular data is the limited possibilities for sharing them. A first issue is that, unlike aggregated data, **granular information is hardly harmonised across sectors and/or countries**, not least because entity-level records are often multidimensional and uncoded. As a result, there is no common and universal framework for dealing with granular data that would ensure consistency in definitions, classifications and methodologies. This is in sharp contrast with macroeconomic statistics, which typically rely on a common methodological framework to ensure comparability and consistency. A telling example is the System of National Accounts, which is almost universally adopted and provides a systematic way to analyse economic phenomena.

A related issue is the **limited standardisation of granular data** that prevents their adequate modelling and exchange. This calls for establishing international standards in order to develop common data structures, definitions, classifications and identifiers. More fundamentally, such standards would allow for interoperability across domains and organisations, hence promoting the findability, accessibility and reuse of granular data (UNECE (2024c)). Fortunately, important initiatives are under way to address this point. For instance, the Data Documentation Initiative (DDI) and XBRL (eXtensible Business Reporting Language) are international standards designed to, respectively, describe socioeconomic surveys, censuses and other microdata collection activities and support business reporting. Moreover, one key project endorsed by global authorities after the GFC has been the launch of the Legal Entity Identifier (LEI), a reference code to uniquely identify legally distinct entities that engage in financial transactions. Furthermore, the Statistical Data and Metadata

eXchange (SDMX) standard that supports the compilation of macro statistics has been adapted to facilitate their reconciliation with the underlying micro sources – in particular with the recent introduction of the new SDMX version 3.0 (Nikoloutsos and Sirello (2023)). Yet, the adoption of statistical standards for granular data remains extremely limited and heterogeneous. For example, the use of SDMX for working with micro data is still in infancy among central banks, mostly due to the lack of sufficient international guidelines (Bogdanova and Buffet (2023)). Other examples include the still limited coverage of the LEI and the need for a better mapping of existing micro data standards.

In addition, **several barriers prevent the sharing of granular data.** Those barriers can be technical, particularly in low- and middle-income countries, reflecting differences in technological infrastructure, capacity and statistical systems. Concrete examples include faulty repository systems, download issues and lack of harmonised software and file formats (UNSD (2023)). A perhaps more fundamental obstacle relates to the ethical challenges (eg privacy) posed by the dealing with individual records, as noted above, as well as cultural issues (eg "silo approaches"; IFC (2015)).

Policymaking institutions, in particular central banks, are however increasingly aware of the need to make progress on data-sharing, not least to better monitor evolving global trends and the risks and vulnerabilities they represent (IFC (2023b)). At the global level, critical undertakings have been initiated in this regard. First, the recommendations adopted by the G20 in the context of the third phase of the Data Gaps Initiative (DGI-3) call for the development of an international micro data-sharing standard as well as for the promotion of a sound and better access to information, including granular data, through common taxonomies (IMF et al (2023)). Second, a common task force, gathering central banks, statistical offices and international organisations, has been established to deliver guidelines and enhance the use of SDMX for micro data. Notwithstanding this progress, the adoption of a uniform approach to standardise or harmonise granular data-based statistical concepts remains a challenge, requiring strong coordinated efforts and international collaboration.

Explainability and transparency for use in policy

Properly communicating statistics to policymakers is a constant challenge (IFC (2024)). The difficulties are multiplied when dealing with granular data sets, which are typically complex, multidimensional and very large, limiting people's ability to derive clear and actionable insights out of a mass of data points.

These **issues can be reinforced by using sophisticated data techniques**. Examples include ML/AI-based approaches that can sometimes be seen as very opaque systems ("black boxes"), making it challenging to understand how decisions are made or why certain outcomes are produced. Moreover, any lack of transparency can erode trust, especially when insights drawn from the data can significantly impact individuals' lives, such as in healthcare, finance or criminal justice. For central banks, as well as more generally for policy institutions and the public sector, traceability and explainability therefore sit at the core of their decision-making processes. Fortunately, a number of approaches have been developed to demonstrate how granular information has been used to get some insights and support certain decisions. As illustrated by the recent Bank of Spain project related to bitcoin, in particular, the use

of a SHAP (SHapley Additive exPlanations; cf Lundberg and Lee (2017)) value estimation method can help to increase transparency and interpretability of ML models.

A further complication is that **granular data can inadvertently perpetuate biases and discrimination**, for instance if the data used to train AI systems reflect societal biases or contain inaccuracies. Moreover, the ongoing avalanche of (micro-level) information, combined with the limited traceability of data techniques and the increased attention to "alternative facts" driving good statistics out of policy debates, raises the risk of eroding public knowledge (IFC (2024); Garrett (2024)). Maintaining fairness, equity and transparency in decision-making processes is thus essential. The key is to recognise that the new techniques need to be properly managed to avoid these issues, such as algorithmic bias, and prevent unfair outcomes for certain demographic groups and discrimination. One solution put forward by Haghighi and Taillefer is to develop a dedicated data and analytics risk management framework in central banking. More generally, and as argued by Pablo García Silva, Vice-Governor of the Central Bank of Chile, the challenges above call for using the new data available in today's society in a smart way, by exercising discernment when generating and disseminating them to public authorities.

References

Aastveit, K, T Fastbø, E Granziera, K Paulsen and K Torstensen (2020): "Nowcasting Norwegian household consumption with debit card transaction data", *Norges Bank Working Paper*, no 17.

Aladangady, A, S Aron-Dine, W Dunn, L Feiveson, P Lengermann and C Sahm (2021): "From transaction data to economic statistics: constructing real-time, high-frequency, geographic measures of consumer spending", in K Abraham, R Jarmin, B Moyer and M Shaprio (eds), *Big data for twenty-first-century economic statistics,* National Bureau of Economic Research, pp 115–145.

Albagli, E, A Fernández, J Guerra-Salas, F Huneeus and P Muñoz (2023): "Anatomy of firms' margins of adjustment: evidence from the Covid pandemic", *Central Bank of Chile Working Papers*, no 981.

Aprigliano, V, G Ardizzi and L Monteforte (2019): "Using payment system data to forecast economic activity", *International Journal of Central Banking*, vol 15, no 4, pp 55–80.

Araujo, D (2023): "gingado: a machine learning library focused on economics and finance", *BIS Working Papers*, no 1122.

Araujo, D, G Bruno, J Marcucci, R Schmidt and B Tissot (2023): "Data science in central banking: applications and tools", *IFC Bulletin*, no 59.

Aurouet, D, D Del Giudice, A De Sanctis, J Franke, J Herzberg, M Osiewicz, R Peronaci and C Willeke (2023): "Indicators of granular exposures to climate-related physical risks for central banks' analytical purposes", in S Arslanalp, K Kostial and G Quirós-Romero (eds), *Data for a greener world: a guide for practitioners and policymakers*, IMF, chapter 4, pp 59–77.

Baker, S, N Bloom and S Davis (2016): "Measuring Economic Policy Uncertainty", *Quarterly Journal of Economics*, vol 131, no 4, pp. 1593–636.

Benigno, G, J di Giovanni, J Groen and A Noble (2022): "A new barometer of global supply chain pressures", *Federal Reserve Bank of New York Liberty Street Economics*, 4 January.

Bogdanova, B and B Buffet (2023): "Towards a full integration of SDMX in the data value chain – the present and the future of SDMX tools", presentation at the 2023 SDMX Global Conference, November.

Buckman, S, A Shapiro, M Sudhof and D Wilson (2020): "News sentiment in the time of Covid-19", *Federal Reserve Bank of San Francisco Economic Letter*, no 2020-08, 6 April.

Buda, G, V Carvalho, S Hansen, J Mora, Á Ortiz and T Rodrigo (2022): "National accounts in a world of naturally occurring data: a proof of concept for consumption", *Cambridge Working Papers in Economics*, no 2244.

Chetty, R, J Friedman, M Stepner and Opportunity Insights Team (2024): "The economic impacts of Covid-19: evidence from a new public database built using private sector data", *Quarterly Journal of Economics*, vol 139, no 2, pp 829–89.

Chodorow-Reich, G (2020): "Regional data in macroeconomics: some advice for practitioners", *Journal of Economic Dynamics and Control*, vol 115, 103875.

Croushore, D and T Stark (2001): "A real-time data set for macroeconomists", *Journal of Econometrics,* vol 105, no 1, pp. 111–30.

Cui, M, G Jiang and B Mosweu (2023): "Swimming in the data lake: an application to NielsenIQ Homescan", *IFC Bulletin*, no 59.

De Beer, B and B Tissot (2021): "Official statistics in the wake of the Covid-19 pandemic: a central banking perspective", *Theoretical Economics Letters*, vol 11, no 4, August.

Dilip, A and B Tissot (2024): "Development and maintenance of a security-by-security database", *IFC Guidance Note*, no 5, May.

Draghi, M (2016): "Central bank statistics: moving beyond the aggregates", speech at the Eighth ECB Statistics Conference, Frankfurt am Main, 6 July.

European Central Bank (ECB) (2022): *The Eurosystem Integrated Reporting Framework: an overview*, September.

ECB and Eurostat (2024): *ESS-ESCB Quality assessment report on statistics underlying the Macroeconomic Imbalance Procedure*, July.

Gabaix, X (2011): "The granular origins of aggregate fluctuations", *Econometrica*, vol 79, no 3, pp 733–72.

Garrett, A (2024): "The devil, the detail, and the data", *Journal of the Royal Statistical Society*, Series A: statistics in society, qnae063.

Huynh, K, H Lao, P Sabourin and A Welte (2020): "What do high-frequency expenditure network data reveal about spending and inflation during Covid-19?", *Bank of Canada Staff Analytical Notes*, no 2020-20.

Inter-Agency Group on Economic and Financial Statistics (IAG) (2017): *Update on the Data Gaps Initiative and the outcome of the Workshop on Data Sharing*, March.

International Monetary Fund (IMF), Inter-Agency Group on Economic and Financial Statistics and Financial Stability Board Secretariat (2023): *G20 Data Gaps Initiative 3 workplan: people, planet, economy – delivering insights for action*, March.

Irving Fisher Committee on Central Bank Statistics (IFC) (2015): "Data-sharing: issues and good practices", *IFC Report*, no 1.

——— (2017): "Big data", *IFC Bulletin*, no 44.

——— (2018): "Central banks and trade repositories derivatives data", *IFC Report*, no 7.

——— (2020): "Computing platforms for big data analytics and artificial intelligence", *IFC Report*, no 11.

——— (2021a): "Micro data for the macro world", *IFC Bulletin*, no 53.

——— (2021b): "Use of big data sources and applications at central banks", *IFC Report*, no 13.

——— (2022): "Machine learning in central banking", *IFC Bulletin*, no 57.

——— (2023a): "Data science in central banking: applications and tools", *IFC Bulletin*, no 59.

——— (2023b): "Data-sharing practices", *IFC Guidance Note*, no 3, March.

——— (2024): "Communication on central bank statistics", *IFC Bulletin*, no 60.

Israël, J-M and B Tissot (2021): "Incorporating microdata into macro policy decision-making", *Journal of Digital Banking*, vol 6, no 3, pp 233–50.

Jahangir-Abdoelrahman, S and Tissot, B (2023): "The post-pandemic new normal for central bank statistics", *Statistical Journal of the IAOS*, vol 39, no 3, pp 559–72.

Križman, I and B Tissot (2022): "Data governance frameworks for official statistics and the integration of alternative sources", *Statistical Journal of the IAOS*, vol 38, no 3, pp 947–55.

Lundberg, S and S Lee (2017): "A unified approach to interpreting model predictions", in *Advances in neural information processing systems* 30 (NIPS).

Nikoloutsos, S and O Sirello (2023): "SDMX, an international standard for micro data", presentation at the 2023 SDMX Global Conference, November.

Schubert, A (2021): "Big data for policymaking in economics and finance: the potential and challenges", in P Nymand-Andersen (ed), *Data science in economics and finance for decision makers*, pp 2–3.

Shapiro, A and D Wilson (2017): "What's in the news? A new economic indicator", *Federal Reserve Bank of San Francisco Economic Letter*, no 2017-10, 10 April.

United Nations Economic Commission for Europe (UNECE) (2024a): *Data stewardship and the role of national statistical offices in the new data ecosystem*, ECE/CES/STAT/2023/4.

——— (2024b): *In-depth review of linking data across domains and sources*, Conference of European Statistics, ECE/CES/2024/5.

——— (2024c): *Data governance framework for statistical interoperability*, High-Level Group for the Modernisation of Official Statistics, March.

United Nations Statistics Division (UNSD) (2023): *Microdata dissemination: Summary report*, Microdata Dissemination Task Force of the Inter-Secretariat Working Group on Household Surveys, February.

Wibisono, O, H Ari, B Tissot, A Widjanarti and A Zulen (2019): "Using big data analytics and artificial intelligence: a central banking perspective", *Capco Institute Journal of Financial Transformation*, 50th edition, "Data analytics", pp 70–83.

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Introductory remarks[1]

Pablo García Silva

Vice-Governor, Central Bank of Chile,
and Chair of the IFC

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Introductory remarks
Granular data: new horizons and challenges for central banks
Pablo García Silva, Deputy Governor of the Central Bank of Chile, Chair of the Irving Fisher Committee on Central Bank Statistics
July 15th 2023

Good morning. On behalf of the Irving Fisher Committee, I would like to express my sincere appreciation for your participation and the insightful contributions you are bringing to this conference held on the occasion of the congress of the International Statistical Institute. Many thanks also to our host, the Bank of Canada, for making this happen, especially to Tetiana and Maryam and their dedicated team, and of course to the IFC secretariat at the BIS.

Today's opportunity to reflect on Granular Data and the new horizons and challenges associated for Central Banks is a good example of how we can all collectively progress and innovate in the world of Central Banking. The new perspectives brought about by the sharing of our own experiences is essential to further deepen our understanding of the issues involved and stimulate our future work using micro level information.

Granular data is in particular key to better understand economic developments, design public policies, and in turn assess their impact on the economy. Accordingly, central Banks are, as policy-making institutions but also as important collectors of micro-level information, progressively incorporating more granular insights into their day-to-day work. Let me provide a few ideas on the opportunities that this trend presents:

First, granular data can greatly improve our understanding of the functioning of the national economy. By analyzing individual-level data, central banks can gain deeper insights on economic behaviors, allowing for more accurate considerations of the impact of policies and economic shocks over firms and individuals. This makes it easier to detect heterogeneities among economic agents and also to update policies more rapidly. For instance, firm-level data enables a more precise understanding of the factors driving business performance, investment decisions, employment dynamics, among other things. This knowledge helps in turn central banks to formulate more effective policies as they can be better targeted to specific sectors or groups of firms, as was evident during the recent Covid-19 pandemic.

Second, the incorporation of various sources of granular information can support the production of new, more timely indicators that can usefully complement traditional official statistics. The incorporation of micro data has already enabled several institutions to monitor economic activities in real-time and detect early warning signs of imbalances or vulnerabilities. The insights gained this way have helped central banks to, for instance, produce more accurate and timely macroeconomic

forecasts, enhance the efficiency of their policy decisions and improve their evaluation of the overall performance of the economy.

Third, granular information can also be instrumental to extend the scope of our analyses beyond national borders, by providing new insights on international phenomena. In particular, micro-level data can help central banks to better identify cross-border connections and dependencies, which are vital for understanding the functioning of the global economic network and its knock-on effects on national countries. For instance, going more granular can enhance our understanding of international trade flows, supply chain dynamics, and financial vulnerabilities and contagion effects. Such knowledge is essential for monitoring global risks, assessing potential systemic disruptions and preparing coordinated policy responses. Let me in particular stress the important role played by central banks and financial supervisors in exchanging micro-level insights, for instance to produce comprehensive global financial statistics under the umbrella of the BIS or to promote knowledge-sharing of national experiences under the framework of the Irving Fisher Committee.

Fourth, the greater attention to granular data provides an excellent opportunity for central banks to further refine their focus on data management issues. Implementing robust systems to protect the privacy, confidentiality, quality and integrity of granular data is a key objective, not least to further improve the reputation of our institutions and the adherence to internationally agreed statistical principles and standards. In particular, the micro data revolution calls for developing new and better data governance frameworks to cover the collection, storage, processing, and dissemination of statistical information. It is a key opportunity to set clear guidelines and protocols that uphold data quality and in turn fortify public trust in what central bank are doing.

Let me stop here. Today's seminar a well as the forthcoming World Congress of the International Statistical Institute will allow us to discuss these matters in greater details. This represents a unique opportunity for learning, sharing ideas, and networking with peers from around the world. I am sure that the insights gained will help us to develop a more holistic and integrated view on central banks' use of granular data.

This represents an exciting journey, and the stakes are high, not least to ensure public trust in the information we produce an in the policy actions we take. Many thanks to all of you to contribute to these efforts.

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

The secret ingredient of being relevant? Microdata. Making the move from ready-made to fresh statistics[1]

Luís Teles Dias and Lígia Maria Nunes,
Banco de Portugal

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

The secret ingredient of being relevant? Microdata.

Making the move from ready-made to fresh statistics.

Luís Teles Dias[1], Lígia Maria Nunes[2]

[1]*Banco de Portugal*, Statistics Department, Lisboa, Portugal – ldias@bportugal.pt
[2]*Banco de Portugal*, Statistics Department, Lisboa, Portugal – lmnunes@bportugal.pt

Abstract

In the realm of statistics, much like in culinary arts, the quality of the ingredients used can significantly influence the final outcome. The Banco de Portugal has embraced this philosophy by transitioning from traditional, "ready-made" statistical reports to dynamic, microdata-driven insights. This shift is essential for producing relevant and timely information that meets the evolving needs of society, particularly in response to rapid economic changes.

Historically, the Banco de Portugal has been producing and publishing bank interest rate statistics for over three decades, harmonized across the euro area since 2003 under the Monetary and Financial Institutions Interest Rates (MIR) framework. These statistics provided a standardized view but often lacked the granularity required to address specific, emergent issues. The period of low or negative interest rates rendered these statistics less newsworthy, but the landscape changed dramatically with the onset of rising interest rates in mid-2022. The predominantly variable-rate nature of Portuguese housing loans – over 90% indexed to Euribor – exposed many households to significant financial risk, demanding a more detailed statistical approach.

The introduction of new statistical indicators in September 2022 marked a pivotal shift towards leveraging the comprehensive data available in the Portuguese central credit register (CCR). The CCR's vast database, encompassing over 200 attributes per loan contract, provided the granular data needed to develop these fresh statistics. These new indicators offered detailed insights into loans granted by banks to households, which were previously obscured in aggregated data. The new set of indicators addressed e.g., the various subtypes of housing loans, a common typology of interest rates based on different indexants, the distribution by percentiles of the amounts of regular instalments, the differentiation between new and renegotiated loans and the early amortization of loans.

This enhanced statistical output has been well-received, evidenced by a surge in engagement with Banco de Portugal's statistical portal, BP*stat*. The average monthly page views of our press releases on interest rate statistics shot up from around 650 in the 1st half of 2022 (before the publication of the new indicators) to around 1800 in the 2nd half (after the new indicators) and remained at this level in the 1st half of 2023. This fact clearly underscored the public's demand for these new statistics.

The transition to microdata-driven statistics required more than just access to data; it required a user-centric approach and active listening to the public's needs. By prioritizing clarity and relevance in its reporting, the Banco de Portugal ensured that the new indicators were understandable and useful to the general public. For instance, instead of using technical terms like "initial period of rate fixation," the new reports focused on whether interest rates were fixed or variable and how frequently they were reviewed.

Significant findings from these new indicators included a notable shift in the housing loan market: the proportion of the stock of loans with variable rate fell from over 90% to near 70% within a year. The differential between fixed and variable interest rates also decreased, with both

averages converging around 4.20% by May 2023. Additionally, there was a noticeable increase in loan early amortization (either full of partial), reflecting borrowers' responses to rising interest rates.

The impact of these new statistics extends beyond immediate public engagement. They have demonstrated the potential of microdata to enhance the relevance and responsiveness of official statistics. Looking forward, the Banco de Portugal aims to apply this approach to other domains, such as non-financial corporate data, to address issues like potential increases in sales margins amidst inflation.

In conclusion, the Banco de Portugal's move towards utilizing microdata for fresh, detailed statistics exemplifies its commitment to staying relevant and responsive in an era of rapid economic change. This approach not only meets regulatory requirements but also provides the public with vital information, enhancing trust and utility in official statistics. The key takeaway is that relevance in statistics, much like in cuisine, depends on the quality and freshness of the ingredients – data that is timely, detailed, and directly responsive to the needs of users.

The secret ingredient of being relevant? Microdata

Making the move from ready-made to fresh statistics

Luís Teles Dias
Lígia Maria Nunes

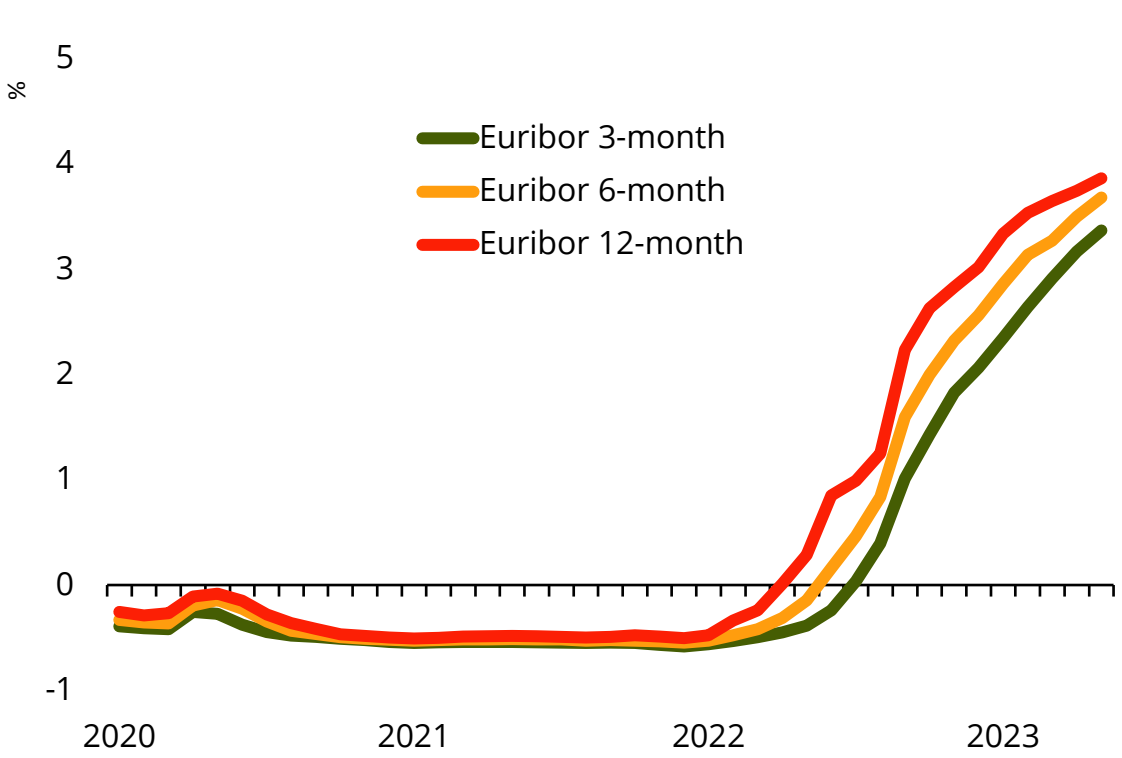
•

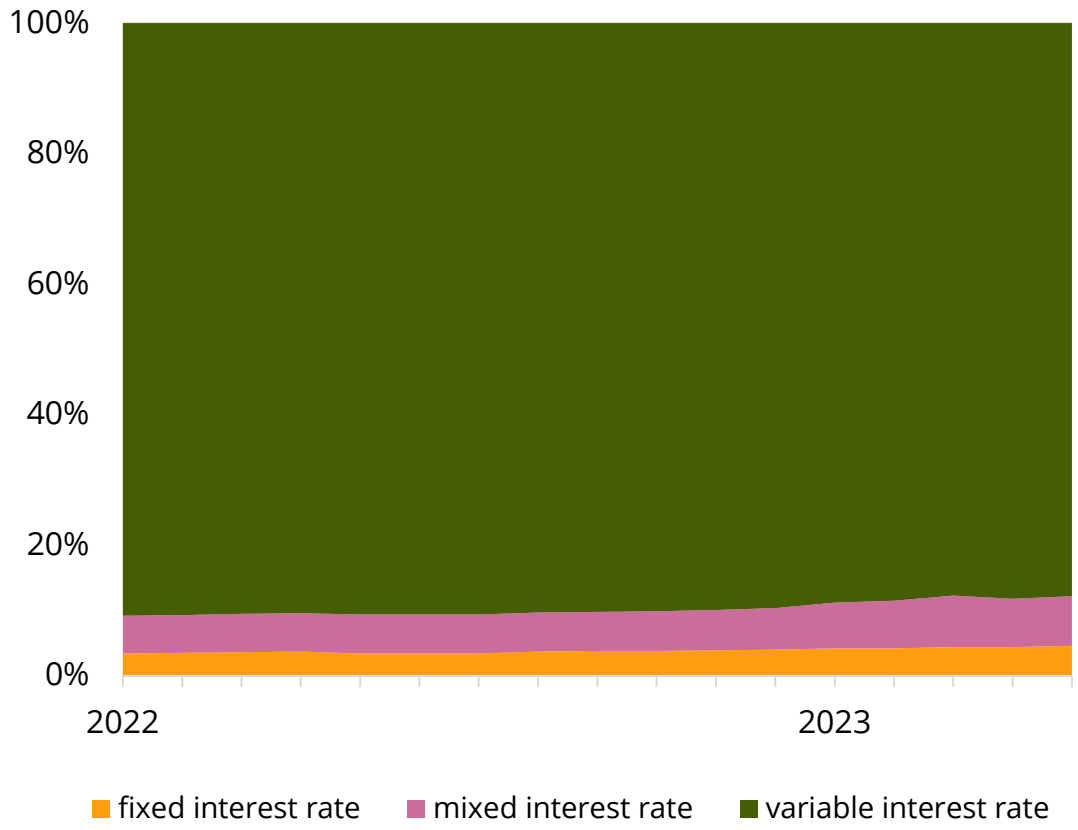
Banco de Portugal



BANCO DE PORTUGAL
EUROSYSTEM

90% of the stock of loans for permanent own housing had a variable interest rate





The Portuguese Central Credit Register has more than 200 variables on every loan above EUR 50.





Better data



Skills on data management and analytics



Identify the hot topics

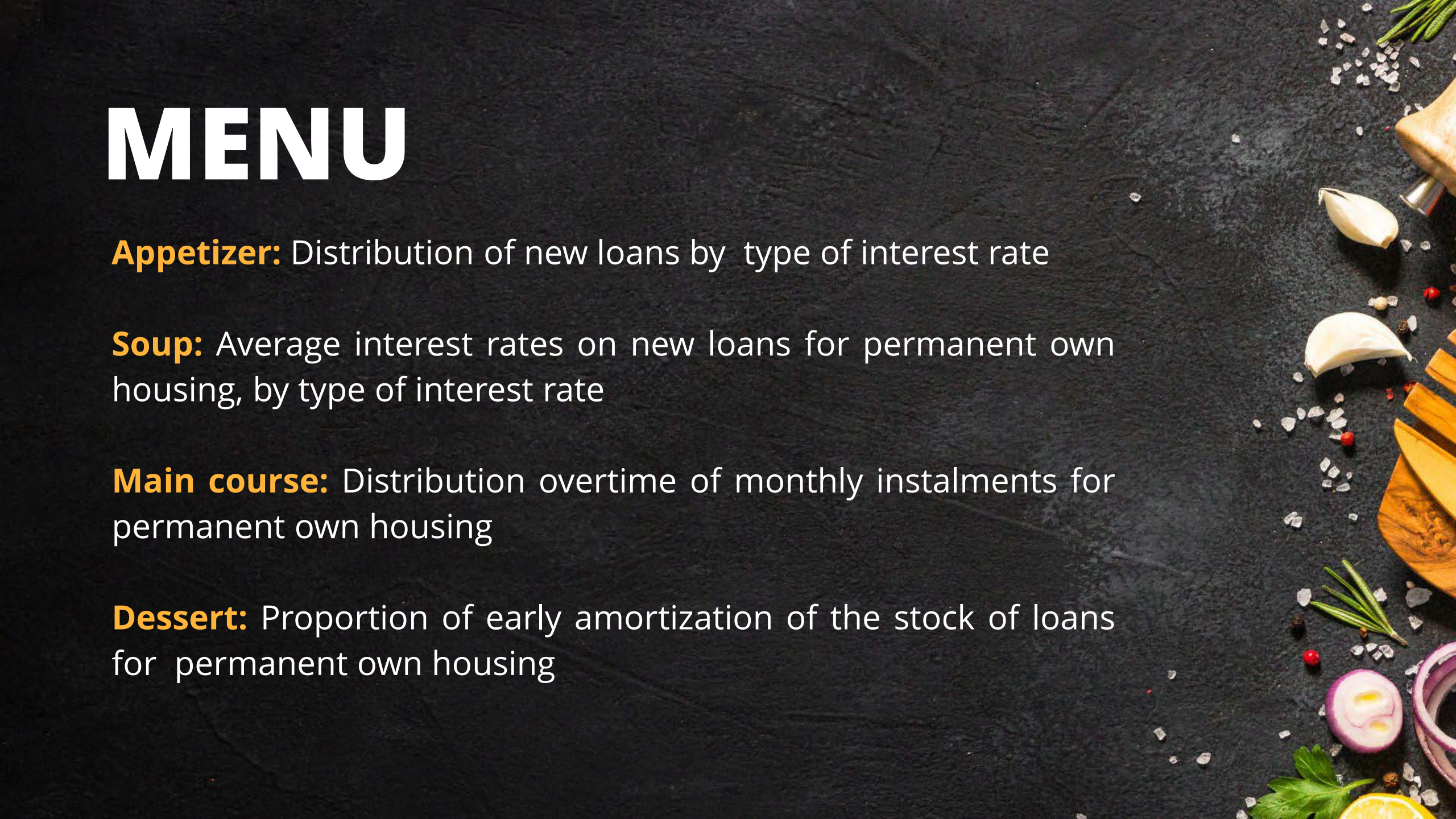
MENU

Appetizer: Distribution of new loans by type of interest rate

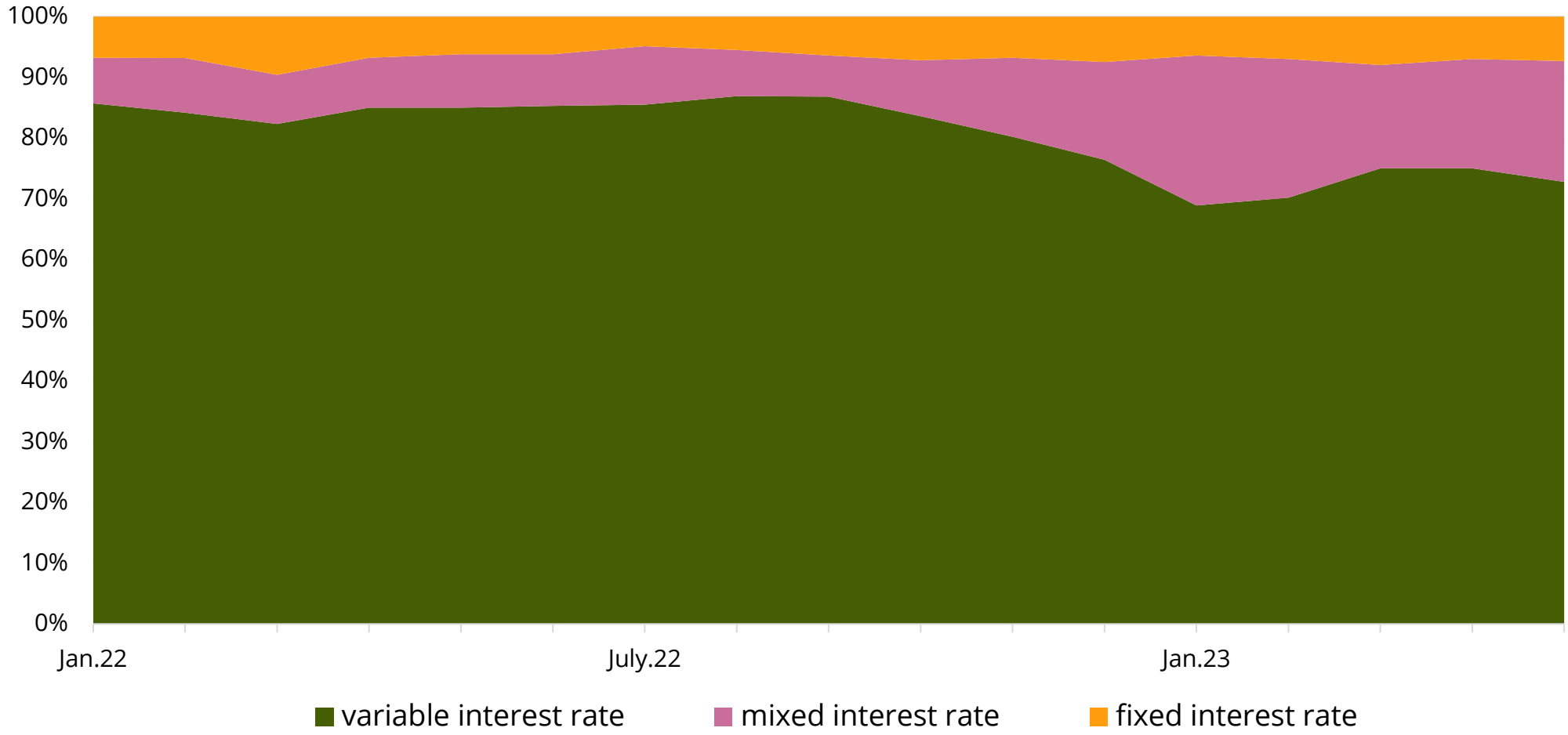
Soup: Average interest rates on new loans for permanent own housing, by type of interest rate

Main course: Distribution overtime of monthly instalments for permanent own housing

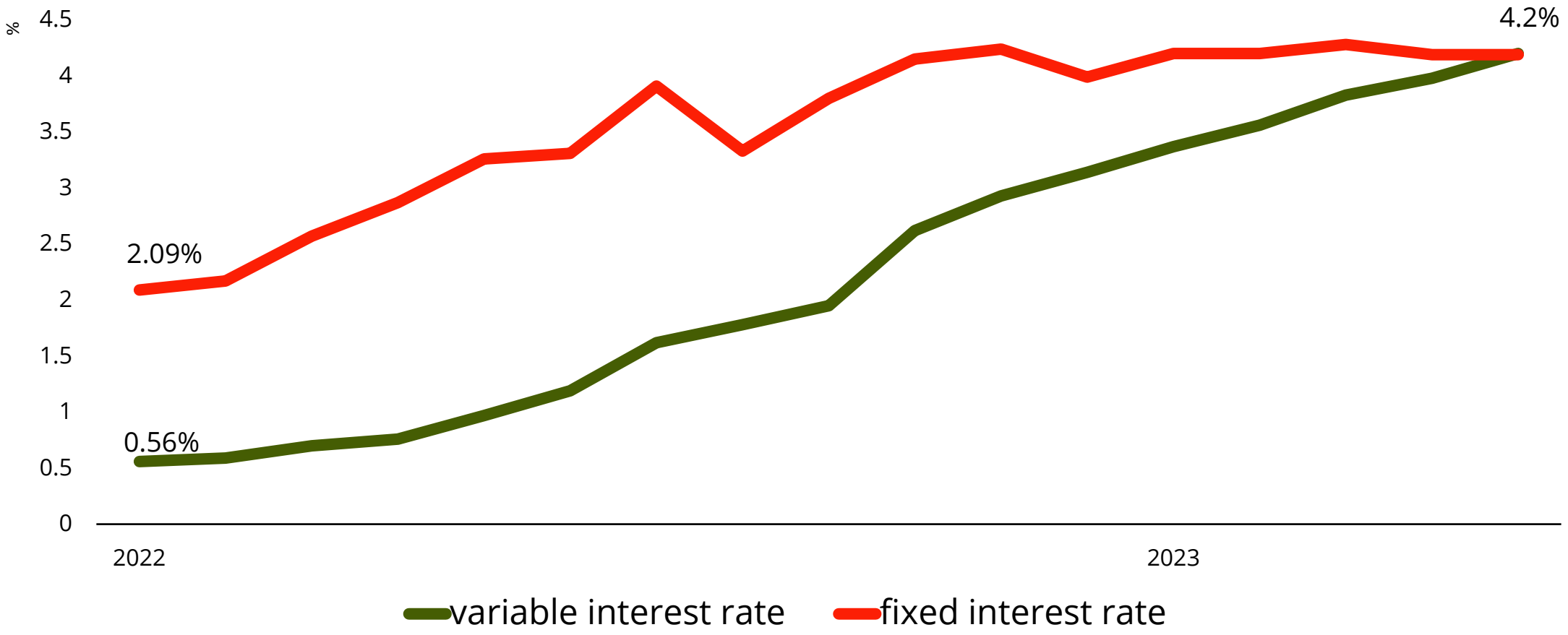
Dessert: Proportion of early amortization of the stock of loans for permanent own housing



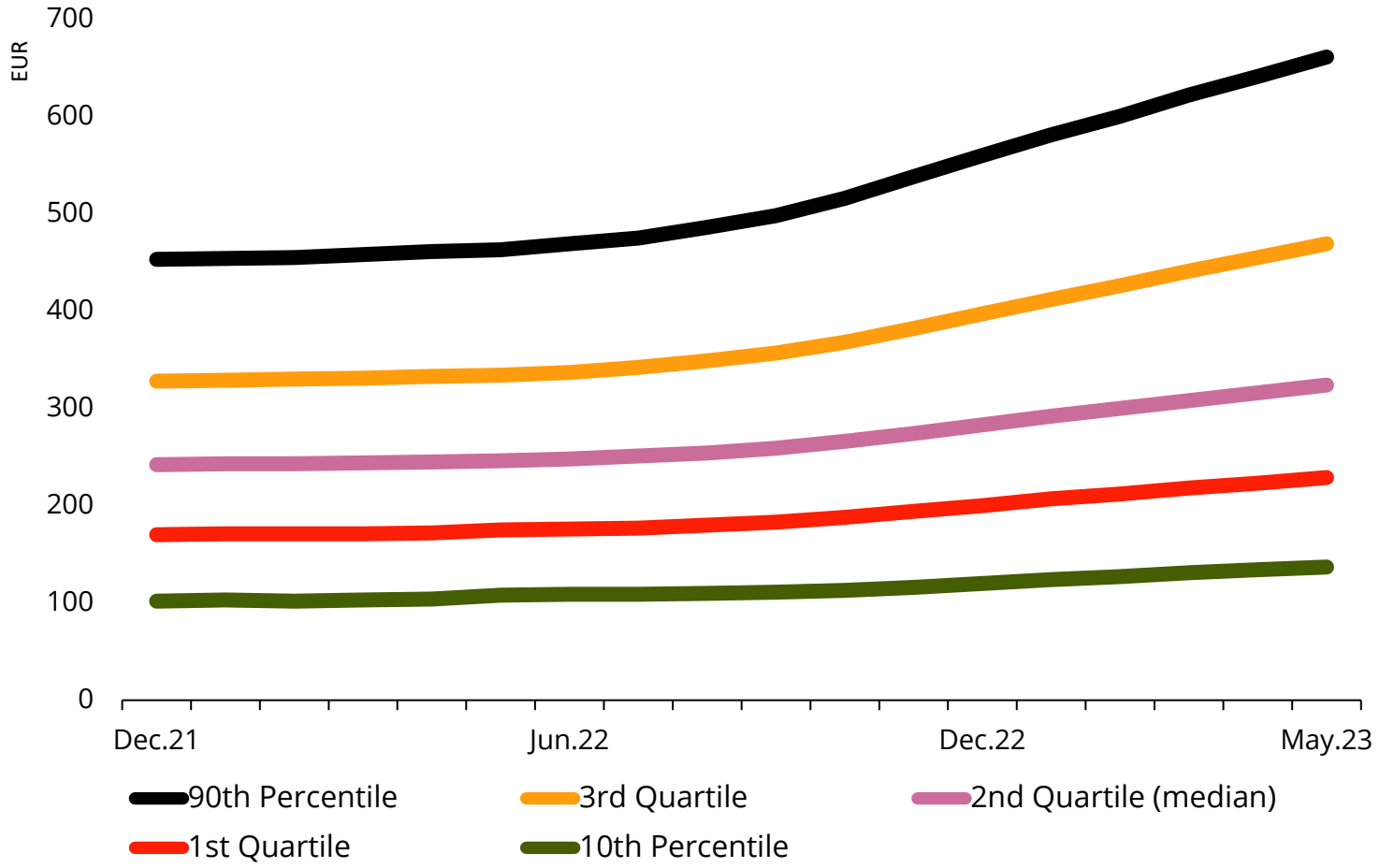
New loans for permanent own housing with variable interest rate decreased from October 2022 onwards



Gap between average interest rates on new loans for permanent own housing with variable and fixed interest rates closed



Monthly repayment of loans for permanent own housing increased, on average, between EUR 35 and EUR 208 between December 2021 and May 2023



+208 EUR
46%

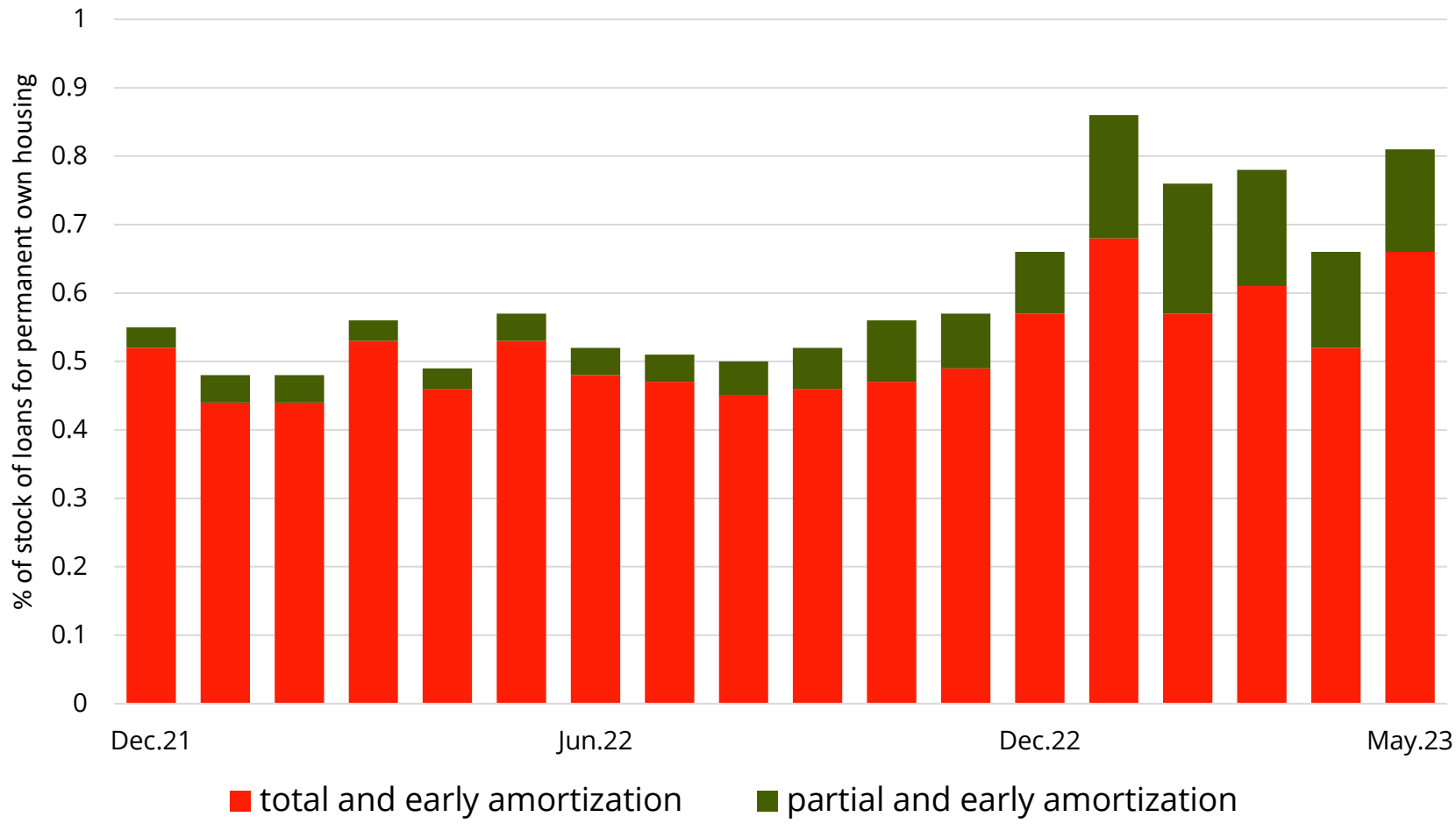
+141 EUR
43%

+82 EUR
34%

+59 EUR
34%

+35 EUR
34%

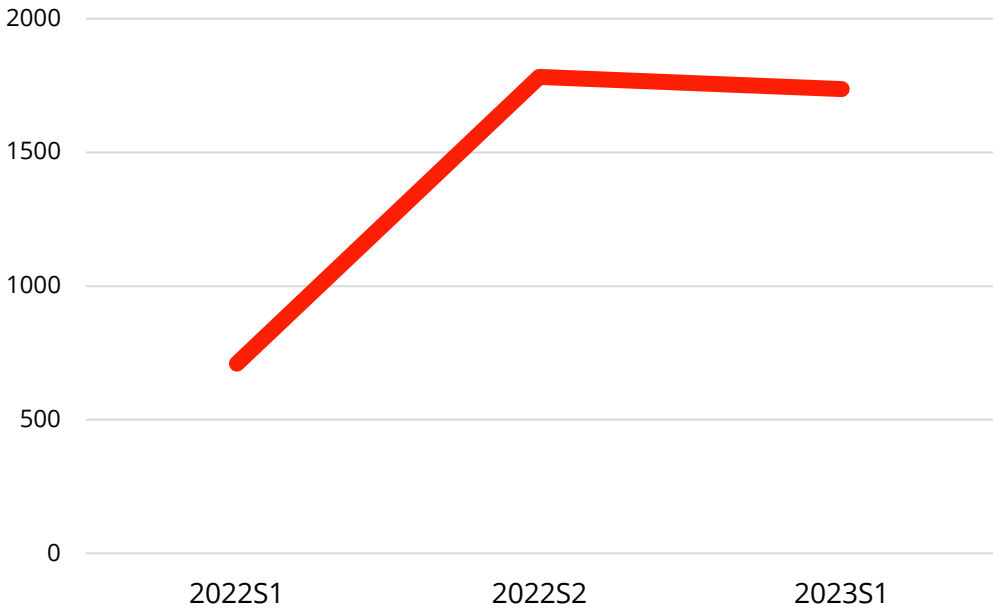
Early amortizations of loans for permanent own housing reached 0.67% of the stock in May, continuing the upward trend observed since September 2022



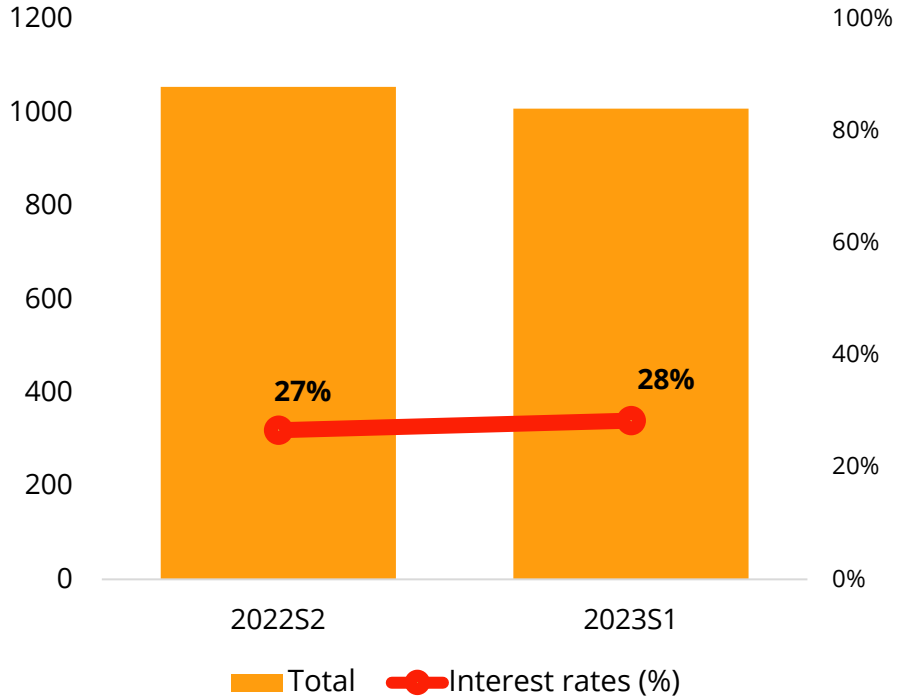
28% of the news published by the media on statistics released by the Banco de Portugal were on interest rates

Statistical press releases – interest rates

Average monthly page views



News published by the media on statistics released by Banco de Portugal





Remaining relevant is about continuing to be users' choice year after year.

SO, SPEND TIME WHERE YOUR USERS SPEND TIME

The secret ingredient of being relevant? Microdata

BoC-IFC satellite seminar on "Granular data: new horizons and challenges for central banks"

Ottawa, July 15, 2023

Luís Teles Dias | ldias@bportugal.pt
Lígia Maria Nunes | lmnunes@bportugal.pt

BANCO DE PORTUGAL
EUROSYSTEM



Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Granular statistical data – Are we charmed with a blessing or a curse?[1]

Jasminka Dodeva Kovaceva,
National Bank of the Republic of North Macedonia

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Granular statistical data – Are we charmed with a blessing or a curse?

Jasminka Dodeva Kovacheva[1]

Abstract

The past has shown that granular statistical data very often comes at the forefront for policy makers. This is particular the case, during or after crisis episodes, when new sources of risks are detected. Individual level data can provide punctuality, clarity and timeliness, as surveys data or aggregated administrative data cannot. However, what is important is how to determine the limit: where is the border between the usefulness and the responsibility for data producers and users; where is the border between the benefit and the burden for data reporters. In this presentation, we investigate the importance behind the collection of granular statistical data for Central banks and the obligation it creates, by tackling the specific experience of the National Bank of the Republic of North Macedonia. The presentation will showcase the granular statistical data we use, how we try to converge from history to new ways of data collection without losing the quality, and how we mitigate the risks associated with managing and using big and detailed databases. There is always a need and a challenge of accepting and introducing new data sources, but there is also a need to preserve what is established, useful and broadly accepted in the country. And the trick is to balance it all and find what the best option is. In the presentation, we show what we think the best option is for our case.

[1] Email: DodevaJ@nbrm.mk, corresponding author. National Bank of Republic of North Macedonia.

The views expressed in this paper are those of the authors and do not necessarily represent the views of the National Bank of Republic of North Macedonia (NBRNM).

Contents

1. Introduction 3
2. What is granular data and how detailed data should be? 4
2.1 What works for us? 6
3. Cost and benefits of data granularity 14
4. Data management 15
5. Final remarks/conclusion 16
6. References 17
Annex 1: Detailed data included in the reporting form for loans 19

1. Introduction

The barriers of statistical research have a heavy duty of providing timely, quality and compatible data so that end-users, such as analysts, policy makers and private sector can make most out of it and can embroider the data into sound decisions for better future. This is an obvious task. However, what is not obvious and is more complex is how to put the theory into practice.

The producers of statistical data are facing many challenges and pressures of responding to the ever-increasing demand of high volume data. Reasons behind this are many, globalization and digitalization being the leaders of the changes in the last few decades. New financial instruments are continuously emerging that could pose higher threat to the financial stability and are putting pressure on defining new regulations. Multinational companies that have trading agreements all over the world that are difficult to capture. Special Purpose entities with activities that are scarce but complex. Economic Unions that provide free, completely liberalized, movements. There are many opportunities offered for economic transactions and all of them have to be statistically captured in order to gather the facts and to provide a valuable insights of the economic (and social) reality. Granularity of data can provide these insights and give answers to many questions; however, there is always a challenge to determine what is sufficient and where is the limit of the granularity. When the granular data becomes a burden instead of an advantage.

In addition, new economic reality provokes introducing new data sources and limiting the “old” ones that are in many cases connected with or originated from certain economic and financial regimes in the countries (trade regimes, restricted financial transactions, pre-approvals and conditions). However, the peculiarity is for the institutions responsible for gathering statistical data to investigate and implement or adopt the combination of both. The need of statistical data should not pose a limit or a delay in the financial transactions, or in the free movement of goods and services, but in the same time, we should not lose the quality and the details of some of the already established statistical systems and cause additional burden to the reporters.

The question of the level of detail of data is something that we will try to cover thought-out the paper and we will try to answer, in certain degree, what works best.

This paper aims to contribute to the ongoing national and international discussions of the relevance of granular statistical data. In addition, to shed light on the risks involved with overburdening the reporters, overflow of information, as well as the responsibility of the data management especially with preserving the individual data and their usage only for statistical purposes.

The paper is organized as follows. The first part covers the definition of granular data and the dimensions it comprises, with a separate part capturing the details and practical aspects in the case of the National Bank of the Republic of North Macedonia. The second part provides insights of the cost and benefit of having detailed data. The third one gives an overview of the need for a complex data management and the ways to mitigate the risks. The last section concludes.

2. What is granular data and how detailed data should be?

In the last decade, the phrase “less is more” has become very popular in many different areas. We have minimalistic style in architecture; design and art in general, used to create larger effect. In addition, it is used very often to allow people to focus on what matters most and not to be distracted by the “irrelevant”. An oxymoron that it is based on the value of simplicity and became a way of creating a life of more.

However, this phrase does not apply in the field of statistics. In statistics, less data means less information for analysts, policy makers and other data users to make solid, data driven decisions. It means less information for quality crosschecks, data comparability and resolving asymmetries. Therefore, it means less quality data and “blind” decision-making. We can freely say nowadays, in statistics “more is more”, and lately we are witnessing a trend of “everybody going granular”.

We see international and country initiatives of enchanting the databases with more details in very different areas, just to name a few: OECD initiative on Municipal Migration Database; or AnaCredit (Analytical Credit Dataset) project set up by the ECB. In addition, on a global level many initiatives are arising for creating unified database for different identifying elements, for example the introduction of the Legal Entity Identifier (LEI) that enables clear and unique identification of legal entities participating in financial transactions or the IMFs initiative for Annual Exchange of Sector Information on Securities Issuers (ESI).

With the rapidly changing nature of how the world works with the globalization and digitalization the need for more granular data is essential in order to understand its complexity and the new phenomenon, new instruments arising.

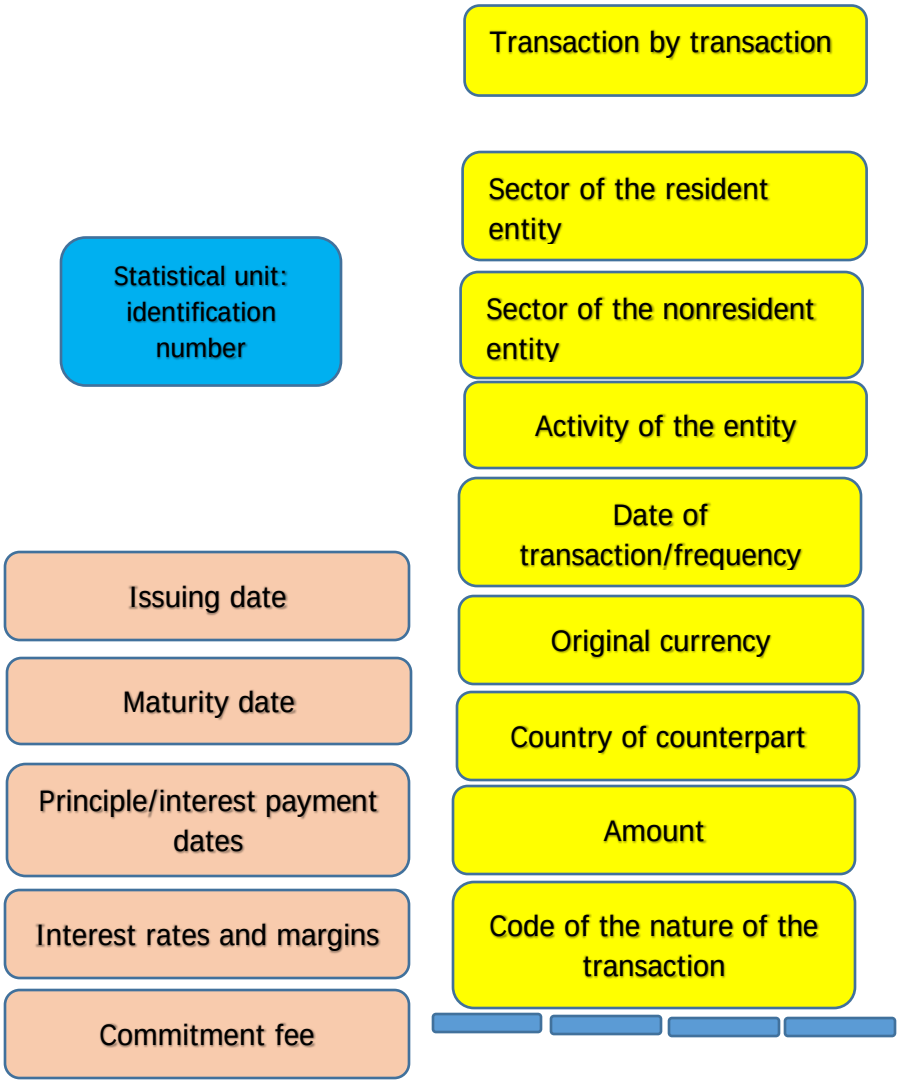
With globalization and multi-international companies gaining pace, the need for data increased even more. The need to understand the international transactions became dependent from the detailed data and the ability to understand their complexity through data linkages. Without having the proper information of how these transactions are conducted, which are the involved parties, what is the type of the transaction and what is their ultimate goal it is difficult to estimate their effect on the economy and how those transactions would influence other entities and instruments. Only granular data can reveal the nature of the integration of the companies in the global economy.

The importance of data is unquestionable, but what is more important in statistical terms is what is considered granular data and how detailed granular data should be.

Granularity is the amount of information contained within the data structure. Granular data usually involves micro-data or data broken down across multiple dimensions of information. In statistical terms, granular data usually comprise data on a level of transaction, instrument and/or account collected on an individual statistical unit.

The level of detailed data is determined primarily by the statistical requests of data, according to the internationally accepted methodology, and secondly, by the demand for additional information for a sound decision-making. Just for an illustration, Picture 1 shows the most common information contained in the granular *external* statistics data set.

Picture 1: Detailed external statistics data



As presented on the picture, the first step is gathering data on transaction-by-transaction level and the identification of the statistical unit through its identification number. Second is to attach all information available from all data sources. In general, this information covers the characteristics of the underlining financial instrument or economic transaction. Most common information of each instrument are: the value of the transaction, sector of resident and non-resident, activity, date of transactions (frequency), original currency, country of the counterpart and code capturing the nature of the transaction. Depending of the data source these dimensions can be broaden with a lot more detailed data, for example data for loans can include information of date of agreement, dates for disbursement, principle and interest payments, interest rate, margins, commitment or other type of fee and charges and so on. Not all instruments have the same information structure, however it is important to determine the ones that are essential and every data source should provide in order to have unified data structure and form integrated database.

Collecting granular data does not comes without a challenge. Data producers should always be careful, because there is a significant risk of information overflow. On the other side, for data analysts there is a trade-off between the advantage of performing a micro level analysis and the risk of losing the focus of the macro perspective. The border between producing, analyzing too many granular data versus aggregated data have to be set very carefully.

2.1 What works for us?

In regards to granularity of statistical data, every country must pinpoint what is important for its specific economic circumstances. The National Bank of the Republic of North Macedonia (NBRNM), as an institution responsible for external statistics data, monetary statistics data and financial accounts data, have long established tradition in collecting data through various data sources. Data are collected through administrative data sources; direct deporting or estimates are used. All data have different structure, which pose a challenge to create one integrated database. The intention is with each newly introduced data source the necessary data structure to be ensured.

The standards for granular data, which are used, are based on the standards agreed and implemented on a global level. These standards result in a large efficiency gains for both data providers and users. Standardization leads to easing the reporting of detailed data, especially because in that way many data reporters could include the reporting obligation within their IT system.

Data sources that the National Bank use are designed to obtain specific information to fulfill the methodological requirements. Many of the direct reporting are relatively newly introduced in order to capture the financial instruments that are gaining importance in recent decades and to provide information according to the up-to-date statistical methodology. On the other side, some data sources have been historically established and only certain changes were implemented. With every introduction of new data source, it is very important that timely and effective interaction between reporters and drafters of statistical reporting regulation exists. Collaboration between public and private sector, as well as public-public collaboration in the case of administrative data, remains the best approach for creating integrated system for statistical reporting.

If we analyze more specifically the data sources used for external statistics data, recently two trends have emerged and have taken pace. With the process of cross-border liberalizations, many data sources have become insufficient to provide full quality data and to capture the economic reality, and even some data sources were completely abolished. In the same time, the process of liberalization created an environment where the statistical reporting have to be limited in order not to be a burden to the realization of the financial transactions. In some cases, this led to decreasing the quality of the reported data and in addition decreasing the possibility of conducting detailed quality crosschecks.

The data sources that the NBRNM use have different structure, containing different level of details. When introducing new direct reporting, the process of defining the information that it would contain is twofold. First, it is essential to start from the international statistical requirements for the specific type of statistical data. Second, how detailed the data source should be also depends on the current relevance (in value terms) of that instrument for the country and its

future prospects. Very often those instruments that are most relevant usually are covered more extensively (in a sense of frequency, level of details, periodicity). What is also very important is enabling an automated process and the possibility of the reporting to be integrated in the IT systems of the reporters, and if possible to be used for other purposes as well (e.g. connected with their accounting, or data that are used for supervisory purposes as well). What is also very important is determining the reporting unit and weather there is a possibility of indirect reporting, or reporting though different institution for e.g. regulators, in order not the burden the units with double reporting.

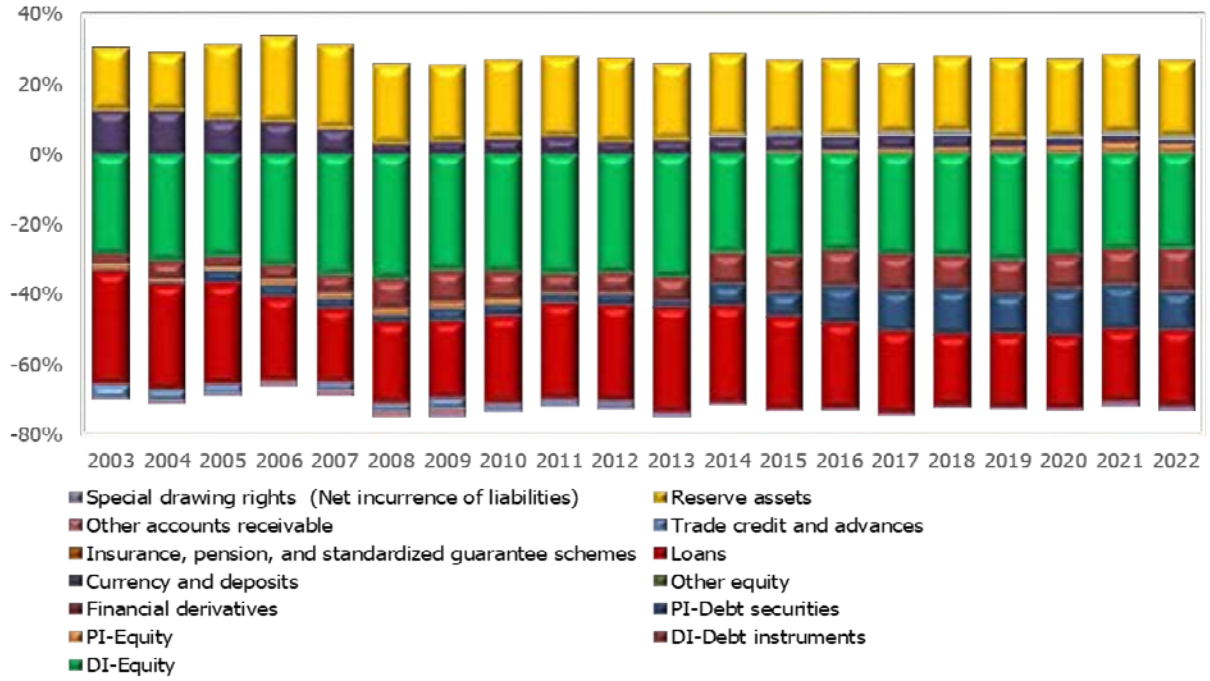
Having this in mind, we have established, or enhanced the historically inherited, several granular data sources. In the following paragraphs, few of the data sources used in external statistics will be elaborated in more details.

Loan by loan

When analyzing the International Investment Position (IIP) of the Republic of North Macedonia it is very notable that "traditional" financial instruments, such as *Loans*, are still one of the most important instruments for foreign financing compared with portfolio investments or financial derivatives (Graph 1). If we add the part of the loans, which are included in the direct investment-debt instrument component, then the importance is even higher. This is to be expected having in mind the status of the country as a developing country.

Graph 1: Structure of International Investment Position of the RNM, by instruments

(in %)

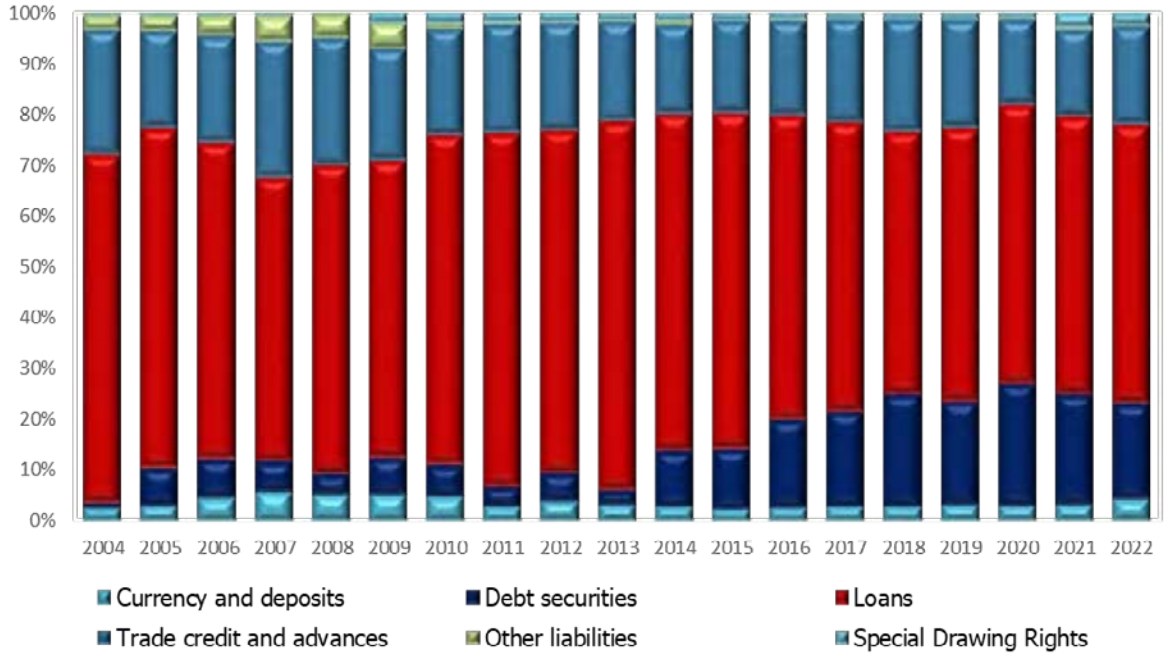


Source: NBRNM

If we analyze the debt instruments within the Gross External Debt Position of the country, the percentage share of the Loans is ranging from 68.6% in 2004, to 72.6% in 2013 and 52% in 2018, which shows relatively constant level over the years.

Graph 2: Structure of Gross External Debt, by instruments

(in %)



Source: NBRNM.

In this environment, the focus is very much on analyzing the effect that this instrument have on the economy and future financial transactions in regards to new disbursements or repayments of debt. The detailed data on loans are used mostly for assessing the ability to service outstanding debt and estimate the future external borrowing, as well as for making economic projections and analysis of a country's overall economic position.

In addition to this need, detailed data on loans are required by international institutions, e.g. by the World Bank for fulfilling the obligations of Debtor Reporting System (DRS). Even though the requirements for loan-by-loan information within this system is only for the public sector, we have applied the same concept for all reporting units. Therefore, the loan component, in our reporting system, is covered at the most detailed level.

According to the regulations (Law on Foreign Exchange, Decision on the manner and the terms of recording and reporting on the concluded credit operations and Manuals for the manner and the conditions for individual reporting of loans taken by residents), each entity that borrows or extend a loan from/to a non-resident have to report that loan to the National Bank within 10 working days after signing the loan agreement. All changes in the agreement and all realized transactions have to be reported within 5 working days. The level of reporting is quite in details,

each reporting set contains several forms: one form is for the general terms of the agreement, second form is for disbursements, third type of form is for schedule of principle payments and the forth is specification in case of multiple creditors in the loan agreement. The most detailed document is the first form. This form contain several information about the loan agreement, as presented on Annex 1.

As can be seen from the picture, all information are on a loan-by-loan basis for each resident, containing information that further provide the basis for additional calculations and deriving information on future schedule of interest payments, stock and flows of accrual interest, arrears, residual maturity. In addition to this, several codes for the type of transactions are applied when recording the disbursement or payments of principal and interest. With this information appropriate classification of the flow within the integrated International Investment Position is ensured, as well as for other statistics for e.g. Financial accounts. All basic elements collected provide a solid ground to create a whole set of information that can be utilized differently, according to the needs.

Historically analyzed, this data source was first introduced because of certain regulatory bans on the movement of the capital to and from abroad, which are usually covered by the term "foreign exchange regime", defined in the Law for foreign exchange operations, but also in order to fulfill the international statistical requirements. Currently, most of the capital transactions, such as credit operations, direct investments, investments of non-residents in securities and in real estate are completely free. Therefore, the recording of loan transactions are for statistical purposes only and it does not require submitting any supporting documents. Having this detailed data provide rich possibilities for different analyses, comparison with other data sources in order to increase the quality, fulfilling the international methodological standards and quickly adopting to the changes in the recommendations.

The ownership of such database very often creates a challenge. This is especially in cases such as classifying personal information of the reporters; requests from other institutions to obtain detailed data for not statistical purposes; occasions where as a convention in order to ensure symmetry among countries certain classifications should be re-located (e.g. "all interbank positions other than securities and accounts receivable/payable are classified under deposits"[2] even though the loan agreement clearly testify otherwise). In addition, as mentioned earlier, maintaining big systems that process and calculates detailed level data on a daily basis, require sophisticated technology and application, and advanced IT human resources.

The decision to collect detailed type of data reports for the Loan component might be seen as too cautious or prudent, but the importance of the instrument on the overall perspective of the economy overwhelms the challenges and drawbacks.

International Transaction Reporting System (ITRS)

Another data source that requires closer attention (not only because of its importance but also for its controversy) is the "historically inherited" International Transaction Reporting System (ITRS). This data source have been originally inherited from the former system in which very rigid foreign exchange regime existed, with clearly defined regulation of all inflows and outflows

[2] Balance of Payments and International Investment Position Manual (BPM6), page 87, 2009, IMF.

from/to abroad. One of the main reasons behind this was the need to sustain the stability of the domestic currency, through controlling and approving the cross-border transactions. Throughout the years, this database full of information became less important for transaction regulation, but became more important for statistical purposes, as one of the main data sources for Balance of Payments compilation (BoP). The system was additionally enriched with separate codes for each type of transaction in order for clear classification within the BoP.

However, with the process of globalization and liberalization, free movement of capital, growth of multinational corporations, and development of foreign financial markets and the surfacing of new financial instruments have led to questioning the sustainability and justification of the ITRS as a statistical data source with adequate coverage. In addition, the difference between the methodological statistical guidance versus the coverage of the data source (e.g. accrual vs payments) have become even broader.

Within the EU, the liberalization of the transactions and the pressure for decreasing the expenses and reporting burden of deposit-taking corporations, as well as speeding the process of realization of each cross-border transaction, led to adopting the EU Regulation 924/2009 on cross-border payments, by the European Parliament and the Council of the European Union. With this regulation, it was foreseen that from January 1st, 2016, the deposit-taking corporations within the EU Member states would not conduct classification of the cross-border transactions for the need of the statistical systems. However, the regulation allows the "Member States to continue to collect aggregated data or other relevant readily available information, provided that such collection has no impact on the straight through processing of the payments and can be fully automated by payment service providers"[3].

The ITRS that the National Bank uses is a so-called "closed system" in which full integration is achieved between stock at the beginning of the period, transactions and stock at the end of the period. It provides data mostly on a daily basis, in original currency. In the last two decades, the National Bank has made significant improvements in the coding list describing the nature of the transaction, in order to be aligned with the international classifications of the transactions and in order for the residents to have easier selection of the code when classifying the transaction. In addition, the National Bank has invested largely into the IT infrastructure for providing fast, easy and secure electronic transfer of data by the banks.

With all being said, how useful is this data source and what kind of data it provides? As mentioned, the reporting provides data on a transaction-by-transaction level, mostly on a daily basis, in original currency of the transaction, without applying any threshold for foreign currency payments[4]. The information it contains mostly capture the ID of the resident legal entity or the code for the resident natural person, the code for the nature of the transaction, the country of the non-resident, an information about the non-resident ("free text"), the ID of the bank through which the transaction is realized.

The biggest advantages of the system are the short deadline in which data are obtained from the reporters (5 working days after making the payment); granular data obtained through census (all

[3] REGULATION (EC) No 924/2009 on cross-border payments in the Community.

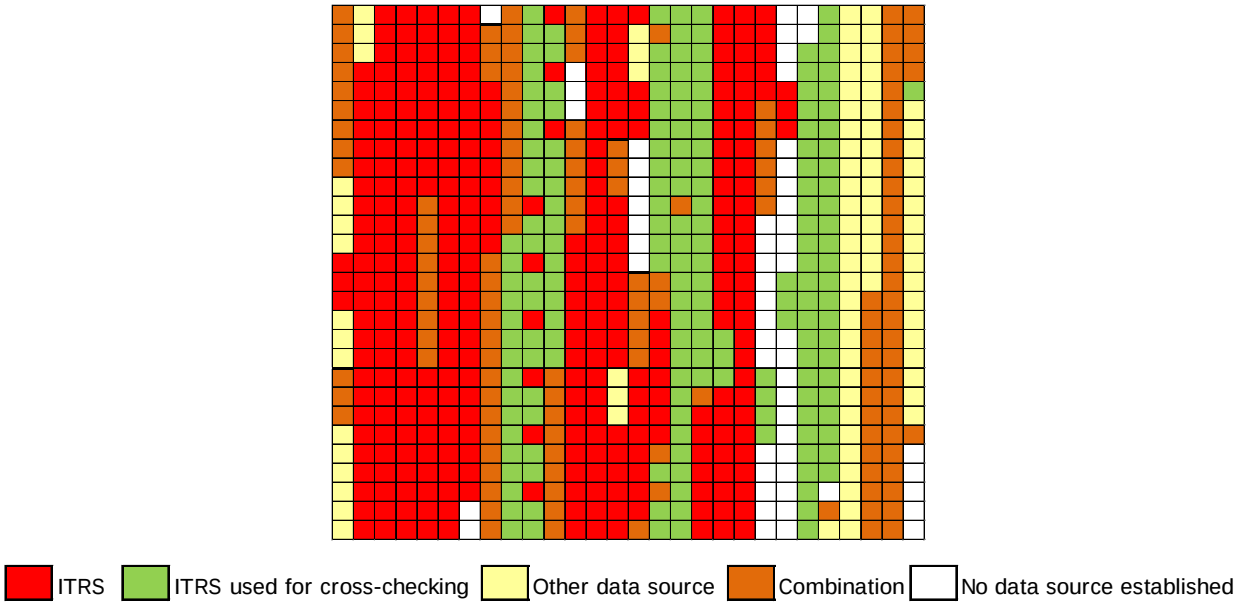
[4] For payments in Macedonian denars, the transactions under 100.000 denars (ca. EUR 1500) are reported on an aggregated basis.

data are reported, without threshold, through total of 14 banks); relatively good data quality, classified according to the statistical requirements; the whole process of reporting is completely automatized; the responsibility of providing correct code of the nature of the transaction is on the resident receiving/sending the money; the companies and the banking sector have already established internal work procedures which is easing the whole process.

The ITRS as a data source is mainly used in the BoP compilation. In addition to being a data source, the ITRS has its application as a source of information for crosschecking of the quality of data received through other data sources. This characteristic is very much appreciated because it provides granular information like no other data source can: who is the reporting unit, when the transaction happened, and why it occurred. What is also very important benefit is providing information of the ID of the legal entities that are involved in certain type of transaction. This information is used for defining the sample of data reporters for a direct reporting (surveying through a sample of companies). These are two of the most important side-benefits from the ITRS that have a direct effect on increasing the data quality.

Picture 2: Data sources - ITRS usage in the Balance of Payments ("The Rubik`s cube")

*Each square present an item from the BoP standard presentation



Source: Authors presentation based on NBRNM data.

Some disadvantages of the ITRS that should be mentioned is the lack of compatibility with the internationally accepted methodological standards. Transactions that are based on payment ("cash") basis are not aligned with the accrual requirements stipulated in the statistical manuals. In the same time, the moment of payment does not capture the moment of change of the economic ownership, which is required for the statistical reporting. In addition, this data source does not capture transactions that do not require payments (e.g. compensation, debt forgiveness) or stock data. One cannot overlook the fact that coding of the transaction could pose delays or

could “slow” the realization of the international payments, especially in the case of inflows where the money are not available to the receiver until the appropriate code is stated. The risk of having this disadvantage should be very carefully managed and if the system is properly set, with an advance IT infrastructure within the Central bank and the commercial banks, it can be mitigated or at least significantly decreased.

Even though there are some shortcomings in this data source, one must agree that it has values that other data sources does not. The trick is to adjust it in order to incorporate the new requirements, but to keep all the advantages it provides. In cases where, as in RNM, the historical habits are so enstamp in the whole work culture of the companies and the system is already incorporated in their accounting systems, being that the ITRS also relies on accounts, is very important to seize the values and utilize the benefits it provides.

Security by security

One granular data source that was relatively newly introduced was the security by security direct reporting for portfolio investments abroad made by residents. Having in mind that portfolio investments became more significant for the country during the recent years and in an absence of a possibility for having a direct access to any global security-by-security database, in July 2014, the National Bank made changes in legislation and introduced new manual for reporting on security-by-security basis.

The reporting was made obligatory for all residents who invest in securities in foreign financial markets, and who at the time of the introduction of the new report, had the possibility to invest according to the Law of foreign exchange operations. Those were deposit-taking corporations, insurance companies, investment funds and pension funds. The reporting to the National Bank is done directly or indirectly through the supervisory agencies. Reporting is ISIN by ISIN, on a monthly basis, and includes all securities (foreign and domestic) in the investor’ portfolio acquired on foreign financial markets.

The reports include mainly detail data for:

- Investor (ID, Name, Institutional Sector);
- Security (ISIN, Currency of issue, Currency of trade, Market and Country of trading, Type of security, Issuing date, Maturity date, Type of interest rate etc.);
- Issuer of the security (Name, Country, Institutional Sector);
- Stock at the beginning of the period (Number/Market price per security at the beginning of the period);
- Transactions (Number/Market price per security of purchased/sold securities);
- Other changes (Price changes, Exchange rate changes, Other changes);
- Stock at the end of the period (Number/Market price per security at the end of the period).

With this reporting, the National Bank established an integrated system of data collection on the stocks and transactions in securities, electronically transferred to the National Bank through web application, by imposing only low burden on the reporters due to the fact that many supervisory agencies collect these data for their own purposes and then report this data to the National Bank. The system provides full reconciliation of data, better quality of data (market value, accrual basis) as well as bigger details of data (unified Data Structure Definition-DSD).

The introduction of this data source enabled the compilation and dissemination of Coordinated Portfolio Investment Survey Data (Assets), with complete breakdown of data by Currency of Denomination of the Securities Held, by Sector of Resident Holder, by Sector of Nonresident Issuer, and cross tabulation by Sector of Resident Holder, and Economy and Sector of Nonresident Issuer for Specified Economies.

With these, the RN Macedonia fulfilled the last SDDS Plus requirement for external statistics data and starting from 28.1.2019 it became a SDDS Plus adherent, together with 18 other countries at that time.

However, new challenges lays a head. Starting from 2019, with the beginning of the second phase of the Stabilization and Association Agreement with the EU, free movement of capital related to portfolio investment was ensured. These means investors in foreign securities can also be individuals and non-financial corporations, which are not captured with the security-by security data source. Currently, these investments are monitored through other data sources (mainly ITRS) and the amounts are not that significant. However, potential data sources are being analyzed.

New Integrated Reporting System for monetary data – Ongoing Project ISIDORA

The need for granular data, mainly for statistical purposes (monetary, financial and external statistics), but also for supervisory purposes, have prompted starting a project for integrated reporting system for monetary statistics data. With the main goal of the project to align the National legislation with the European legislation, the project ISIDORA started in 2016. The intention was set at collecting balance sheet data from deposit-taking corporations (including the Central Bank) and saving houses.

Key elements that were identified as essential for collection, that are not captured with the current system, were:

- currency structure;
- geographical distribution of non-residents' instruments;
- sector classification of non-residents;
- sub-sectoral classification of residents;
- flows data, as opposed to stocks (in order to separate the influence of price and exchange rate changes on the movement of monetary aggregates, as well as other changes in volume).

Many of these dimensions represent a requirement for statistical reporting, however, in order to minimize the reporting burden, the ISIDORA project was broaden in scope by capturing also the requirements for the Credit Registry, as part of the supervisory purposes.

With the implementation of this project, the National bank would achieve a complete reorganization of the current data reporting system (from accounts to instruments described through predefined dimensions and attributes) and an integration of various statistical reports, but also integration part of the supervisory reports. So far, a lot have been done, but many more activities still awaits, such as development of by-laws (decisions and reporting manuals), technical performance of the systems in the National Bank and with reporters, continued collaboration and

holding meetings and workshops with reporters, testing of the system, parallel reporting and finally accepting the system as an official data source.

3. Cost and benefits of data granularity

The cost and the benefits of data granularity can be analyzed from different perspective. The decision on what prevails, the cost or the benefit, largely depends weather that perspective is from the angle of the data collector, data reporter or data user. On average most of the common benefits can be summarized in several different points.

When we consider statistical data, the most important benefit from granular data is the possibility of fulfilling the statistical methodological requirements. Even more so, the availability of microdata provide an opportunity to revise the time-series data in cases of updates in the methodology. In addition, microdata provide a good ground for increasing the quality of data by crosschecking of data between different data sources and the quality checks can be done even on single statistical unit. In addition, the analyses of data, when individual data are available, can be done on a single Balance of Payment report produced for one statistical unit, providing a very direct view of the effects that entity have on the cross border transactions, the possible pressure on the foreign exchange market or foreign reserves assets.

Lately, models for estimation have become very widely used among statisticians, especially when data are scarce or less frequent. They provide information that can easily be applied once the initial parameters are determined. However, what became very much in focus on a global level is also the large detected data asymmetries between countries. Not to neglect the importance of estimation models, but in situations when different countries have different input parameters for the estimates, it is very possible that the asymmetries would increase between those two countries. Having micro level data instead of estimates increase the quality of the analyses and ease the process of detecting the potential differences in data.

In addition, when stating the benefits we cannot overlook the fact that granular data provide an opportunity more detailed data to be published and be available for different kind of users (academia, researchers, students, officials, private sectors, media). It provide possibility of conducting multidimensional analyses and it stipulates larger creativity and determining the causality of the effects.

If we focus on the needs of the data collectors and analysts, at first glance, it is obvious that more granular data means larger possibility of it use. However, knowing the obvious does not mean we should lose sight and neglect the possible traps that the possession and reporting of detailed data can have.

When analyzing the costs, it is important to make a distinction whether we discuss about the costs of the reporters or of the collectors. One of the most common mentioned cost of microdata reporting is the burden of data reporters that it creates. Reporting microdata usually means reporting not only detailed level data but also high frequency data that can be very time consuming process. However, when stating this burden one have to have in mind that it is not

always straightforward to determine what creates more pressure microdata or aggregated data. In cases where the IT structure is enhanced and all reporting is automated and from already available databases then is much easier to transfer the database from the reporting unit to the data collector, than to recalculate, aggregate according to certain rules. However, even in these cases certain aspects have to remain in focus of the data collectors: like making sure not to collect the same information twice or to have a clear vision of the data that are useable (or would become in use in the future) and not to overburden the reporters with excess data.

In order to decrease the reporting burden, while keeping the details of data, it is very important to ensure harmonization and standardization of data. When harmonizing the data on global level, it is very important to keep in mind the interest of all involved parties, as much as possible.

From the aspects of data collectors, the possession of a detailed database does not come without a responsibility. Maintaining the confidentiality of the individual data and having in place profound procedure of how, when and for what purpose these data can be used, is at most importance. Managing these micro databases usually implies that the institutions responsible have to be able to provide sufficient human capital, devoted to managing and preserving these data, as well as developed IT resources that can enable secure process of statistical production (acquisition, production, exploration and dissemination of data).

4. Data management

At the National Bank, the focus today is on producing high quality and granular data and in managing this high-volume of data. Granular data should be properly collected, organized, and managed to maintain its integrity. Data quality and validation processes are crucial when dealing with granular data, because errors on individual level can have a significant impact on the overall analysis.

When working with granular data it is important to adhere to privacy regulations and ensure appropriate data anonymization or aggregation when publishing in order to protect individuals' identities.

In the Republic of North Macedonia, the National Statistical System is defined in the Law of state statistics and in the five-year Statistical Programme. The State Statistical Office and the National Bank (NBRNM) are the main institutions barriers of statistical research, together with several ministries and agencies who are responsible for fewer statistical surveys noted in the five-year statistical programme. The regulatory framework consist of Law on State Statistics and the Law of the National Bank, as well as agreements for cooperation and data sharing between institutions. The legislation, among other things, covers the statistical confidentiality as well. Confidential data are data that allow direct or indirect identification of the statistical unit. All confidential data collected for statistical purposes should be protected under the law.

Special procedures are in place at the Bank that regulate the derogations in which confidential data can be disclosed. The exceptions cover requests by external auditors or regulatory authorities in order to perform their duties, court orders, or when a consent from the reporter is

obtained. Within the institutions, access to the confidential data is allowed only to the employees responsible for collecting, analyzing and managing the data. Measures are in place to ensure data protection (password protection, restricted access, encryption of data or data transfer through special application to ensure privacy). Internal procedures are also established for regular data exchange between departments within the institution or in cases of special requests.

5. Final remarks/conclusion

The conclusions can be summarized in several key points. In our complex world that we live in, having available granular statistical data is a necessity, in order to understand the economic reality. It provides insights of how our economic, financial, social dynamics works. Adding micro data one to another and piling, building more details on top of that can create the picture of the movements and structures in the country and how they are interconnected with the rest of the world, like pieces of a puzzle. Nowadays is not the question on do we need granular data, but the question is more on how to obtain them without creating additional burden, to obtain them with less effort and how to best utilize them and protect them.

What is evenly important to granular data is to have high frequency and quality data. In order to ensure that, having appropriate data sources is essential. The adaptation of different data sources depends entirely of the type of the economy-level of development and the reporting culture in the country. Even though harmonization of data is a necessity for providing similar, aggregated data structure on a global level, “cut-off models” and harmonization of data sources in different countries is not always a good choice. Each data source have some particularities that each country should assess and determine whether it would be a good fit for its case. Even though some data sources are considered “historical”, “not appropriate” and should be neglected, one have to keep in mind the advantages those data sources have and the possibility that little change can ensure keeping their value and continue to use them. The focus should not always be on “one side of the coin”, but to re-think and re-invent the options presented and maximize the already established granular data sources while respecting the needs of all involved parties and having in mind the necessity for ever evolving modernization in statistics.

6. References

Marques C. (2014):"How granular data should be to compile BOP statistics", IFC Bulletin No 37, January.

OECD/Statistics Finland (2021), Measuring Trade in Value Added for Finland", OECD Publishing, Paris.

Balance of Payments and International Investment Position Manual (BPM6), 2009, IMF.

International Reserves and Foreign Currency Liquidity: Guidelines for a Template, IMF, 2013.

External Debt Statistics: Guide for Compilers and Users, IMF, 2013.

OECD Benchmark Definition of Foreign Direct Investment 4th edition BD4.

Regulation (EC) No 184/2005 of the European Parliament and of the Council of 12 January 2005 on Community statistics concerning balance of payments, international trade in services and foreign direct investment. OJ L 35, 8.2.2005, p. 23, as amended by Commission Regulation No 555/2012 (OJ L 166, 27.06.2012).

Commission Regulation (EC) No 601/2006 of 18 April 2006 implementing Regulation (EC) No 184/2005 of the European Parliament and of the Council as regards the format and the procedure for the transmission of data. OJ L 106, 19.4.2006.

Commission Regulation (EC) No 1055/2008 of 27 October 2008 implementing Regulation (EC) No 184/2005 of the European Parliament and of the Council, as regards quality criteria and quality reporting for balance of payments statistics. OJ L 283, 28.10.2008, as amended by Commission Regulation (EU) No 1227/2010 (OJ L 336, 21.12.2010, p. 15-16).

Commission Implementing Regulation (EU) No 228/2014 of 10 March 2014 amending Regulation (EC) No 601/2006 implementing Regulation (EC) No 184/2005 of the European Parliament and of the Council on statistics concerning balance of payments, international trade in services and foreign direct investment, as regards the format and the procedure for the transmission of data. OJ L 70, 11.3.2014, p. 16–17.

Regulation (EU) 2016/1013 of the European Parliament and of the Council of 8 June 2016 amending Regulation (EC) No 184/2005 on Community statistics concerning balance of payments, international trade in services and foreign direct investment. OJ L 171, 29.6.2016, p. 144–152.

Law on the National Bank of the Republic of North Macedonia (Official Gazette of the Republic of Macedonia No. 158/10, 123/12, 43/14, 153/15, 6/16 and 83/18 and Official Gazette of the Republic of North Macedonia No. 110/21).

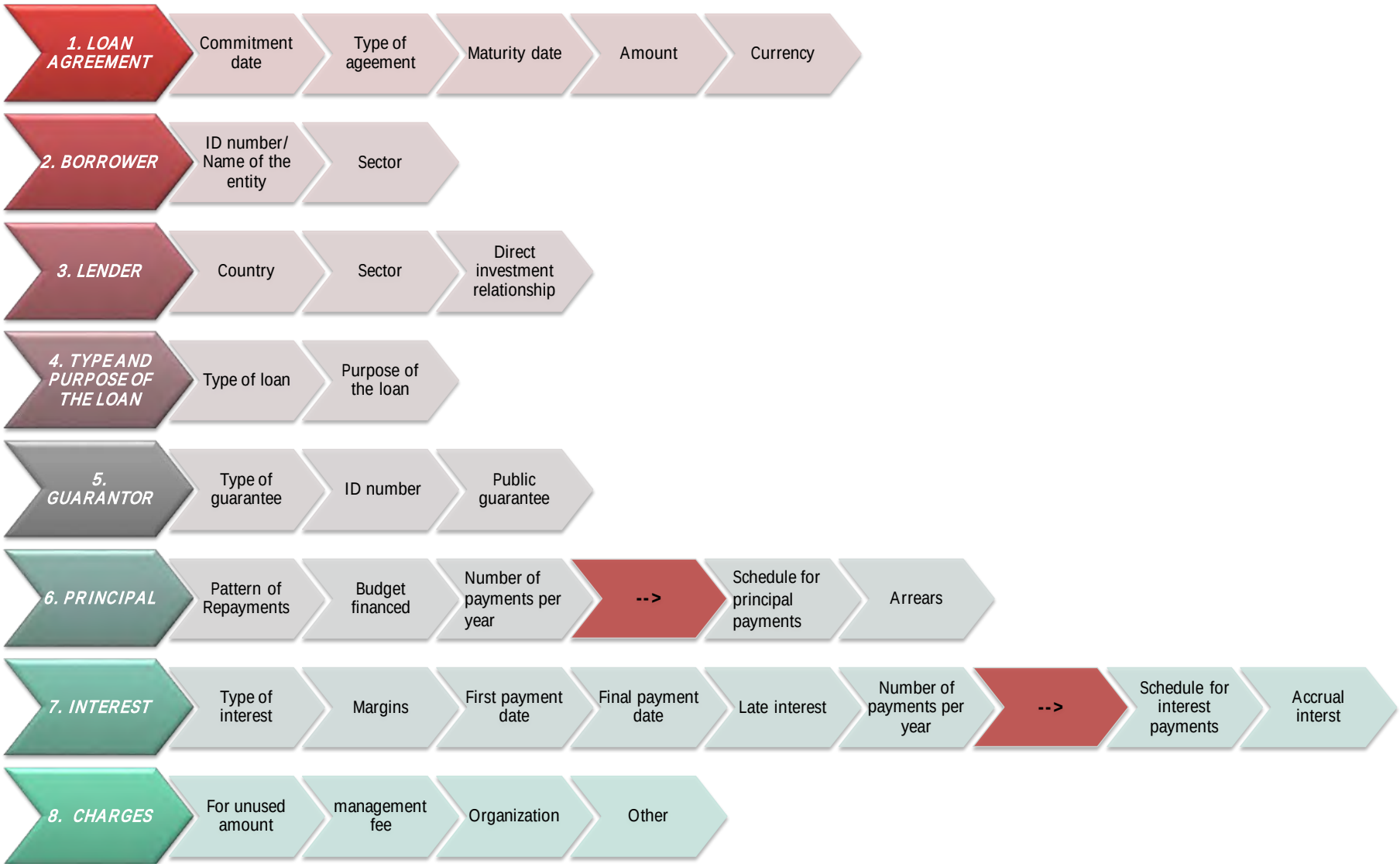
Law on Foreign Exchange Operations (Official Gazette of the Republic of Macedonia No. 34/01, 49/01,103/01, 51/03, 81/08, 24/11, 135/11, 188/13, 97/15, 153/15 and 23/16 and Official Gazette of the Republic of North Macedonia No.110/21).

Decision on the manner and the procedure for reporting on transactions with non-residents (Official Gazette of the Republic of Macedonia No. 87/14).

Decision on the manner and the terms of recording and reporting on the concluded credit operations (Official Gazette of the Republic of Macedonia No. 173/15 and 105/18).

Decision on the manner of performing international payment operations (Official Gazette of the Republic of Macedonia No. 42/11).

Annex 1: Detailed data included in the reporting form for loans





NATIONAL BANK OF THE REPUBLIC OF NORTH MACEDONIA

Granular statistical data – Are we charmed with a blessing or a curse?

Jasminka Dodeva Kovacheva

BoC-IFC Satellite Seminar,

Bank of Canada

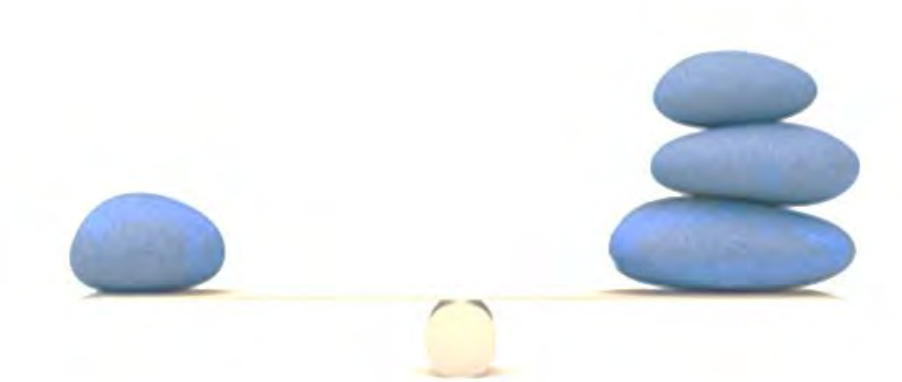
July 15, 2023

Granular statistical data – Are we *charmed* with a blessing or a curse?

Introduction

- "New reality" - globalization and digitalization, new financial instruments, multinational companies, special purpose entities-bring new need for more data;
- How to make a choice which data sources to be used?...
- …and what level of details to be used;
- Is there a prototype?;
- The obligation and responsibility that obtaining granular data caries.

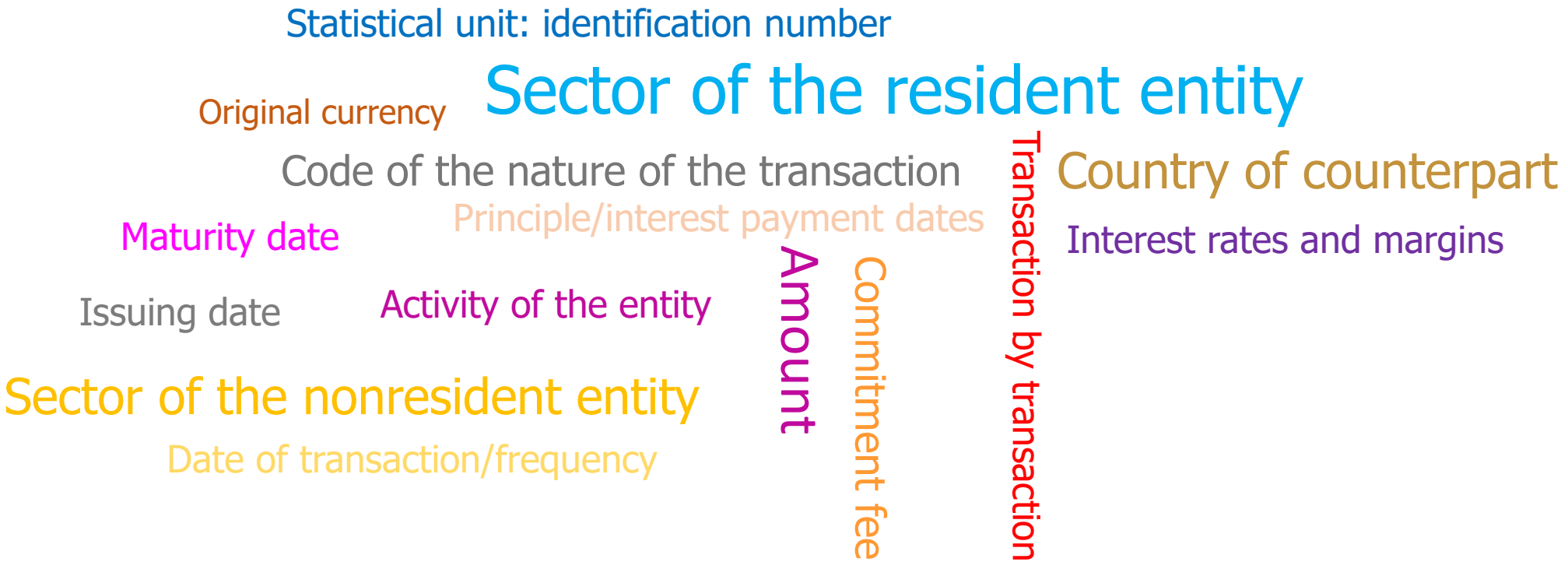
What is granular data and how detailed data should be?



Less is More?

What is granular data and how detailed data should be?

- Global initiatives for granular databases – "everybody is going granular"



* External Statistics Data

What works for us?



What works for us?

- Accepting and implementing internationally accepted standards;
- Electronically reporting and integrated IT systems;
- Maximizing the level of details with minimizing the reporting burden;
- Collaboration between public and private sector, as well as public-public sector;
- Custom made data sources;
- History matters, re-think how to utilize it for the future.

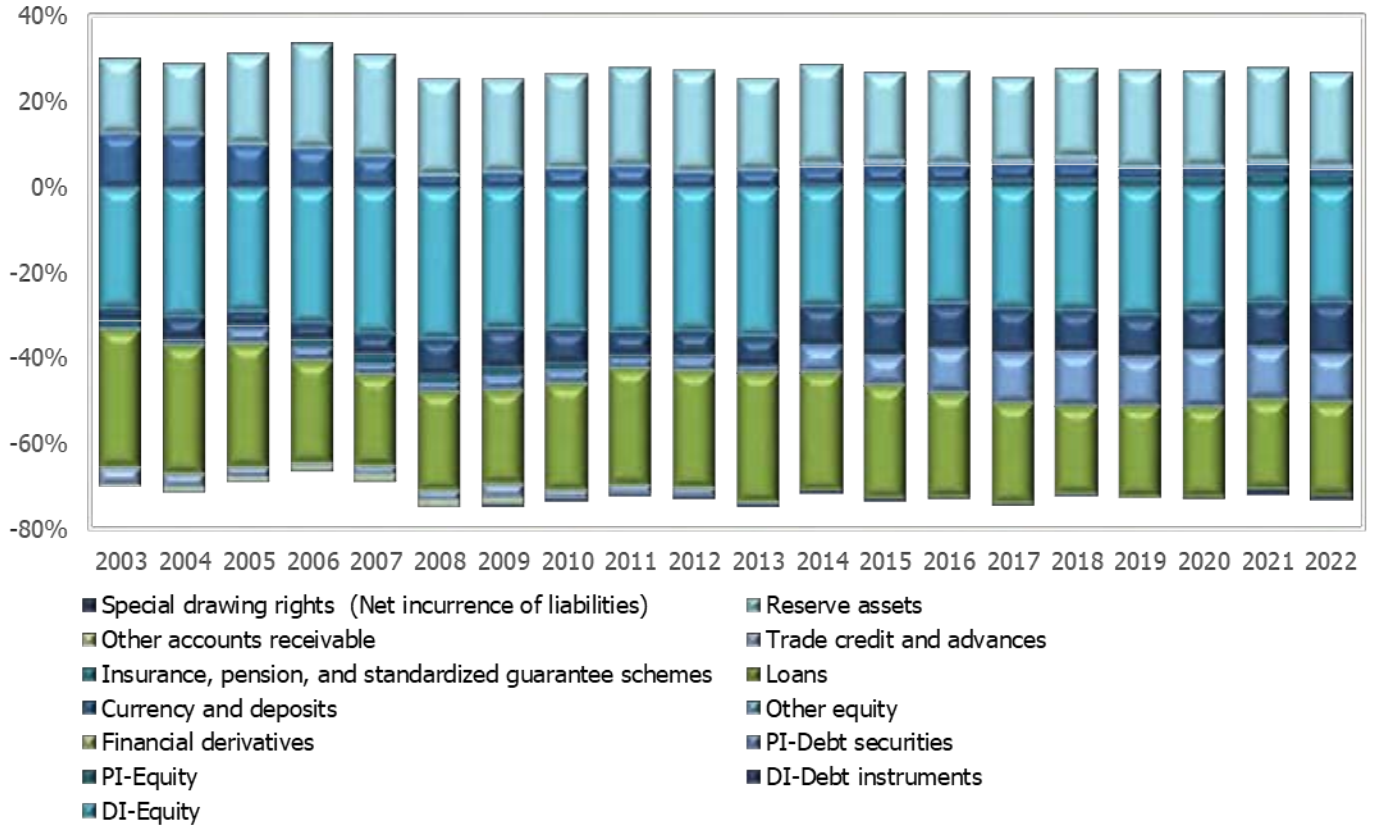
What works for us?

Granular data sources used:

- Loan - by - Loan;
- International transaction reporting system (ITRS);
- Security - by - Security;
- Integrated Reporting System for monetary data – Ongoing Project ISIDORA.

Loan - by - Loan

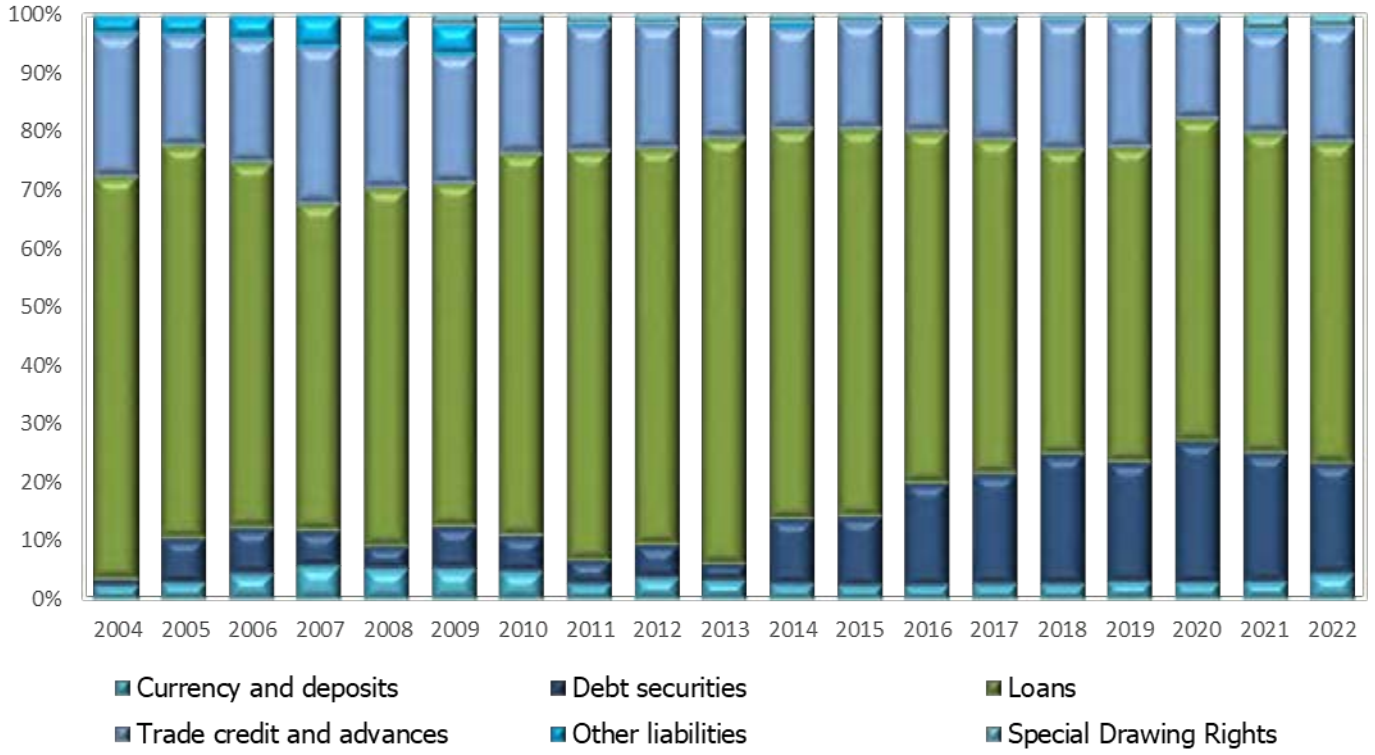
Graph 1: Structure of International Investment Position of the RNM, by instruments



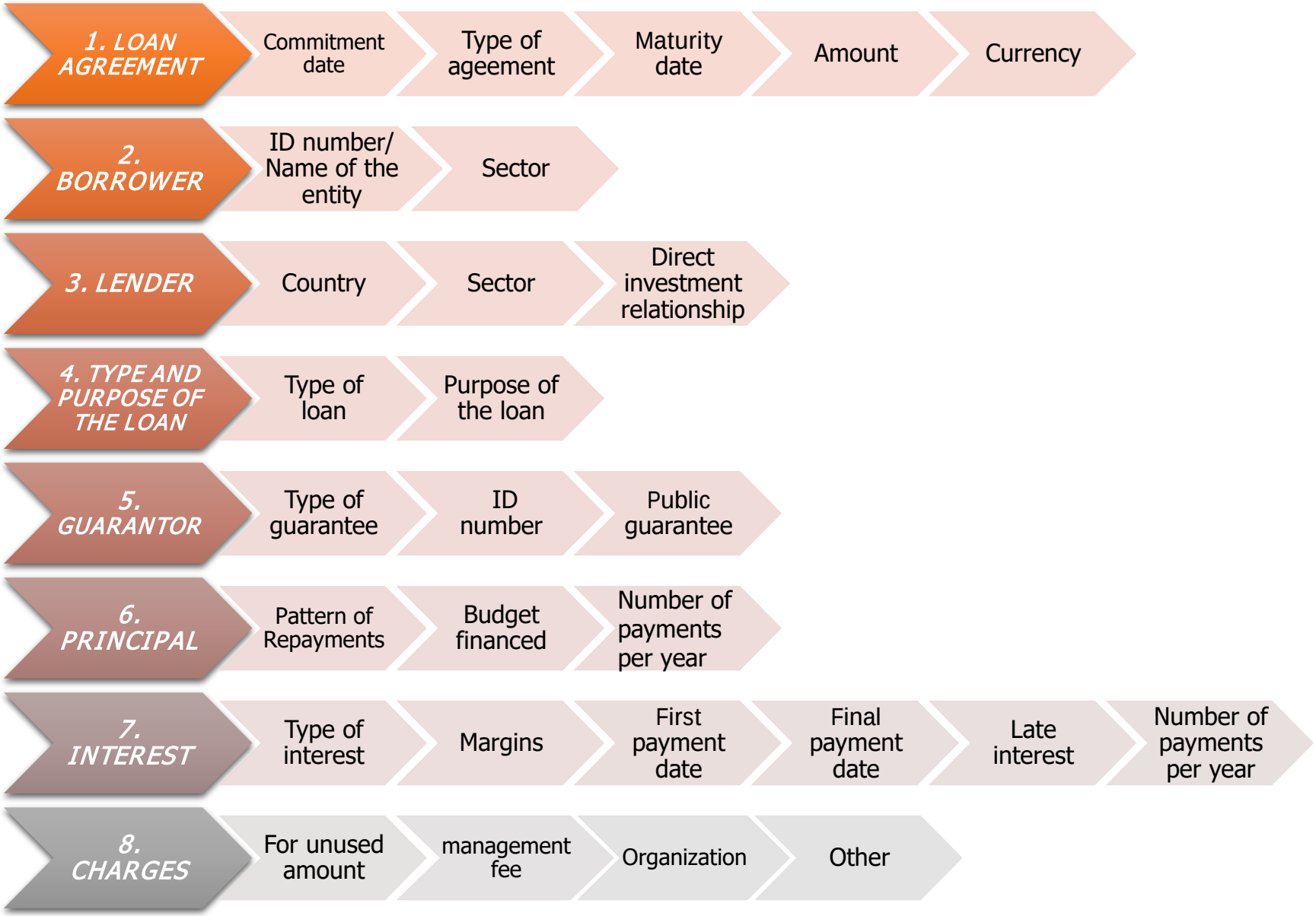
Source: NBRNM

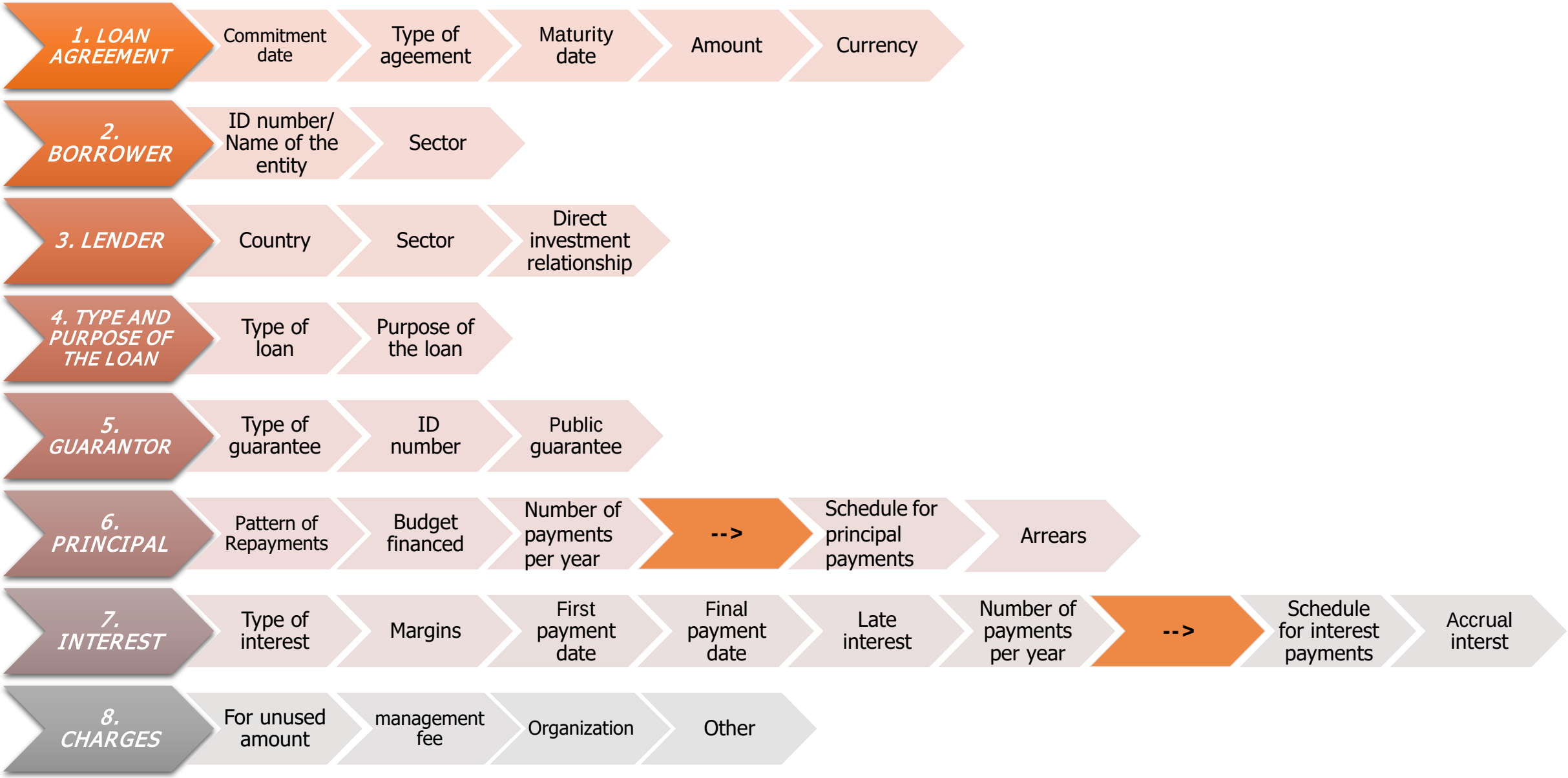
Loan - by - Loan

Graph 2: Structure of Gross External Debt, by instruments



Source: NBRNM


1. LOAN AGREEMENT
Commitment date
Type of ageement
Maturity date
Amount
Currency
2. BORROWER
ID number/ Name of the entity
Sector
3. LENDER
Country
Sector
Direct investment relationship
4. TYPE AND PURPOSE OF THE LOAN
Type of loan
Purpose of the loan
5. GUARANTOR
Type of guarantee
ID number
Public guarantee
6. PRINCIPAL
Pattern of Repayments
Budget financed
Number of payments per year
7. INTEREST
Type of interest
Margins
First payment date
Final payment date
Late interest
Number of payments per year
8. CHARGES
For unused amount
management fee
Organization
Other

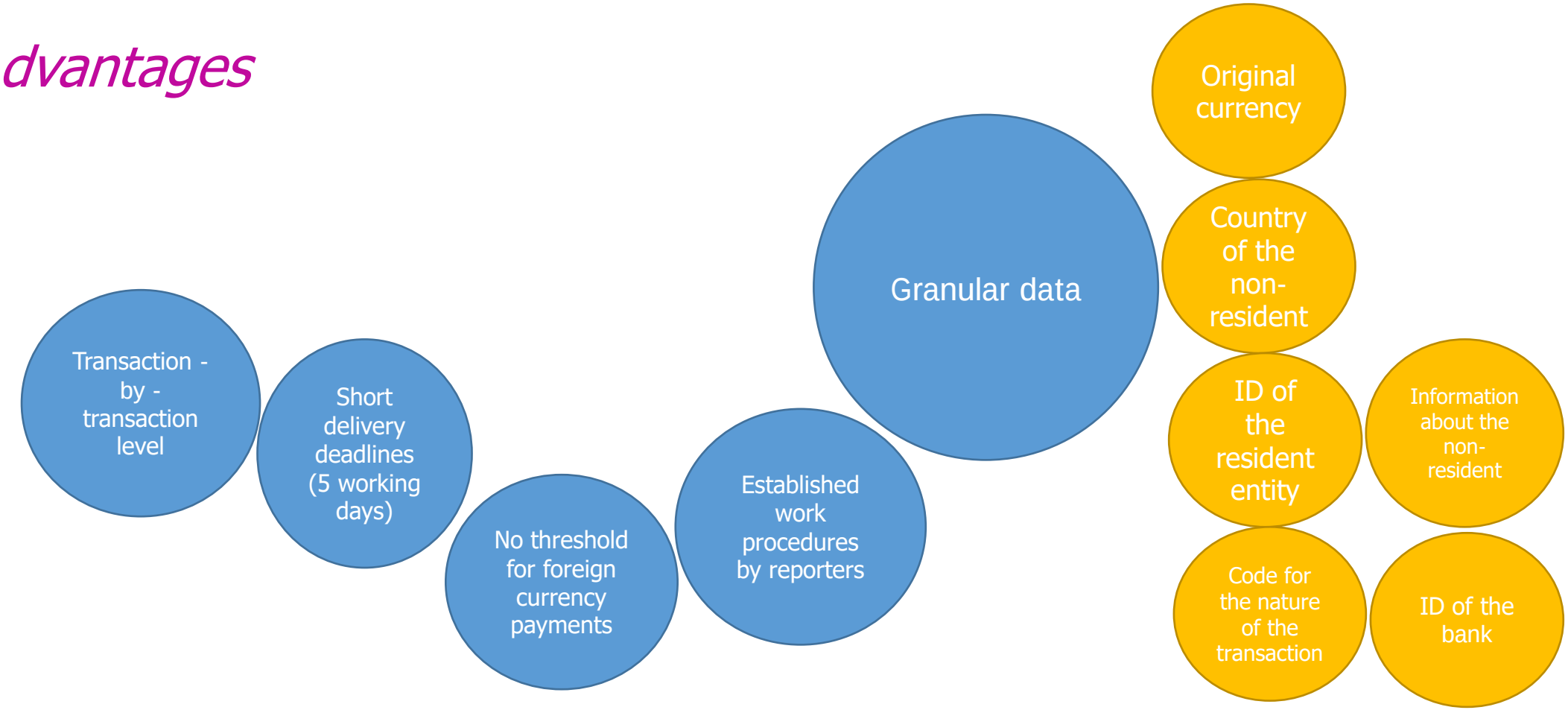

1. LOAN AGREEMENT
Commitment date
Type of ageement
Maturity date
Amount
Currency
2. BORROWER
ID number/ Name of the entity
Sector
3. LENDER
Country
Sector
Direct investment relationship
4. TYPE AND PURPOSE OF THE LOAN
Type of loan
Purpose of the loan
5. GUARANTOR
Type of guarantee
ID number
Public guarantee
6. PRINCIPAL
Pattern of Repayments
Budget financed
Number of payments per year
-->
Schedule for principal payments
Arrears
7. INTEREST
Type of interest
Margins
First payment date
Final payment date
Late interest
Number of payments per year
-->
Schedule for interest payments
Accrual interst
8. CHARGES
For unused amount
management fee
Organization
Other

International Transaction Reporting System (ITRS)

- "Historically inherited system";
- "Closed system" with full integration between stocks at the beginning of the period, transactions and stocks at the end of the period;
- Received mostly on a daily basis;
- Separate codes for each type of transaction (for BoP purposes);
- Electronically transmitted by the deposit-taking corporations.

International Transaction Reporting System (ITRS)

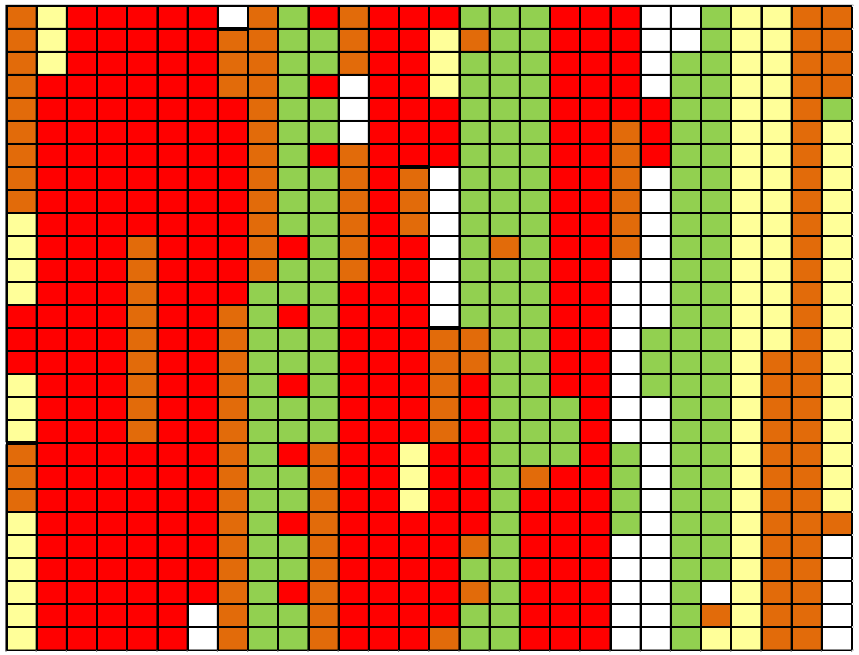
Advantages



International Transaction Reporting System (ITRS)

Picture 2: Data sources - ITRS usage in the Balance of Payments ("The Rubik's cube")

*Each square present an item from the BoP standard presentation

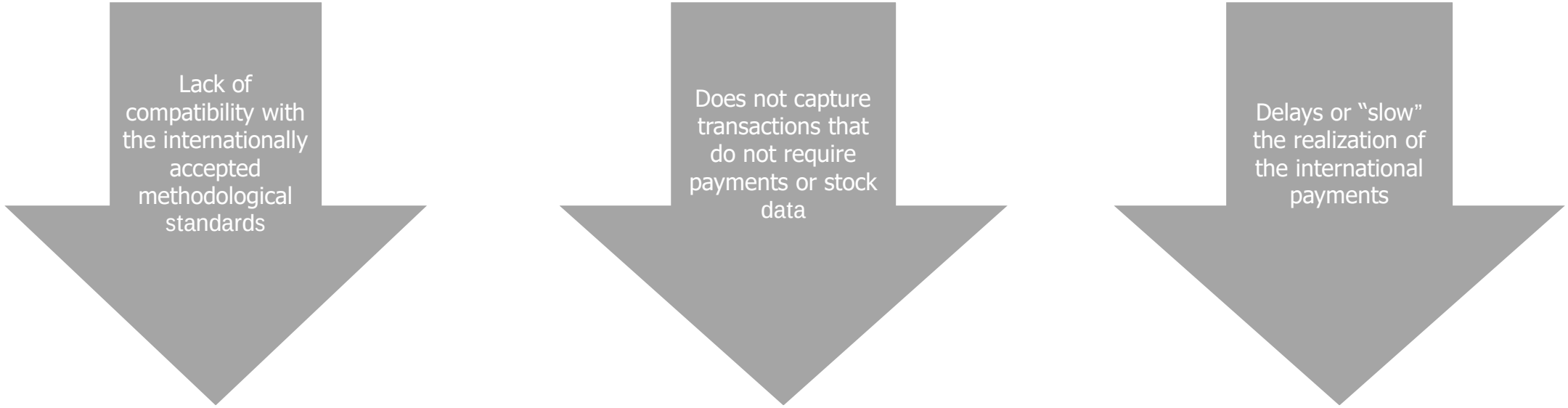


ITRS | ITRS used for cross-checking | Other data source | Combination | No data source established

Source: Author`s presentation based on NBRNM data.

International Transaction Reporting System (ITRS)

Disadvantages



Security - by - Security

- Newly introduced in 2014;
- Portfolio investments aboard – assets;
- ISIN – by – ISIN;
- Monthly reporting (t+20 days), made electronically through web application;
- Direct or indirect reporting made by supervisory agencies of insurance companies, investment funds and pension funds;
- Full reconciliation of data;
- Ensures compilation of Coordinated Portfolio Investment Survey (➡ SDDS Plus adherent, 2019).

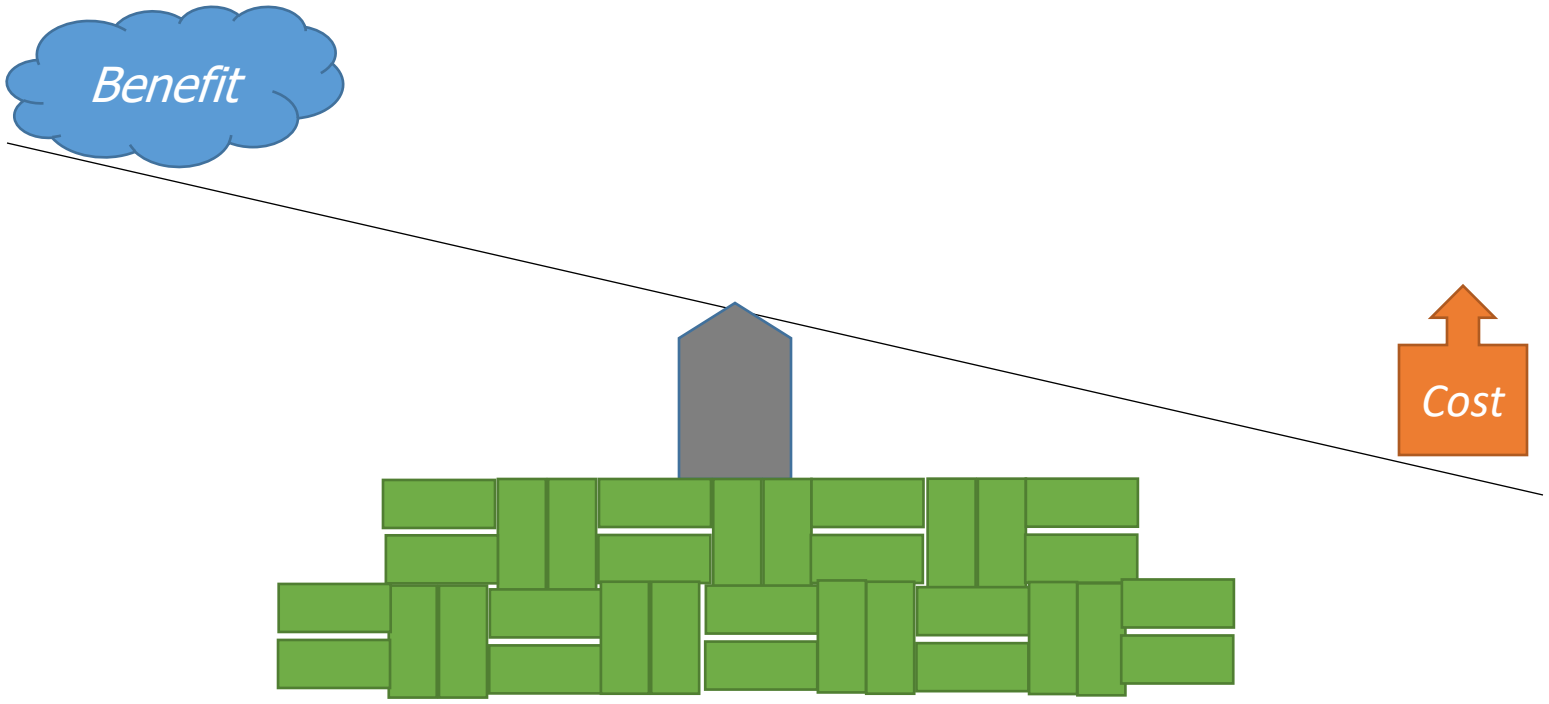
Security - by - Security

- Data are fully reconciled. Forms include data for:
  - Investor (ID, Name, Institutional Sector)
  - Security (ISIN, Currency of issue, Currency of trade, Market and Country of trading,..)
  - Issuer of the security (Name, Country, Institutional Sector,..)
  - Stock at the beginning of the period (Number/Market price per security at the beginning of the period)
  - Transactions (Number/Market price per security of purchased/sold securities)
  - Other changes (Price changes, Exchange rate changes, Other changes)
  - Stock at the end of the period (Number/Market price per security at the end of the period).

A glimpse into the future - ISIDORA

- Ongoing project;
- Collecting monetary data (balance sheet data from deposit-taking corporations, including the Central Bank, and saving houses) for statistical and supervisory purposes (e.g. Credit registry);
- Complete reorganization of the current data reporting system (from accounts to instruments described through predefined dimensions and attributes);
- Detail data structure (currency, country for non-residents, sector classification of non-residents, sub-sectoral classification of residents, flows data, separate the influence of price and exchange rate changes, as well as other changes in volume).

Cost and benefits of data granularity



Cost and benefits of data granularity

Benefits/costs

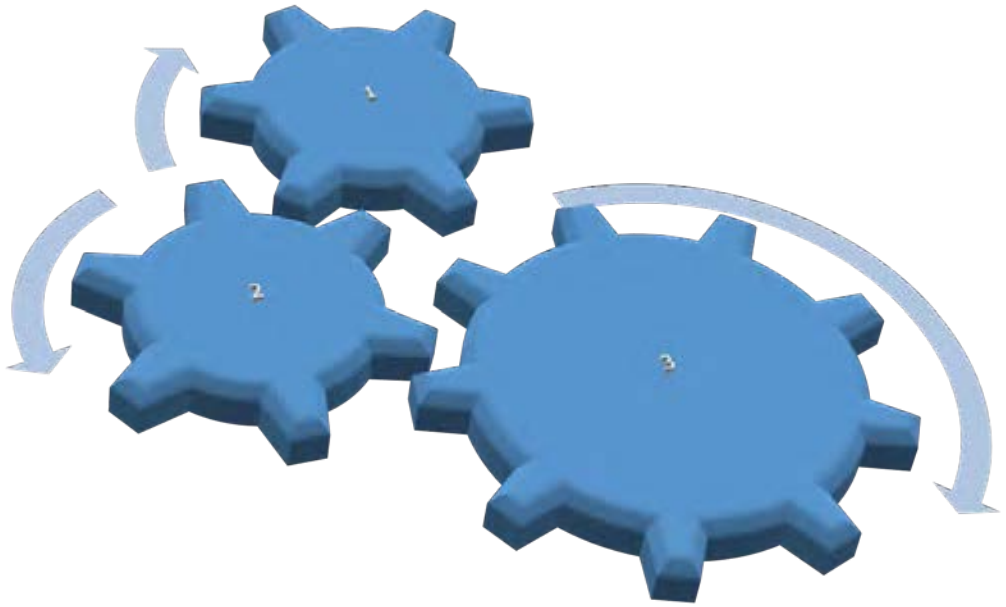
- Fulfilling the statistical methodological requirements;
- Easy to revise the time-series;
- Improved data quality by crosschecking;
- Better data analyses of the effects on the economy on a micro level;
- Decreasing the asymmetries between countries;
- More detailed data available for data users;
- Sometimes, easier transfer of the database from the reporting unit to the data collector;
- High frequency reporting.

Cost and benefits of data granularity

Stating the obvious

- Make sure not to collect the same information twice;
- Have a clear vision of the data that are useable or will become;
- Harmonization and standardization of data;
- Provide possibility for electronic transfer of data - have a developed IT resources;
- Ensure good communication and collaboration with the reporters.

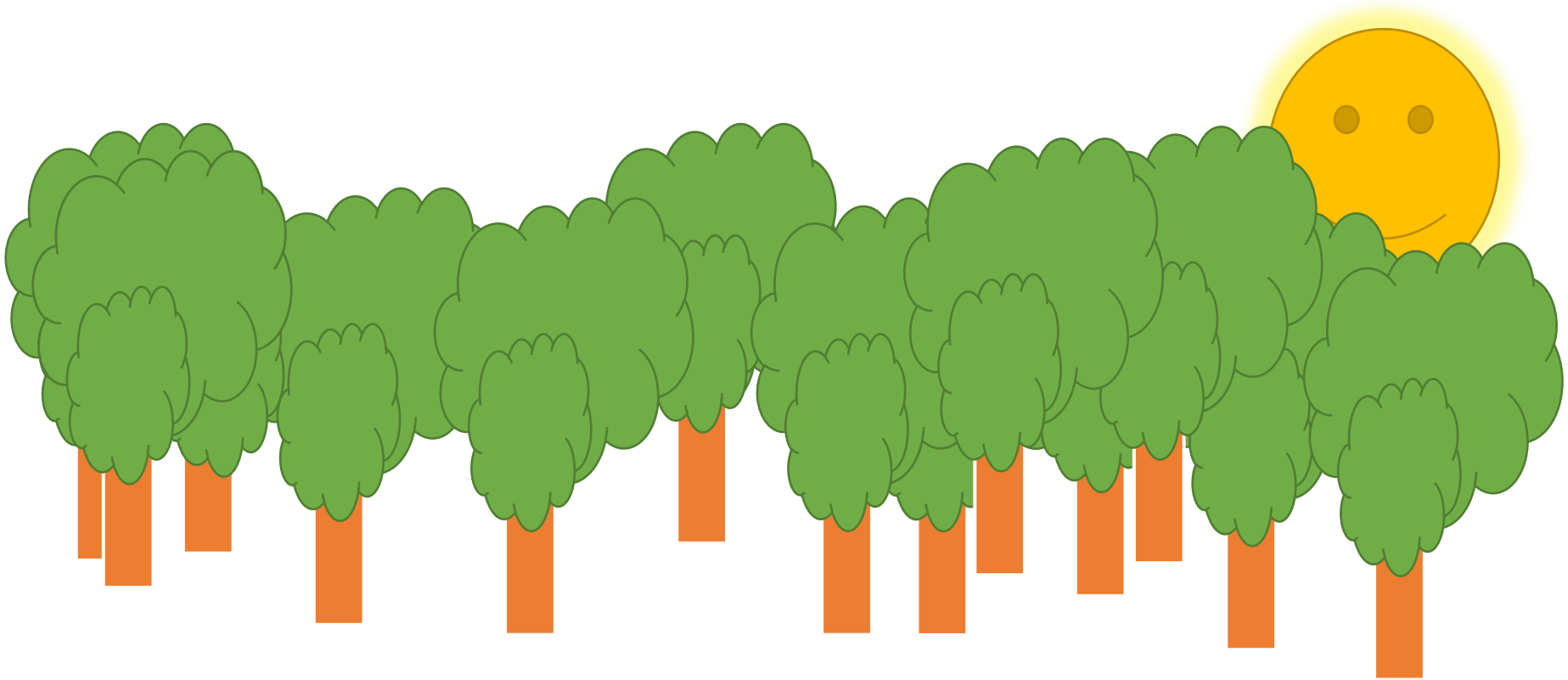
Data Management



Data Management

- The National Statistical System is defined in the Law of state statistics and in the five-year Statistical Programme;
- The regulatory framework consist of Law on State Statistics and the Law of the National Bank, as well as agreements for cooperation and data-sharing between institutions;
- Under the law, confidential data are data that allow direct or indirect identification of the statistical unit, and should be protected;
- Internal procedures and measures are in place to ensure data protection;
- Special procedures regulate the derogations in which confidential data can be disclosed.

Conclusions



Conclusions

- Granular, high frequent and quality data are essential;
- Introduce appropriate data sources;
- Adjust the data sources to the country;
- Apply international harmonization of the data structure, less to the data sources;
- Re-think and re-invent the “historically inherited” data sources;
- Consider the data reporters needs;
- Forward looking approach, modernization, by keeping the old charm.

Thank you

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

On the need for granular, forward-looking physical risk data to support policy analysis[1]

Jens Mehrhoff,
International Monetary Fund

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

STATISTICS

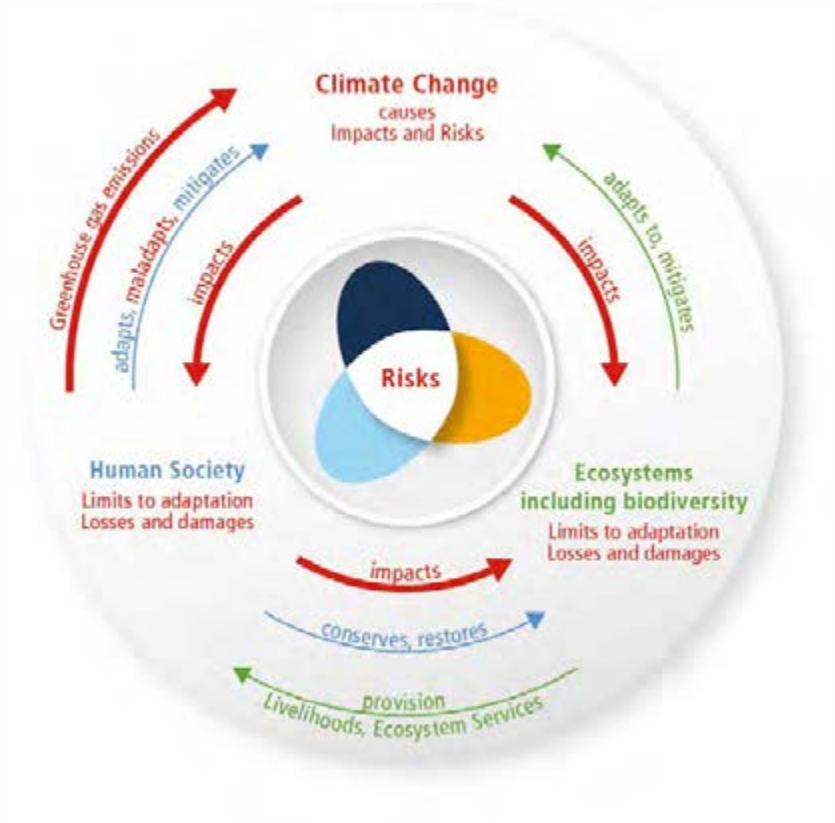
On the need for granular, forward-looking physical risk data to support policy analysis

JULY 15, 2023

Jens Mehrhoff
Real Sector Division

The views expressed herein are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

The definition of risk



Source: https://www.ipcc.ch/

- According to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), ***risk* is the potential for adverse consequences**.
- In the context of **climate change *impacts***, risks result from dynamic interactions between **climate-related hazards with the exposure and vulnerability to the hazards** (~ "physical risk").
  - **Hazard:** The potential occurrence of a physical event or trend.
  - **Exposure:** The presence in places and settings that could be adversely affected.
  - **Vulnerability:** The propensity or predisposition to be adversely affected.

The definition of risk (cont'd)

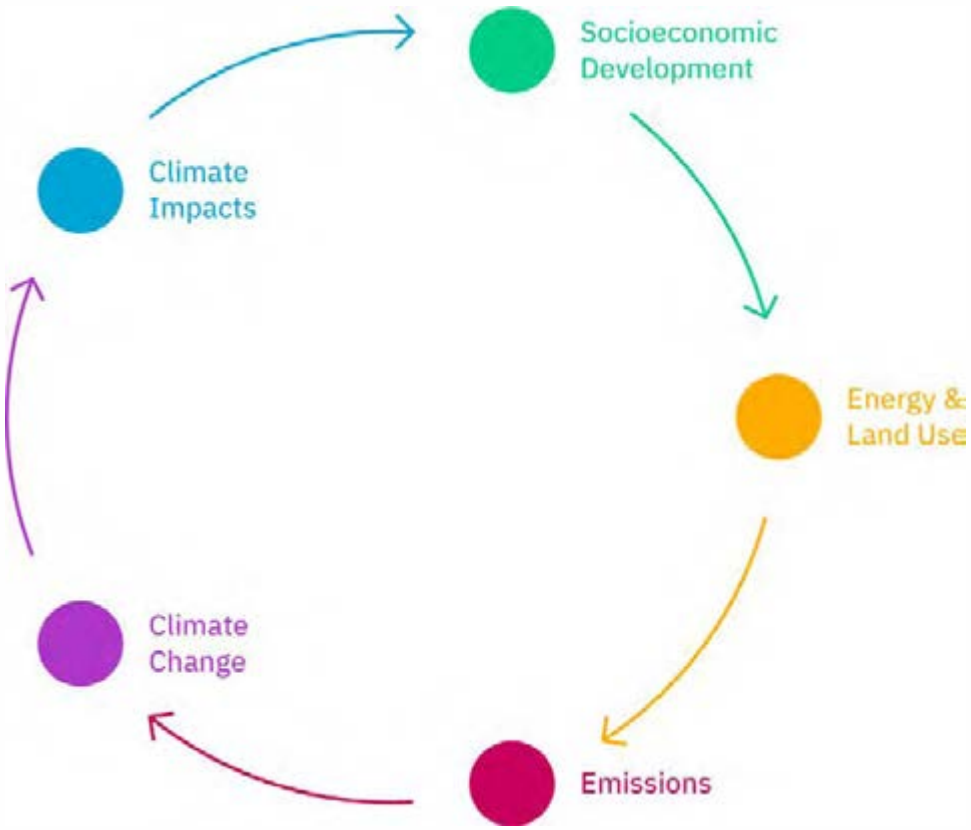
- The revised **definition of risk clarifies and refines** previous definitions.
  - The "core" definition of *risk* is **"the *potential* for adverse consequences" – not to describe physical hazards** (not "the risk *of* …").
  - Risk should be used where **potential adverse consequences and the uncertainty** relating to those consequences are explicitly considered.
  - And to improve the ability for decision-makers to **understand and manage risk**.

- Example: The **risk from flooding** to human and ecological systems is caused by
  - the flood **hazard** – the frequency and/or magnitude of flood events,
  - the **exposure** of the system affected – e.g., topography, or infrastructure in the area potentially affected by flooding –, and
  - the **vulnerability** of the system – e.g., design and maintenance of infrastructure, existence of early warning systems.

Physical risk

- In much of the business and financial literature, the term “physical risk” relates to those derived from the **hazard × exposure × vulnerability framework**.
- ***Physical* risks** affect the economy in two ways.
  - **Acute impacts** from extreme weather events can lead to business disruption and damages to property.
  - There is some evidence that with increased warming they could also lead to persistent longer-term impacts on the economy.
  - These events can increase underwriting risks for insurers, possibly leading to lower insurance coverage in some regions, and impair asset values.
  - **Chronic impacts**, particularly from increased temperatures, sea levels rise, and precipitation, may affect labor, capital, land, and natural capital in specific areas.
  - These changes will require a significant level of investment and adaptation from companies, households and governments

What are climate change scenarios?



Source: https://climatescenarios.org/primer

- Climate change scenarios are not about predicting the future, but come in the form of **projections of what can happen or pathways of how to reach certain goals**.
- **Socioeconomics and climate change are inseparably intertwined.**
  - Socioeconomics define for example our land and energy needs,
  - which are closely linked to emissions.
  - Increased emissions lead to higher GHG concentrations,
  - which leads to climate change and its impacts.
  - Climate impacts close the cycle as they will also influence our socioeconomics.

Extreme events

- With an increase in global mean temperatures, extreme events like droughts, floods, and heatwaves become **more frequent and severe**.
- Under climate change, e.g., the **mean temperature rises**, and the **distribution shifts to higher temperatures**. Hot days that have been rare in the past occur now more frequently and reach even higher temperatures.
- Additionally, the **variance increases**. Formerly average temperature become less frequent, and the distribution spreads out. Hot temperatures become even more frequent and extreme. Cold days still occur but become rarer.
- Thresholds of what constitutes an extreme event can be defined **relative to the preindustrial climate**. This allows examining changes under temperatures that already occur today, as well as a glimpse at future global warming levels.
- **Uncertainties** across climate and impact models originate from various stages in the modelling chain. Still, the overall picture remains clear: the world is **(already) increasingly exposed to extreme events**.

Climate scenario analysis

- The distinct nature of climate risks poses a challenge to **standard risk assessment approaches**.
  - Climate risks have **long time horizons with high uncertainty** about how policy and socioeconomic factors might evolve;
  - they are **global and economy-wide** in nature; and
  - they are complex, **varying from region to region and sector to sector**.
- These **distinct characteristics are not captured** by risk assessment approaches that rely on top-down modeling and historical trends, are narrowly focused, and assume the structure of the economy and financial system remain unchanged.
- **Scenario analysis** is an essential tool to overcome these challenges. It provides a flexible **"what-if" framework** for exploring how the risks may evolve in the future.
- These scenarios can help a wide range of players better understand how climate factors will **drive changes in the economy and financial system**.

Data needs and data sources

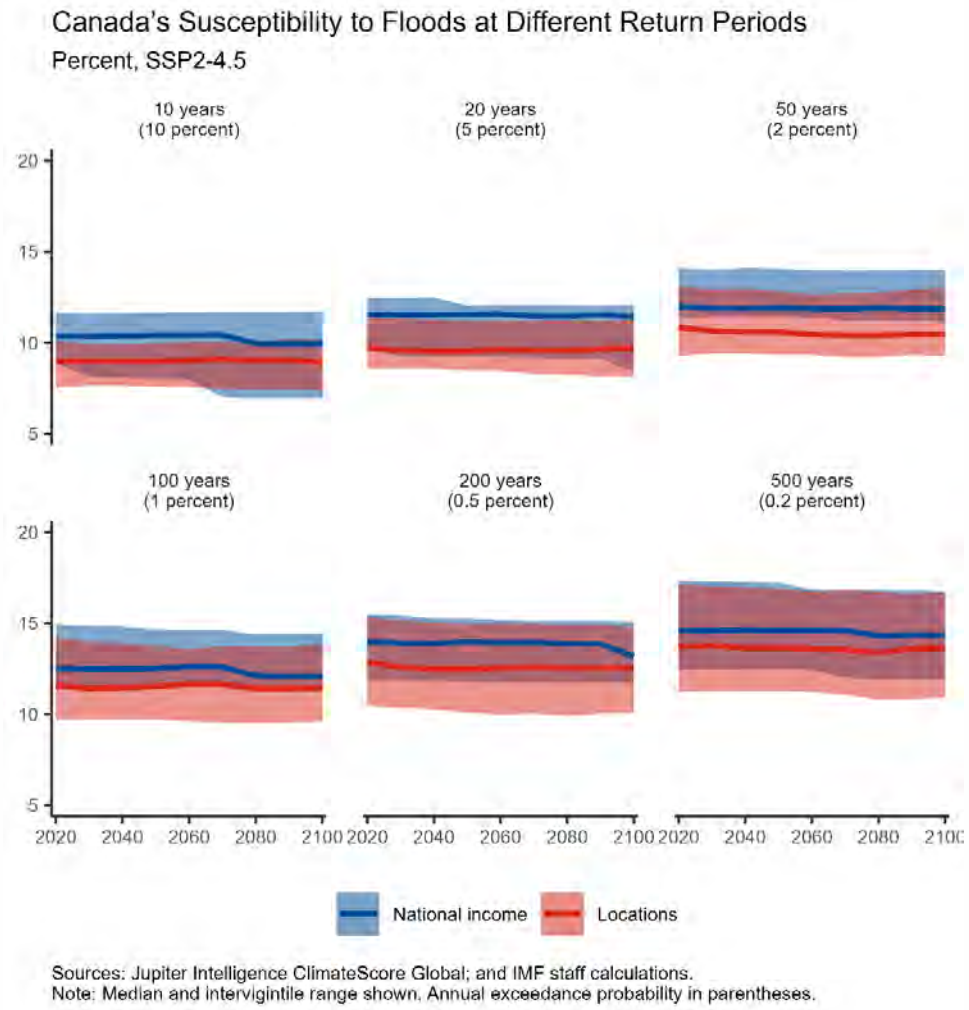
- **"Zero" layer:** Geographic resolution, e.g., topographical map with elevation.
- **Hazard layer:** Geospatial data for different climate-related hazards, e.g., temperature, precipitation, or drought; different metrics for each hazard (e.g., return periods) and uncertainty measures (e.g., confidence intervals).
- **Exposure layer:** Gridded data on population (urban and rural), GDP/capital stock, and infrastructure (real properties and agricultural land).
  - **Grids often coarse**, e.g., 5 arcminutes equal 9.277 km at the equator; global average area of a grid cell is about 55 km^2 (21 mi^2).
- Also, **geospatial data** on banks' loans and investments, firms' and households' balance sheets, and insurance exposure and (un)insurability.
- **Vulnerability layer:** Hazard-specific damage functions, linking the intensity of individual hazards to economic damages.

➢ Need for **public, global, geospatial, forward-looking data and tools**.

Data needs and data sources (cont'd)

- **"Zero" layer:** Google Maps API and GADM shapefile.
- **Hazard layer:** Jupiter Intelligence (vendor data).
  - **Geographical coverage:** 100,000 geospatial locations, covering 97 percent of global GDP (at 15 arcmin resolution) – **example: Canada, Ottawa**.
  - **Hazards:** inland and coastal flooding, high winds, extreme heat, drought, hail, wildfires, precipitation, and extreme cold – **example: flood depth at 200-year return period** (0.5 percent annual exceedance probability).
  - **Emissions scenarios:** "low" (SSP1-2.6), "intermediate" (SSP2-4.5), and "very high" (SSP5-8.5) – **example: SSP2-4.5**.
  - **Years:** 2020 to 2100 in five-year increments, and a historical baseline (1986–2005) – **example: 2050**.
- **Exposure layer:** Socioeconomics compatible with the Shared Socioeconomic Pathways (at 5 arcmin resolution) – **example: gridded GDP** from Murakami et al.
- **Vulnerability layer:** None (yet). ☹

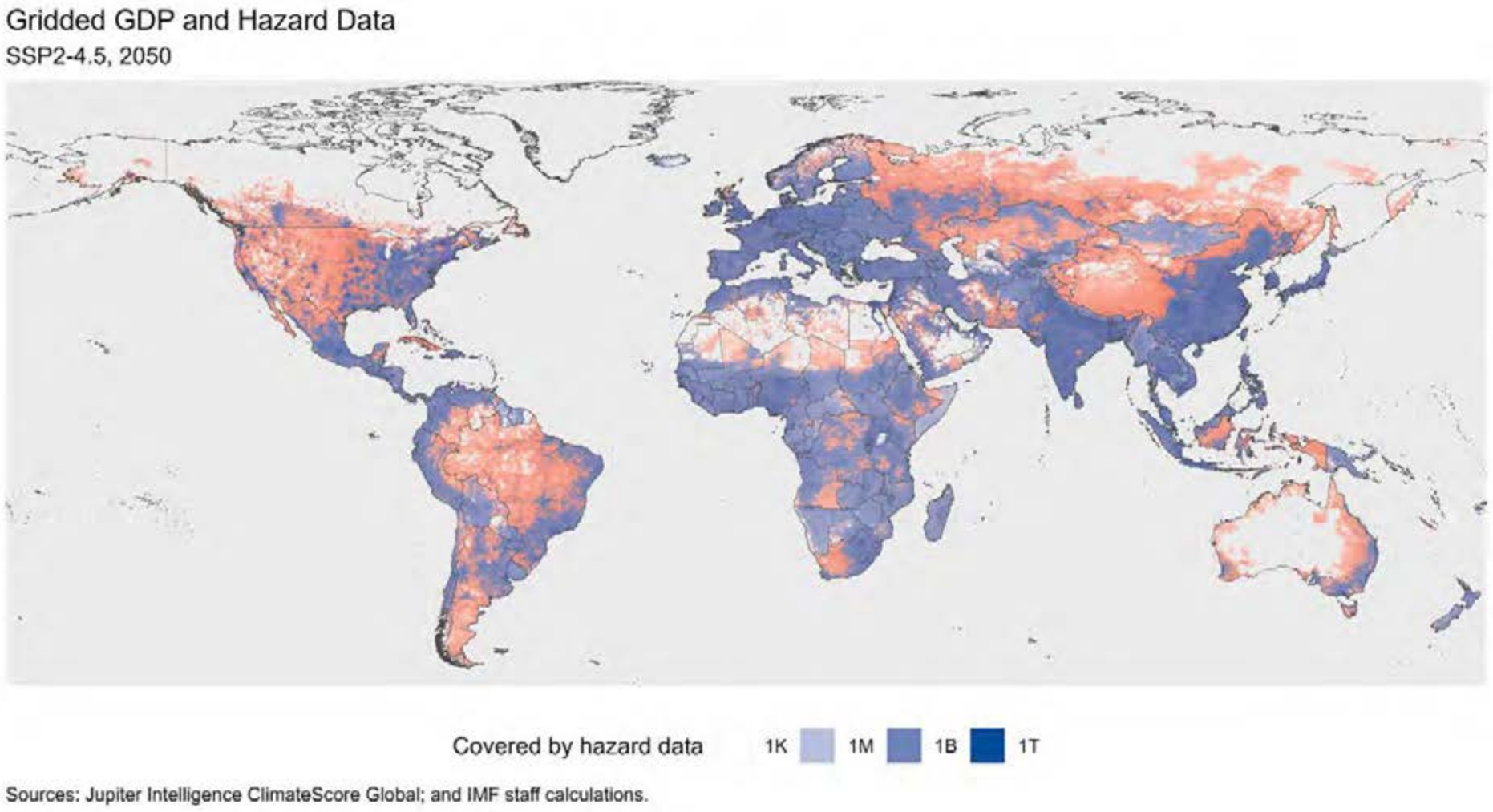
Digression on macro indicators



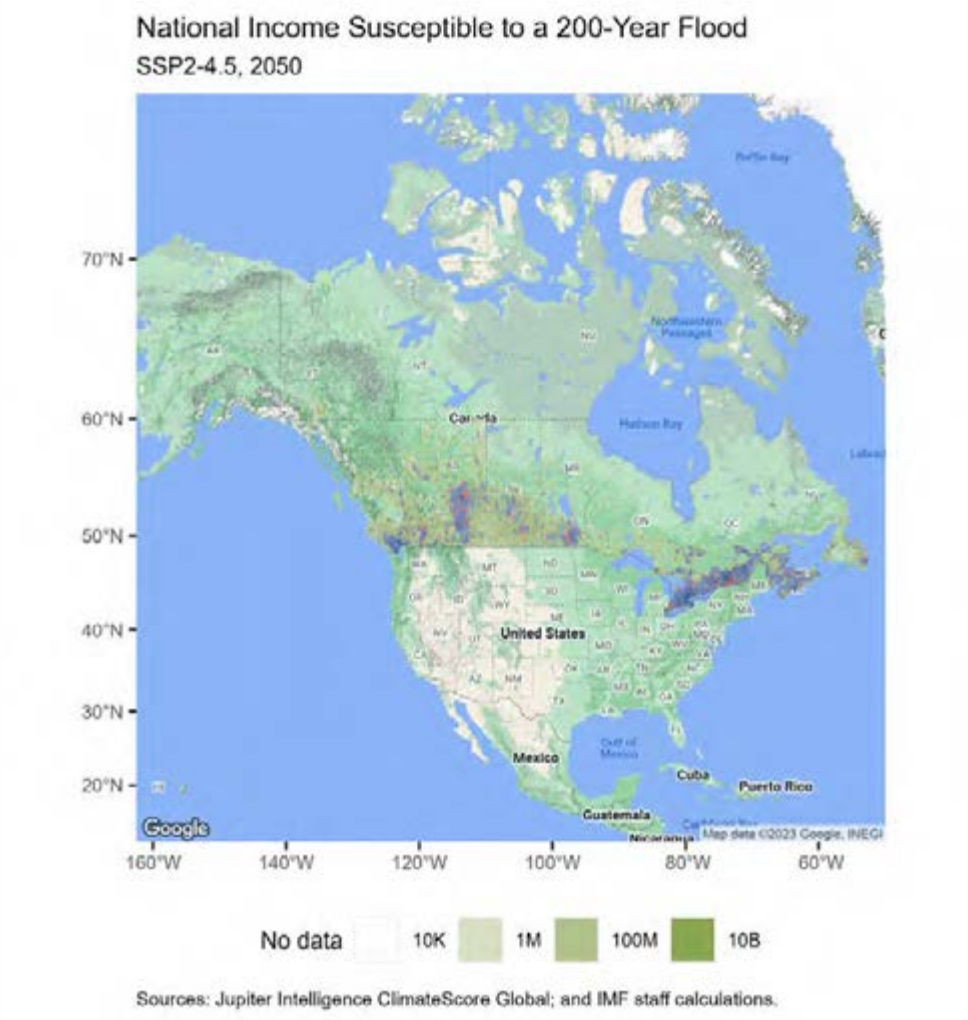
- Obviously, the **potential impact of a rarer event would be larger**, but that is more of a level than a climate change effect.
- **Susceptibility** – not necessarily equal to risk(!) – **increases in all scenarios** but more so for higher emissions scenarios.
- While **Canada's GDP exposure tends to fall over time**, this is not the case for the (fixed) locations.
- This is hence due to **GDP moving away from more hazardous areas** – this is not possible in all regions or countries though.

➢ More details in IFC-sponsored **IPS 205 – Sustainable finance statistics**, Wednesday 2 pm, room 213.

GDP exposure and hazard(s) data

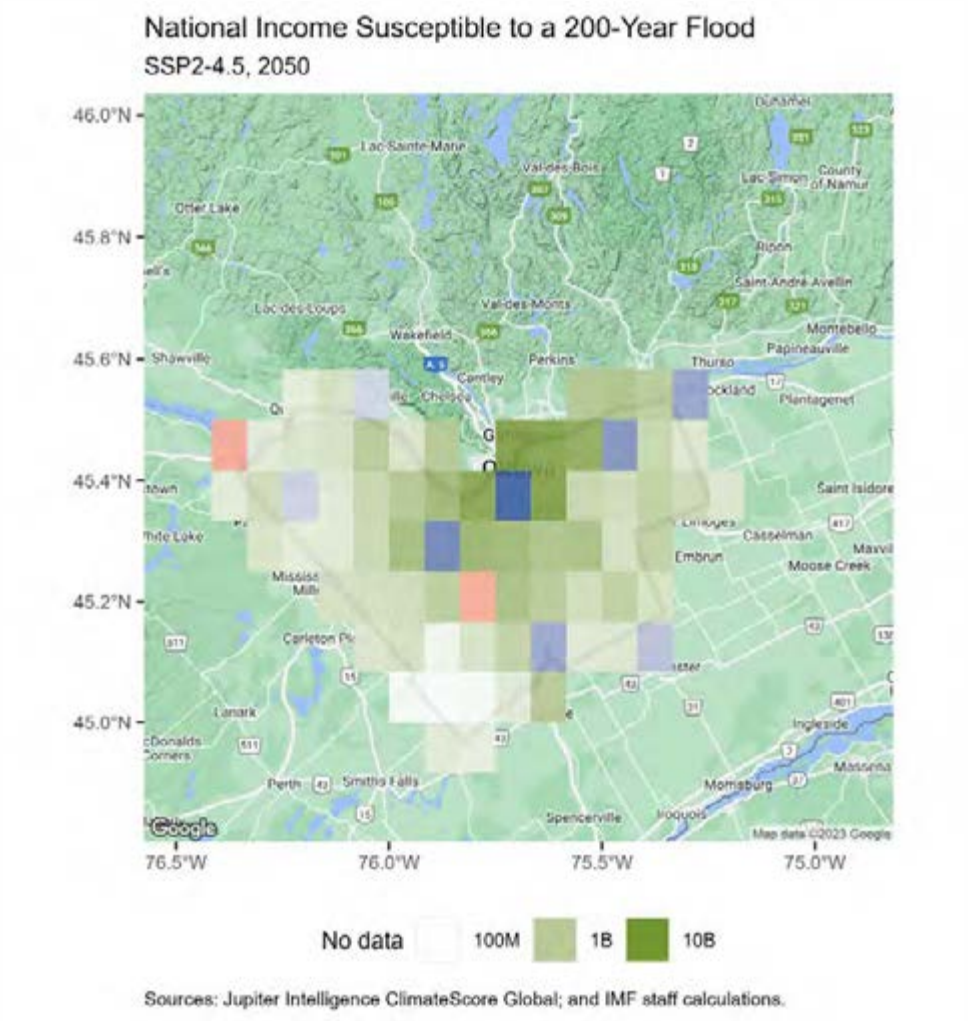


Flood hazard and GDP exposure



- A **grid cell** can be
  - **not susceptible** to floods at a 200-year return period in 2050 under SSP2-4.5;
  - **susceptible** to floods; or
  - **not covered** by the hazard data.
- The **hazard data are geospatial observations (places on a map)**, and are assumed representative for the grid cell they are contained in.
- The **GDP coverage for Canada** (at 15 arcmin resolution) **is 94 percent**.
- The **aggregate susceptibility**, as shown on the slide before, is **14 percent for GDP** in 2050 under SSP2 and **12 percent for the (fixed) locations**.

Flood hazard and GDP exposure (cont'd)



- The variation in the distribution of hazards **within a region or country** matters.
- However, the coverage of the 100,000 locations the IMF secured access to is **somewhat limited at a granular resolution** (not only for cities, also for small islands states, for example).
- For Ottawa, **only 1 percent of GDP** would be susceptible to floods.
- But the point in the hazard data set closest to the Bank of Canada is about 1.5 km south-southeast – and thus also **not in the same proximity to the Ottawa River**.
- Yet, the **grid cell represented by this one observation** accounts for almost 12 percent of the city's GDP.

Discussion

- Need for **public, global, geospatial, forward-looking physical risk data** to support policy analysis.
  - Requires **bias correction and statistical downscaling** – see e.g., World Bank Climate Change Knowledge Portal: https://climateknowledgeportal.worldbank.org/
- Need for **interactive tools** to make geospatial data (more) accessible.
  - **Dynamic data visualization**, e.g., zoom in/out, and add/remove layers; and
  - **Aggregation across layers**, e.g., filter by area of interest, and combine hazards and exposures.
- There is a wide range of **existing data sets and resources** (including open-source data sets, tools, and models, along with methods to improve data granularity), and it is important to identify which ones can be used to best support the climate risk assessment.
- A **country-specific approach** that is tailored to the needs and objectives identified by local stakeholders should be adopted, **factoring in the country's specificities and (data) constraints**.

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Payment habits during Covid-19: Evidence from high-frequency transaction data[1]

Tatjana Dahlhaus and Angelika Welte,
Bank of Canada

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel — 2021-43

Last updated: September 9, 2021



Payment Habits During COVID-19: Evidence from High-Frequency Transaction Data

by Tatjana Dahlhaus[1] and Angelika Welte[2]

[1] Canadian Economic Analysis Department
[2] Currency Department
Bank of Canada, Ottawa, Ontario, Canada K1A 0G9

tdahlhaus@bankofcanada.ca, awelte@bankofcanada.ca

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

ISSN 1701-9397

©2021 Bank of Canada

Acknowledgements

The views expressed in this paper are those of the authors and do not represent the views of the Bank of Canada. We thank Eric Chue, Walter Engert, Allan Gregory, Kim Huynh, Francisco Rivadeneyra, Leo Sabetti, Rodrigo Sekkel and seminar participants at the Bank of Canada for useful comments and suggestions.

Abstract

We investigate how the COVID-19 pandemic has changed consumers' payments habits in Canada. We rely on high-frequency data on cash withdrawals and debit card transactions from Interac Corp. and Canada's Automated Clearing Settlement System. We construct daily measures of payment habits reflecting cash usage, average transaction values, and the share of transactions in which the customer or card holder and the acquiring machine (ATM or POS) are of the same bank. Using simple dummy regressions and local projection models, we assess how these indicators of payment habits have changed with the evolution of the COVID-19 pandemic. We find evidence that during the pandemic consumers adjusted their behaviour by avoiding frequent trips for cash withdrawals and point-of-sale purchases and making fewer transactions for higher amounts. They also made smaller-value cash withdrawals compared with the value of card payments, which could reflect a reduced use of cash for point-of-sale transactions. Consumers also made relatively more withdrawals from ATMs that are linked to their financial institution (on-us transactions). Finally, we highlight that estimates of economic activity based on card data alone could be biased if shifts in payment habits are not taken into account. We estimate that debit card payments might have overstated consumer expenditure growth by up to 7 percentage points over the course of the pandemic.

Topics: Coronavirus disease (COVID-19); Payment clearing and settlement systems; Recent economic and financial developments; Domestic demand and components
JEL codes: C22, C55, D12, E21, E42, E52

1 Introduction

The COVID-19 pandemic and the measures to contain it have shifted how much consumers spend, what they spend on and which payment methods, in particular cash, they use when they spend. Understanding these changes is essential to the core mandates of many central banks as they need to assess the state of the economy and issue currency in the form of banknotes or coins. The rapid developments during COVID-19 have highlighted that traditional indicators often lack immediacy and convey an incomplete picture. Since the start of the COVID-19 pandemic, the Bank of Canada and other central banks have thus accelerated the use of alternative data sources. For example, Huynh et al. (2020) obtained alternative weights for the consumer price index from high-frequency expenditure network data on card payments.

This paper uses high-frequency transaction data to assess how consumers' payment habits have adjusted to the COVID-19 pandemic. Understanding changes in payment behaviour is not only important for monitoring cash demand but also allows us to shed light on the reliability of this transaction data in assessing the state of the economy. A shift in payment habits may affect the dynamics of consumer spending as measured by card payments. For example, shifts from cash towards debit cards could imply that actual spending could be overestimated by card payments. Quantifying how cash use has evolved relative to electronic payment methods during the pandemic is thus key in understanding the link between high-frequency transaction data and consumption.

In particular, we use a comprehensive dataset based on data for ATM withdrawals and point-of-sale (POS) debit card transactions from two payment infrastructure providers in Canada, Interac Corp. and the Automated Clearing Settlement System (ACSS). We construct daily measures of payment habits, such as the share of cash transactions and average transactions values. The transaction data from Interac and ACSS also allows us to obtain a measure of so-called "on-us" transactions—those transactions in which the customer or card holder and the acquiring machine like ATM or POS are of the same bank. We discuss how these measures have evolved throughout the pandemic and support

these observations with simple dummy regressions. Further, we assess the link between the COVID-19 pandemic and payment habits more formally using local projection models at daily frequency. We obtain impulse responses for each measure of payment habits to changes in indicators of the pandemic such as new confirmed cases and the stringency of government-imposed response measures. Finally, we provide insights on how a shift from cash towards debit cards affects the dynamics of card transactions, and we calculate the potential bias if debit card payments were used as a proxy for consumption expenditure.

Our work contributes to several strands of ongoing research: First, recent studies have illustrated the usefulness of electronic payments data in monitoring the state of the economy for a number of countries, e.g., Galbraith and Tkacz (2013), Galbraith and Tkacz (2018), Chapman and Desai (2021) for Canada, Aastveit et al. (2020) for Norway, Cavallo (2020) for the US, and Aprigliano et al. (2019) for Italy. These papers mostly assess whether high-frequency card transaction data can improve forecasts of economic variables such as GDP and consumption. However, they abstract from the role of cash and how shifts in payment behaviours may influence the reliability of this data for monitoring the economy.

Second, our paper contributes to the literature analyzing the response of cash following shocks that directly or indirectly affect its use. This literature has documented how payment habits change following government-imposed measures directly restricting cash use. One prominent example is the Indian demonetization in 2016, when the Indian government removed the legal tender status of 86% of the Indian currency in a surprise announcement and introduced new banknotes with legal tender status. The stated goals of this policy were to curb the underground economy and remove counterfeit banknotes from circulation. Lahiri (2020) considers the aggregate change in indicators of payment habits such as currency in circulation and the use of digital payments. He concludes that currency in circulation recovered within months and the adoption of digital payments was lower than expected. Further, Chodorow-Reich et al. (2019), using geographical variation in the severity of demonetization, show that a sharp, temporary decline in currency reduced ATM withdrawals and economic activity while encouraging adoption of alternative payment technologies.

Another case is the Greek government's imposition of daily ATM withdrawal limits

during the Greek debit crisis in 2015, which lasted until 2019. At the same time, measures were put in place to increase electronic payments and discourage cash. While electronic payment methods have substantially increased at the POS (European Payments Council, 2019), Greece is still among the most cash-intensive countries in the EU (Esselink and Gijsel, 2017).

The COVID-19 pandemic is different from the former cases in the sense that there has been no imposition of direct restrictions on cash behaviour. Chen et al. (2020) find that cash use fell in the early stages of the pandemic when lockdown measures were in place and increased during the reopening in the summer of 2020, indicating that consumers may have adjusted their payment behaviour without direct restrictions on cash use.

Our analysis shows that with the COVID-19 pandemic, cash withdrawals, especially withdrawals from ATMs not belonging to the cardholder's bank, have dropped in comparison to debit card transactions. This decline is more pronounced for the total transaction value, although the average dollar value of both cash withdrawals and card transactions has increased. At the same time, a larger share of POS debit card transactions now involves a consumer and a merchant who do not have their accounts with the same financial institution (FI).

We find evidence that consumer behaviour adjusted to the pandemic by avoiding frequent trips for cash withdrawals and POS purchases and making fewer transactions for higher amounts. The increased amounts for ATM withdrawals (especially during the first lockdown) could also reflect an increase in cash demand related to precautionary motives. Further, consumers withdrew less cash compared to card payments, which may reflect a reduced use of cash for point-of-sale transactions. In addition, consumers made relatively more withdrawals from ATMs that are linked to their FI (on-us transactions), possibly because they planned their withdrawals or because they could not access white-label ATMs that are often located inside of businesses. The larger share of on-us transactions could also be explained by reduced mobility which kept consumers close to their "home bank." These reasons are consistent with a decrease of the on-us share during the reopening of businesses and the lifting of mobility restrictions. Finally, the share of non-on-us debit POS transac-

tions increased during the pandemic. Consumers are probably unaware of whether a card payment is non-on-us or on-us. However, shifts towards e-commerce and larger retailers as well as closures in certain industries might have impacted FIs' issuing and acquiring lines in an asymmetric way, hence leading to changes in their non-on-us shares. Microdata from consumers and businesses or FIs could deliver more insights.

While generally our measures of payment habits have not yet returned to pre-pandemic levels, their behaviour seems to be linked to the evolution of the pandemic. For example, the use of cash increased over the summer of 2020 (when there were relatively low and stable COVID-19 case counts) and decreased again towards the end of the year (when Canada entered the second wave). However, it is not yet evident from our dynamic analysis whether this behaviour is in line with temporary changes or a sign of a persistent shift away from cash and also towards withdrawals at FI-owned ATMs.

Finally, our findings highlight that estimates of economic activity based on card data alone could be biased if shifts in payment behaviour, such as a change in cash use, are not taken into account. For example, growth in consumer spending could be mis-estimated from card transaction data as payment choices seemed to change during the pandemic. We estimate that debit card payments might have overstated expenditure growth by up to 7 pp over the course of the pandemic.

The paper proceeds as follows. In Section 2 we describe the transaction data and the COVID-19 indicators used in our analysis. Section 3 introduces our measures of payment habits and provides first insights on their evolution amid COVID-19. Section 4 shows a more formal analysis based on dummy regressions and local projection models. We also assess by how much consumption expenditures could be biased were they proxied by debit card transactions alone. Section 5 concludes.

2 Data

We start this section by introducing the high-frequency transaction data based on Interac and ACSS ABM withdrawals and debit card transactions. We then present the indicators

that we use to describe the severity and state of the COVID-19 pandemic.

2.1 Transaction Data

Data on cash withdrawals and debit card transactions comes from two payment infrastructure providers, Interac Corp. and the ACSS. We start by describing the schematic presented in Figure 1. To distinguish the roles of these infrastructures, we explain on-us transactions. Electronic payment card transactions such as cash withdrawals at ABMs and debit cards transactions at a POS terminal involve four parties. On the so-called issuing side there are two parties: the cardholder and the financial institution (FI) where the bank account linked to the card is held. On the acquiring side, there are also two parties: the owner of the POS terminal or the ABM and the financial institution of the owner. On-us transactions involve the same financial institution on the acquiring side and the issuing side. For example, if someone withdraws cash at an ABM that belongs to the same bank that issued the card, then the transaction is on-us. If the person uses their card at a merchant who has their account with a different financial institution, then the transaction is not on-us.

Interac Corp. operates the main debit card network in Canada, which consists of the shared Inter-Member Network (IMN) for ABM withdrawals and another shared IMN for debit purchases at the point of sale. All ABM withdrawals and debit purchases, on-us and non-on-us, made with Interac debit cards are processed through the shared IMN.[1] Merchants also need a payment processor to clear and authorize debit purchases. The ACSS further clears the non-on-us transactions that are processed and authorized by the IMN and for which the issuer and acquirer are different members of the shared network.[2]

Besides Interac cash withdrawals and the Interac debit payments at a POS terminal discussed here, Interac Corp. also offers Interac online debit payments and Interac e-transfers. E-transfers can be used to transfer funds using the mobile phone number or email address of the recipient. Likewise, the ACSS also clears other non-on-us e-transfers and other

[1] IMN participants are either direct connectors with a physical connection to the IMN or indirect connectors whose transactions are routed through a direct connector.

[2] Many Interac debit cards in Canada are co-badged as Visa Debit cards. Transactions at a physical POS made with these cards that are entered as Visa Debit are also processed by Interac, and the same distinction between on-us and non-on-us applies.

electronic items such us pre-authorized debit and credit, as well as paper items (cheques).

We will now explain in more detail the streams, or transaction flows, used in our analysis. **Interac**: For cash transactions, the Interac data includes withdrawals at ABMs connected to the IMN. These ABMs could be on the premises of an FI, FI-owned offsite, or white-label ABMs.[3] The POS purchases in the Interac dataset consist of Interac debit payments made at a physical payment terminal, including chip- and pin-authorized transactions and contactless Interac Flash transactions. The data is already adjusted for refunds and cases where an ABM did not contain enough banknotes to cover the requested amount. Data is available for every calendar day, including weekends and holidays.

ACSS: In the ACSS terminology, the transactions considered in our analysis are called point-of-service payments because they most commonly involve using a payment card at a physical service point which is either an ABM or a payment terminal. ABM transactions are split into two streams: the first one is the "pull" that happens when a cardholder requests a certain amount of cash from an ABM. The second one is an adjustment stream that accounts for situations where the ABM does not contain enough banknotes to cover the requested amount. Similarly, two separate streams exist for in-store debit payments, one for payments and one for refunds. Each stream provides the dollar value and the number of transactions. We subtract these refunds and adjustments from purchase quantities and requested ABM quantities, respectively, to obtain the numbers used in this paper .

ACSS data is available on a daily frequency for the days when the ACSS operates. The ACSS operates Monday to Friday except for holidays. Relative to the time when the cardholder initiated the transaction at the point of service, the ACSS clears the transaction with a delay.[4][5]

[3] See CBA: https://cba.ca/abm-market-in-canada.

[4] A transaction on a day when the ACSS is open is assumed to clear on the first day after the transaction that the ACSS is also open. If the transaction takes place on a day when the ACSS is closed, it is assumed to clear on the second day after the transaction that the ACSS is also open.

[5] Financial institutions may prefer on-us transactions because they do not experience a delay in settlement and because they are not exposed to the risk that another financial institution cannot meet its obligation (Northcott, 2002).

2.2 COVID-19 Indicators

To measure the pandemic's intensity and evolution, we consider a few alternative indicators: government stringency and COVID-19 cases. The data for government stringency and COVID-19 cases is taken from the COVID-19 dataset collection maintained by Our World in Data (Roser et al., 2020). The COVID-19 government stringency index is part of the Oxford COVID-19 Government Response Tracker (see Hale et al., 2020).[6] It is a composite measure of nine responses indicators—school closures, workplace closures, cancellation of public events, restrictions on public gatherings, closures of public transport, stay-at-home requirements, public information campaigns, restrictions on internal movements, and international travel controls—rescaled to a value from 0 to 100 (100 = strictest).[7] Daily data for confirmed COVID-19 cases originates from the COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University.

Figure 2 shows the evolution of these indicators over our sample period (January 1st, 2019–February 26th, 2021). The first wave of the pandemic in the spring of 2020 led to strict government measures that were eased partially over the summer due to relatively stable new case counts. In the fall, Canada entered the second wave as new cases and the positivity rate increased.

3 Measures of Payment Habits and First Insights

3.1 Cash-Card Ratios

To measure one of the aspects of payment habits related to cash usage, we calculate the cash-card ratio. The cash-card ratio has been used as a proxy for the share of cash transactions at the POS (see, e.g., Rocco, 2020). The ratio of cash withdrawals to card transactions (cash-card ratio) is constructed by dividing the amount of ABM cash transactions by the amount of POS debit transactions. We observe that the cash-card ratio is higher in the

[6] It is important to note that this index simply records the strictness of government policies. It does not measure or imply the appropriateness or effectiveness of a country's response.

[7] If policies vary at the sub-national level, the index is shown as the response level of the strictest sub-region.

Interac data than in the ACSS data (Figure 3) before and during the pandemic. This is explained by our later discussion of the ACSS-Interac ratios; namely, the proportion of on-us ABM withdrawals in 2019 was about three to four times larger than the proportion of on-us POS payments in the same time period. In the Interac data, we find therefore that in the pre-pandemic period the value of banknotes withdrawn relative to debit card payments was about 43%, while the ratio is approximately 13% in the ACSS data. With the onset of the pandemic in March 2020, the value and volume ratios drop in both data series but remain below pre-pandemic averages until the end of our sample. Their behaviour also seems to be linked to the evolution of the pandemic. For example, the cash-card ratio increased somewhat over the summer of 2020 (when there were relatively low and stable COVID-19 case counts) and decreased again towards the end of the year (when Canada entered the second wave).

In addition to transaction data, central banks mostly rely on surveys to estimate the share of cash transactions. To that end, we compare the cash-card ratio based on debit card transactions to estimates derived from Canadian survey data for the year 2017. Specifically, Henry et al. (2018) estimate the value share of cash transactions at the POS at 16% and the volume share at 33%, relative to debit card shares of 26% for the value and 24% for the volume. From these estimates, the value share of cash transactions would be around 38%, obtained by dividing the cash value share of 16% by the sum of the value shares of cash and debit (26%). By contrast, the cash-card ratio in the Interac data would imply a value share of 30%.[8]

Several factors may explain the 8 pp higher cash share in the survey data than in the Interac data: First, while Henry et al. (2018) report that ABM withdrawals are the most common way to obtain banknotes, bank teller withdrawals and cash-back withdrawals during a debit card purchase also play a significant role. Second, using withdrawals in the cash-card ratio ignores the possibility that the same banknote is used in several cash transactions before it is deposited at an FI. On the other hand, there are also factors that would support a higher cash-card ratio constructed from point-of-sale transactions in the survey

[8]The cash-card ratio implies that the value of ATM withdrawals is 0.43 times the value of debit card transactions. Hence the value share of cash is $0.43/(1 + 0.43) = 0.30$.

data than the cash-card ratio constructed from ATM withdrawals. First, individuals also withdraw and hold cash for non-transactional purposes (savings, precautionary motives). Second, cash is also used for person-to-person transactions. Third, Henry et al. (2018) observe a downward trend in the value share of cash transactions, which might indicate that the cash-card ratio at the POS in 2019 might be lower than in the 2017 survey.

3.2 Average Transaction Values

We also consider average transaction values as another aspect of payment habits. Average transaction values are calculated as the transaction value for a given day divided by the transaction volume for the same day. In 2019, a typical ATM withdrawal amount was around $178 in the Interac data and about $145 in the ACSS data (see Figure 4). The data shows an upward shift in the size of withdrawal between January and February 2020. The gap between Interac and ACSS withdrawal increased at the onset of the pandemic. Also, both series seem to co-move with the intensity of the pandemic. By contrast, the typical debit card transaction processed by Interac was around $39, close to the typical value of $38 for the ACSS. Overall, both POS averages increase and remain highly correlated over the duration of the pandemic thus far.

3.3 ACSS-Interac Ratios

We calculate the ACSS-Interac ratios for ATM withdrawals and POS transactions to gauge the non-on-us transactions. To do so, we create a second harmonized dataset from the Interac streams by aggregating processed amounts by the day that they would have been cleared by the ACSS. For example, we aggregate Interac transactions from Saturday, Sunday and Monday and compare them to ACSS transactions for Tuesday. Tuesday's Interac transactions are compared to Wednesday's ACSS observations. This dataset allows us to compare Interac and ACSS data at the daily frequency and to calculate the ratios.

Figure 5 shows the Interac-ACSS ratios over time. For 2019, 22% of the value of ABM transactions was cleared by the ACSS; correspondingly, 78% of the value was contributed by on-us transactions. For the volume, 28% of the withdrawals were cleared by the ACSS (72%

on-us transactions). In other words, cardholders most commonly make on-us withdrawals, and they withdraw higher amounts during on-us transactions than during non-on-us transactions.

A likely explanation is that cardholders can often avoid account fees and other charges if they make withdrawals at an ABM affiliated with the issuer of their card. Also, bank-owned ABMs might allow for higher withdrawal amounts since the owner of the white-label ATM might set the maximum withdrawal amount below the maximum set by the card issuer.

During the pandemic, the ACSS-Interac ratios for cash plunged during the first lockdown, increased during the reopening, and fell again during the second lockdown (Figure 5, red lines). Turning to POS debit payments, the share of non-on-us POS debit transactions was quite steady around 80% for the value. While the volume share averaged around 83% in 2019, it increased during the second half of that year. The cardholder is usually unaware of which FI will acquire a card payment, but changes in the concentration on the issuing or acquiring side can still drive this share of non-on-us card payments. In contrast to cash, the ACSS-Interac ratio for these transactions steadily increased over the course of the pandemic (Figure 5, blue lines).

4 Empirical Approach

To analyze the changes in payment habits throughout the COVID-19 pandemic more formally, we use two approaches. First, we employ a static dummy regression. Then, we estimate a local projection model to assess the dynamic effects of our COVID-19 indicators.

4.1 Static Model: Dummy Regressions

We use a simple dummy regression to explore the change in the variables of interest during different phases of the pandemic. This allows us to quantify the shift in each COVID-19 phase compared to the pre-pandemic period (2019 to early 2020) that we observed in the previous section.

We roughly define four phases: lockdown 1 (March 2020–April 2020), reopening 1 (May

2020–October 2020), lockdown 2 (November 2020–January 2021), reopening 2 (February 2021).[9] We construct four dummies, one for each phase. We control for seasonal patterns and trends by including day-of-week, week-of-month, and month-of-year dummies.[10] We regress the variables of interest described in the former section on the COVID-19 phase dummies, seasonal dummies, and a constant over a daily sample covering January 1st, 2019–February 26th, 2021.

The cash-card ratios both in value and volume terms decreased over the pandemic. Cash-card value ratios were up to 9 pp and 6 pp lower for Interac and ACSS transactions, respectively.

Average transaction values increased during the pandemic. For cash withdrawals, the average withdrawal size increased by up to 31% in the Interac data and by up to 15% in the ACSS data. In both datasets, the average withdrawal amount was highest during the first lockdown and decreased afterwards (while remaining above pre-pandemic levels). Similarly, the average value of a POS debit card payment increased by 11% and 12% for the Interac data and ACSS data, respectively, during the first lockdown and remains elevated.

Overall, the ACSS-Interac ratio decreased over the pandemic for cash and increased for card transactions. For cash, the ratio was up to 6 pp and 4 pp lower for the value and the volume, respectively. For cards, the increase was up to 4 pp for the value and up to 6 pp for the volume (Table 2).

4.2 Dynamic Model: Local Projections

To evaluate the dynamic effects and their persistence during the pandemic, we use the local projection impulse response estimator introduced by Jordà (2005). We obtain the impulse responses of our variables of interest to the COVID-19 indicators by estimating the following

[9]The phases are motivated by the behaviour of cases and stringency across Canada. For example, we roughly define the lockdown phases as periods when COVID-19 cases are increasing and provincial governments such as Ontario imposed lockdown measures. Since lockdowns have not been imposed at the national level, the phases may have varied at the provincial level. The phases defined here are, thus, a rough generalization at the national level.

[10]In the case of Interac, we also need a day-of-month dummy to fully control for seasonal patterns. Since the ACSS data comes at the business-day frequency, we also include dummies for the day of the business week.

regressions over our daily sample from January 1st, 2019–February 26th, 2021:

$$y_{t+h} = a_h + \sum_{i=1}^{p} b_{i,h} y_{t-1} + c_h x_t + e_{t+h}, \quad h = 1, ..., 100, \tag{1}$$

where y_{t+h} is our variable of interest (cash-card value and volume ratios, logarithms of average ABM and POS transaction values, ACSS-Interac ABM value and volume ratios, and ACSS-Interac POS value and volume ratios), x_t is one of the two exogenous COVID-19 indicators (logarithm of COVID-19 cases or logarithm of the government stringency index), and p is the number of lags. We assume that the COVID-19 indicators are exogenous since it is unlikely that transactions affect these indicators (by affecting the transmission of the virus) at the daily frequency. However, endogeneity concerns can become relevant at lower frequencies.

As before, we control for seasonal patterns by including day-of-week, day-of-month, week-of-month, and month-of-year dummies. To render the responses of ACSS and Interac variables comparable in terms of timing, we use the Interac data harmonized to business-day frequency. The estimation includes 10 lags, corresponding to two weeks.[11] We consider horizons of up to 100 business days.

Figure 6 shows impulse responses for the cash-card value and volume ratios to the two COVID-19 indicators. We find that a 100% increase in new cases results in a drop of 0.4 pp on impact in the cash-card value ratio. The responses reach their peak at about -0.9 pp and -1.5 pp for the ACSS and Interac transactions, respectively. While the response of the cash-card ratio measured by ACSS data reverts after about 40 business days (although not to pre-shock levels), the cash-card value ratio measured by Interac data continues to fall.

Similar patterns arise when looking at the responses to an increase in the stringency index. The Interac cash-card value ratio falls by between about 0.7 pp and 2.5 pp following a 100% increase in the stringency index over a horizon of about 80 business days. The ACSS cash-card value ratio decreases by between 0.8 pp and 1.7 pp over a horizon of about 40 business days. Again, the responses of the cash-card value ratios based on Interac and

[11] Note that responses are similar when we consider a smaller number of lags.

ACSS start to diverge after about 40 business days.

Similar patterns are observed if we consider the responses of the cash-card volume ratios to an increase in cases and the stringency index. However, the responses following an increase in cases revert somewhat to pre-shock levels, albeit not significantly.

Turning to the dynamic responses of ABM average transaction values, we find that following a 100% change in the stringency index, the ABM average transaction value measured by ACSS increases by about 2 pp over all horizons (see Figure 7, panels a and b). The response of ABM average transaction values based on Interac is similar on impact but reaches a maximum of about 8 pp after about 40 business days. Similar responses arise, although slightly less pronounced, following a percent change in COVID-19 cases. The observed difference between the dynamic responses of ABM average transaction values measured by ACSS and Interac is in line with the point made earlier: following the onset of the pandemic, there were relatively more and higher-value on-us ABM withdrawals versus non-on-us.

POS average transaction values obtained from ACSS and Interac behave very similarly in response to a change in stringency and cases (Figure 7, panels c and d). While persistently positive, the responses start to revert after about 40 business days.

We now discuss the responses of the ACSS-Interac ratios reflecting the behaviour of non-on-us transactions over the pandemic (Figure 8).[12] A change in stringency decreases the ACSS-Interac ratio for ABM withdrawals in both value and volume terms. The responses start to revert after about 40 business days, although they do not recover to pre-shock levels. The observed dynamic responses of the ACSS-Interac ratio, therefore, support our claim that the differences in responses between ACSS and Interac data over the pandemic are related to the distinct behaviours of non-on-us and on-us transactions throughout the pandemic.

[12]For the sake of brevity, we show only impulse responses to a change in stringency. The responses to a change in COVID-19 cases are very similar.

4.3 Adjustment to include cash transactions

How cash has evolved relative to card payments is key in understanding the link between high-frequency transaction data and household spending. Estimates of economic activity based on card transaction data alone could be biased if shifts in payment behaviour, such as a change in cash use, are not taken into account. For example, growth in consumer spending could be mis-estimated from card transaction data as payment choices seem to have changed during the pandemic. In what follows, we illustrate how the cash-card ratio is relevant for linking card payment data to nominal household expenditure. We provide insights on how a shift from cash towards debit cards affects the dynamics of cards payments and calculate the potential bias if debit card payments were used as a proxy for consumption expenditure.

We construct two proxies for growth in nominal expenditure, i.e., one based on the value of POS debit card transactions alone and one adjusted for cash (sum of cash and debit card transactions).[13] In what follows, we base our results on the Interac data because it has a better coverage of transactions compared to the ACSS data and also because the cash share in the Interac data is closer to, e.g., the cash share based on survey data in Henry et al. (2018). In particular, the growth rates in expenditure between time t_1 and t_2 based on the value of POS debit card transactions alone (g_{card}) and adjusted for cash (g_{adj}) are given by:

$$
\begin{aligned}
g_{card} &= \frac{Card_{t_2}}{Card_{t_1}} - 1; \\
g_{adj} &= \frac{Card_{t_2} + Cash_{t_2}}{Card_{t_1} + Cash_{t_1}} - 1 = \frac{1 + y_{t_2}}{1 + y_{t_1}}(g_{card} + 1) - 1,
\end{aligned}
$$

where y_t denotes the cash-card ratio. Note that in the formula for g_{adj}, the quantity $1 + y_t$ is the inverse of the card share at time t since $1 + y_t = 1 + \frac{Cash_t}{Card_t} = \frac{Card_t + Cash_t}{Card_t}$.

The difference between g_{adj} and g_{card} reflects the adjustment due to including cash

[13] Note that therefore our calculations reflect the assumption that all withdrawn cash is used for consumption.

transactions. It can be expressed as:

$$g_{adj} - g_{card} = \frac{1 + y_{t_2}}{1 + y_{t_1}}(g_{card} + 1) - (g_{card} + 1) = \frac{y_{t_2} - y_{t_1}}{1 + y_{t_1}}(g_{card} + 1).$$

Note that $g_{card} + 1 \geq 0$ and $1 + y_{t_1} \geq 0$; hence the direction of the adjustment is given by the sign of $y_{t_2} - y_{t_1}$, the difference in cash-card ratios. Now assume that a series of COVID-19 shocks $x_{t'}$ occur at times t', $t' < t_2$. We decompose the cash-card ratio as $y_{t_2} = \tilde{y}_{t_2} + \Delta_{t_2}$. Here Δ_{t_2} is the sum of all the impulse responses to the COVID-19 shocks (changes in COVID-19 indicators), $\sum_{t'} c_{t_2 - t'} x_{t'}$, and $\tilde{y}_{t_2}$ is the component of y_{t_2} that cannot be explained by these shocks. We can thus rewrite the adjustment as:

$$g_{adj} - g_{card} = \underbrace{\frac{\Delta_{t_2}}{1 + y_{t_1}}(g_{card} + 1)}_{\text{Impulse response to COVID-19}} + \underbrace{\frac{\tilde{y}_{t_2} - y_{t_1}}{1 + y_{t_1}}(g_{card} + 1)}_{\text{residual}}.$$

This decomposition represents the adjustment as the sum of a component explained by COVID-19 shocks and a residual.

Figure 10 shows the evolution of the two growth rates based on transactions in the Interac data since the beginning of 2019. Time is measured in days, and we set $t_1 = t_2 - 364$ to obtain year-over-year (y/y) daily growth rates.[14] Until the onset of the pandemic and during the early phases of the first lockdown, daily y/y card-based and adjusted growth rates are similar, and the adjustment is around -1 pp, likely reflecting a pre-pandemic trend towards card payments. The cash-card ratio falls sharply until July 2020 and reaches -7 pp. While the y/y card-based growth and adjusted growth rates picked up during the summer, the adjusted growth rate still remained below the card-based one. In particular, card-based growth rates would imply positive growth for some of the second half of 2020, although adjusted growth rates might indicate that nominal spending remained below 2019 levels. The adjustment remained at -7 pp from June to August 2020, then increases to -1 pp in December 2020 and falls again to -6 pp in February 2021.

Further, we decompose the cash adjustment for the y/y growth rate during the first 11

[14] Using 364 as the time step instead of 365 or 366 (for a leap year) ensures that weekday effects are factored out.

months of the pandemic into the COVID-19 component and the residual. Since impulse responses are only available for the first 100 business days, we must make assumptions about the shocks for the remaining time period. In Scenario A the shock is assumed to be a simple moving average after the first 100 days, corresponding to persistent shocks. In Scenario B, the simple moving average is multiplied by an attenuation factor of 0.975 for each business day. In Scenario B, the assumption is that shocks are not persistent after 100 business days.

We first consider shocks to the newly reported COVID-19 cases. Scenario A as shown in Figure 11 shows that COVID-19 shocks explain most of the cash adjustment for the period from late May 2020 to early December 2020. However, the first months of the pandemic and also the period of late December 2020 to early January 2021 are characterized by smaller adjustments, although the COVID-19 component is of similar magnitude as in the aforementioned period. A pickup in the cash-card ratio (despite COVID-19 shocks) drives this difference and might be due to non-transactional motives to withdraw cash. For example, precautionary cash holdings increased at the beginning of the pandemic (Chen et al., 2020). Given restrictions on shopping outlets, consumers might also have given cash gifts during the holiday season.

Scenario B for attenuating shocks behaves identically to Scenario A until the 100th business day of the pandemic, by construction. However, afterwards the COVID-19 shocks explain only about half of the cash adjustment between August and December 2020. In contrast, the COVID-19 component explains a large part of the adjustments in December 2020 and January 2021, and a smaller part in February 2021.

The decomposition of the cash adjustment into a COVID-19 stringency component and a residual are shown in Figure 12. While this decomposition is similar to the previous one, the model with attenuating shocks would imply that other factors behind the discrepancy between card-based and adjusted growth rates become more pronounced in the later stage of the pandemic, as the residual falls to around -4 pp.

5 Conclusion

High-frequency data on ATM cash withdrawals and card transactions reveals that, since the onset of the pandemic, cash withdrawals, especially withdrawals from ATMs not belonging to the cardholder's bank, have dropped in comparison to debit card transactions. This decline is more pronounced for the dollar value, although the average dollar value of both cash withdrawals and card transactions has increased. At the same time, a larger share of POS debit card transactions now involves a consumer and a merchant who do not have their accounts with the same FI.

We find evidence that during the pandemic consumers adjusted their behaviour to avoid frequent trips for cash withdrawals and POS purchases. Instead, they performed fewer transactions but for higher amounts. They reduced the value of cash withdrawals compared with the value of card payments, and they performed a smaller number of cash transactions than card transactions. Consumers also made relatively more withdrawals from ATMs that linked to their FI (on-us), possibly because they planned their withdrawals or because they could not access white-label ATMs that are often located inside of businesses. The larger share of on-us transactions could also be explained by reduced mobility which kept consumers close to their home bank. These reasons are consistent with a decrease of the on-us share during the reopening of businesses and the lifting of mobility restrictions. However, neither the cash-card share nor the non-on-us share recovered to pre-pandemic levels during the reopening, and both fell again during the second lockdown. It is not yet evident from our dynamic analysis whether this is due to the short time period between the two lockdowns or is a sign of a persistent shift towards card payments and also towards withdrawals at FI-owned ATMs.

The share of non-on-us debit POS transactions increased during the pandemic. Consumers are generally unaware of whether a card payment is non-on-us or on-us. However, shifts towards e-commerce and larger retailers as well closures in certain industries might have impacted the FIs' issuing and acquiring lines in an asymmetric way, hence leading to changes in their non-on-us shares. Microdata from consumers and businesses or FIs could

deliver more insights.

Since card network data has increasingly been used to understand shifting expenditure patterns during COVID-19, the cash-card ratios in the Interac and ACSS data can be used to understand the direction of the bias of the estimated expenditure growth based on card network data. While some caveats apply, the decline in the cash-card ratio might be an indicator that expenditure growth estimated from card network data alone is biased upwards.

References

Aastveit, K. A., Fastbø, T. M., Granziera, E., Paulsen, K. S. and Torstensen, K. N. (2020). Nowcasting norwegian household consumption with debit card transaction data, *Working Paper 17/2020*, Norges Bank.

Aprigliano, V., Ardizzi, G. and Monteforte, L. (2019). Using Payment System Data to Forecast Economic Activity, *International Journal of Central Banking* **15**(4): 55–80.

Cavallo, A. (2020). Inflation with Covid Consumption Baskets, *NBER Working Papers*, National Bureau of Economic Research.

Chapman, J. and Desai, A. (2021). Using Payments Data to Nowcast Macroeconomic Variables During the Onset of COVID-19, *Staff Working Paper 2021-2*, Bank of Canada.

Chen, H., Engert, W., Huynh, K., Nicholls, G., Nicholson, M. and Zhu, J. (2020). Cash and COVID-19: The impact of the pandemic on demand for and use of cash, *Staff Discussion Paper 2020-6*, Bank of Canada.

Chodorow-Reich, G., Gopinath, G., Mishra, P. and Narayanan, A. (2019). Cash and the Economy: Evidence from India's Demonetization, *The Quarterly Journal of Economics* **135**(1): 57–103.

Esselink, H. and Gijsel, L. H.-v. (2017). The use of cash by households in the euro area, *Occasional Paper Series 201*, European Central Bank.

European Payments Council (2019). Greece's steady progress towards a cashless society, https://www.europeanpaymentscouncil.eu/news-insights/insight/greeces-steady-progress-towards-cashless-society. Accessed: 2021-08-15.

Galbraith, J. and Tkacz, G. (2013). Analyzing Economic Effects of September 11 and Other Extreme Events Using Debit and Payments System Data, *Canadian Public Policy* **39**(1): 119–134.

Galbraith, J. W. and Tkacz, G. (2018). Nowcasting with payments system data, *International Journal of Forecasting* **34**(2): 366–376.

Hale, T., Webster, S., Petherick, A., Phillips, T. and Kira, B. (2020). Oxford COVID-19 Government Response Tracker, *Technical report*, Blavatnik School of Government.

Henry, C., Huynh, K. and Welte, A. (2018). 2017 Methods-of-Payment Survey Report, *Staff Discussion Paper 2018-17*, Bank of Canada.

Huynh, K., Lao, H., Sabourin, P. and Welte, A. (2020). What do high-frequency expenditure network data reveal about spending and inflation during covid-19?, *Staff Analytical Note 2020-20*, Bank of Canada.

Jordà, O. (2005). Estimation and inference of impulse responses by local projections, *American Economic Review* **95**(1): 161–182.

Lahiri, A. (2020). The great indian demonetization, *Journal of Economic Perspectives* **34**(1): 55–74.

Northcott, C. A. (2002). Estimating settlement risk and the potential for contagion in canada's automated clearing settlement system, *Staff working papers*, Bank of Canada. **URL:** *https://EconPapers.repec.org/RePEc:bca:bocawp:02-41*

Rocco, G. A. . A. N. . G. (2020). A game changer in payment habits: evidence from daily data during a pandemic, *Questioni di Economia e Finanza (Occasional Papers) 591*, Bank of Italy.

Roser, M., Ritchie, H., Ortiz-Ospina, E. and Hasell, J. (2020). Coronavirus Pandemic (COVID-19, *Technical report*, OurWorldInData.org.

URL: *https://ourworldindata.org/coronavirus*

Tables and Figures

Table 1: *Dummy Regressions of Selected Variables: Cash-Card Ratios and Average Transaction Values*

		Cash-Card Value Ratio	Cash-Card Volume Ratio	log(ABM) Average Value)	log(POS) Average Value)
Lockdown 1	Interac	-0.040*	-0.025*	0.306*	0.110*
	ACSS	-0.041*	-0.012*	0.154*	0.118*
Reopening 1	Interac	-0.085*	-0.029*	0.234*	0.091*
	ACSS	-0.045*	-0.012*	0.079*	0.083*
Lockdown 2	Interac	-0.081*	-0.027*	0.230*	0.093*
	ACSS	-0.050*	-0.012*	0.049*	0.049*
Reopening 2	Interac	-0.109*	-0.031*	0.231*	0.121*
	ACSS	-0.059*	-0.014*	0.051	0.114 *
# of observations	Interac	752	752	752	752
	ACSS	528	528	528	528

*Notes: * denotes significance at the 1% level based on standard errors robust against heteroskedasticity and autocorrelation (HAC) in the residuals. Daily sample from January 1st, 2019 –February 26th, 2021.*

Table 2: *Dummy Regressions of Selected Variables: ACSS-Interac Ratios*

	ABM Value	ABM Volume	POS Value	POS Volume
Lockdown 1	-0.057*	-0.040*	0.003	0.010
Reopening 1	-0.040*	-0.013*	0.016	0.030*
Lockdown 2	-0.057*	-0.025*	0.029*	0.044*
Reopening 2	-0.054*	-0.024*	0.044	0.055*
# of observations	528	528	528	528

*Notes: * denotes significance at the 1% level based on standard errors robust against heteroskedasticity and autocorrelation (HAC) in the residuals. Daily sample from January 1st, 2019–February 26th, 2021.*

Figure 1: *Network Schematic of Interac and ACSS*

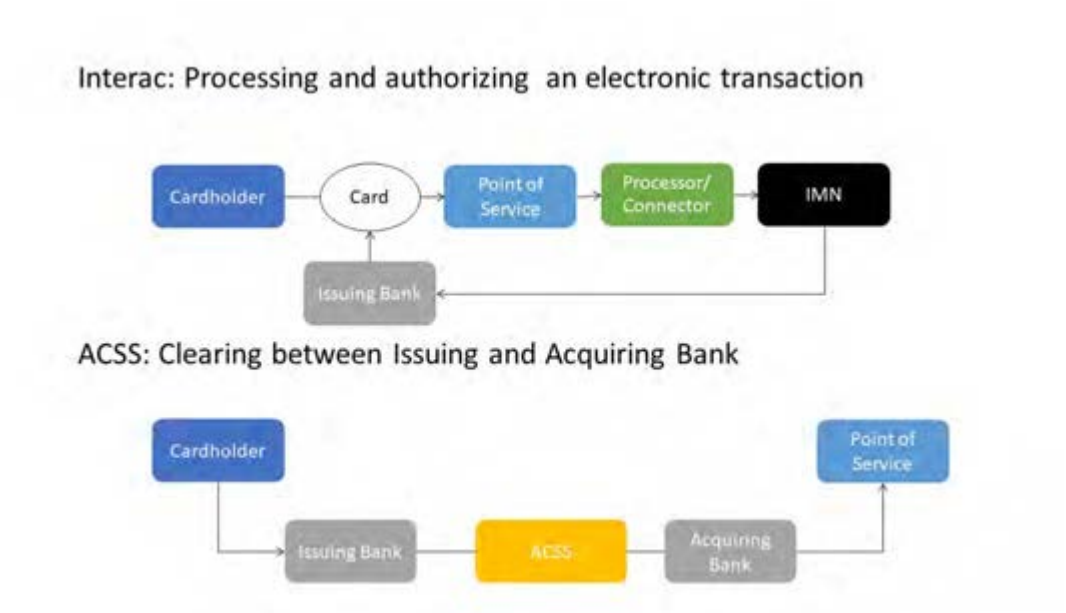


Figure 2: *COVID-19 Indicators: Daily New COVID-19 Cases and Stringency*

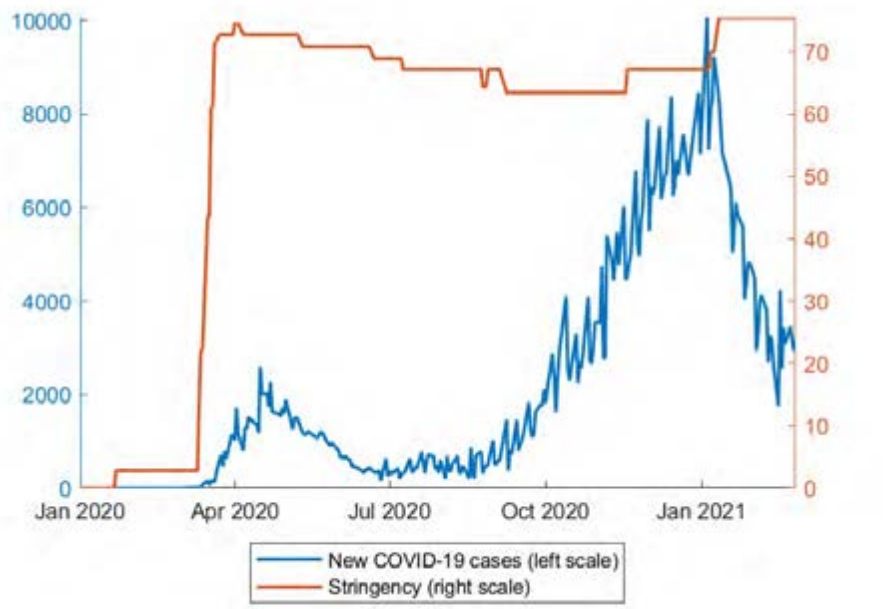
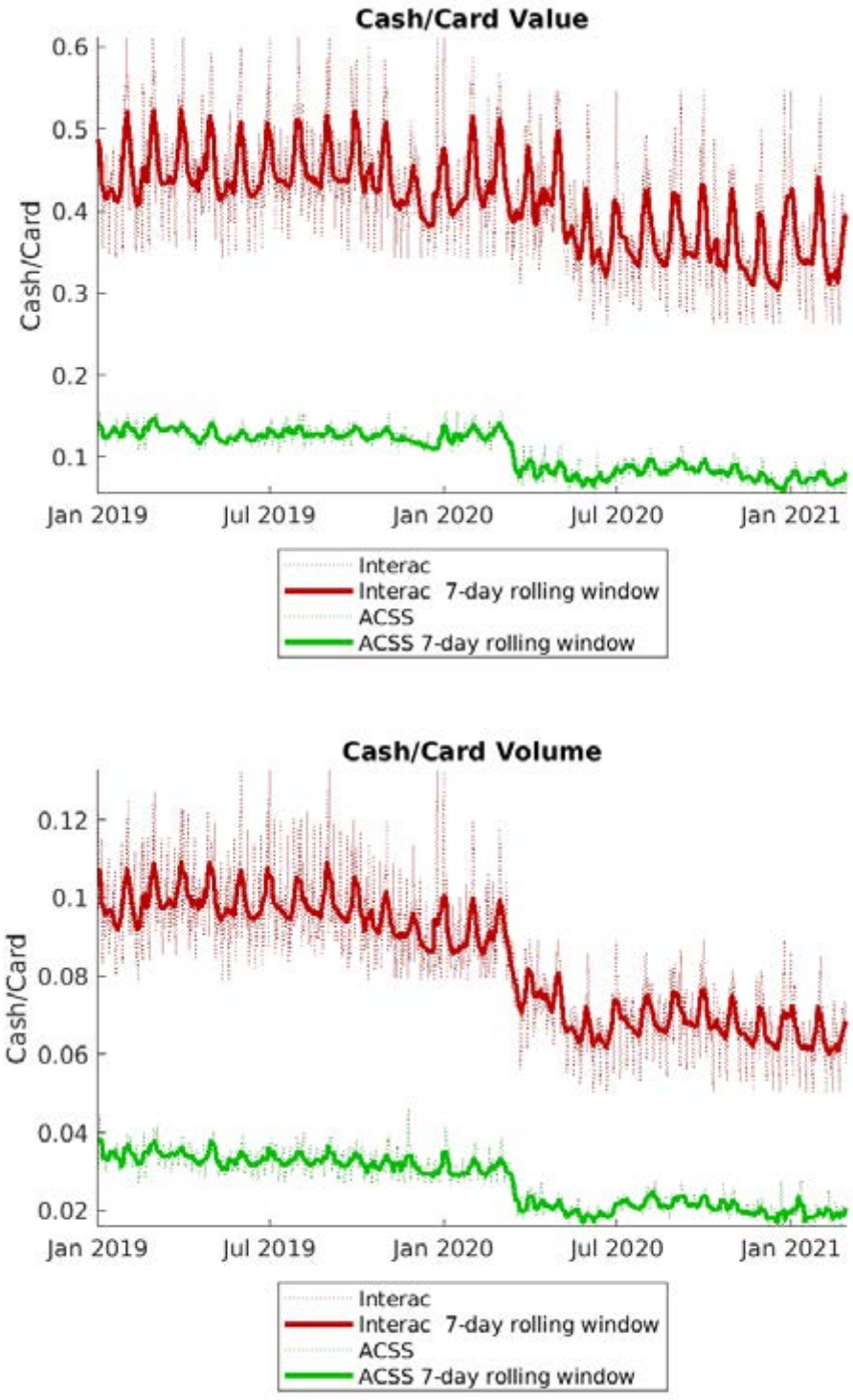
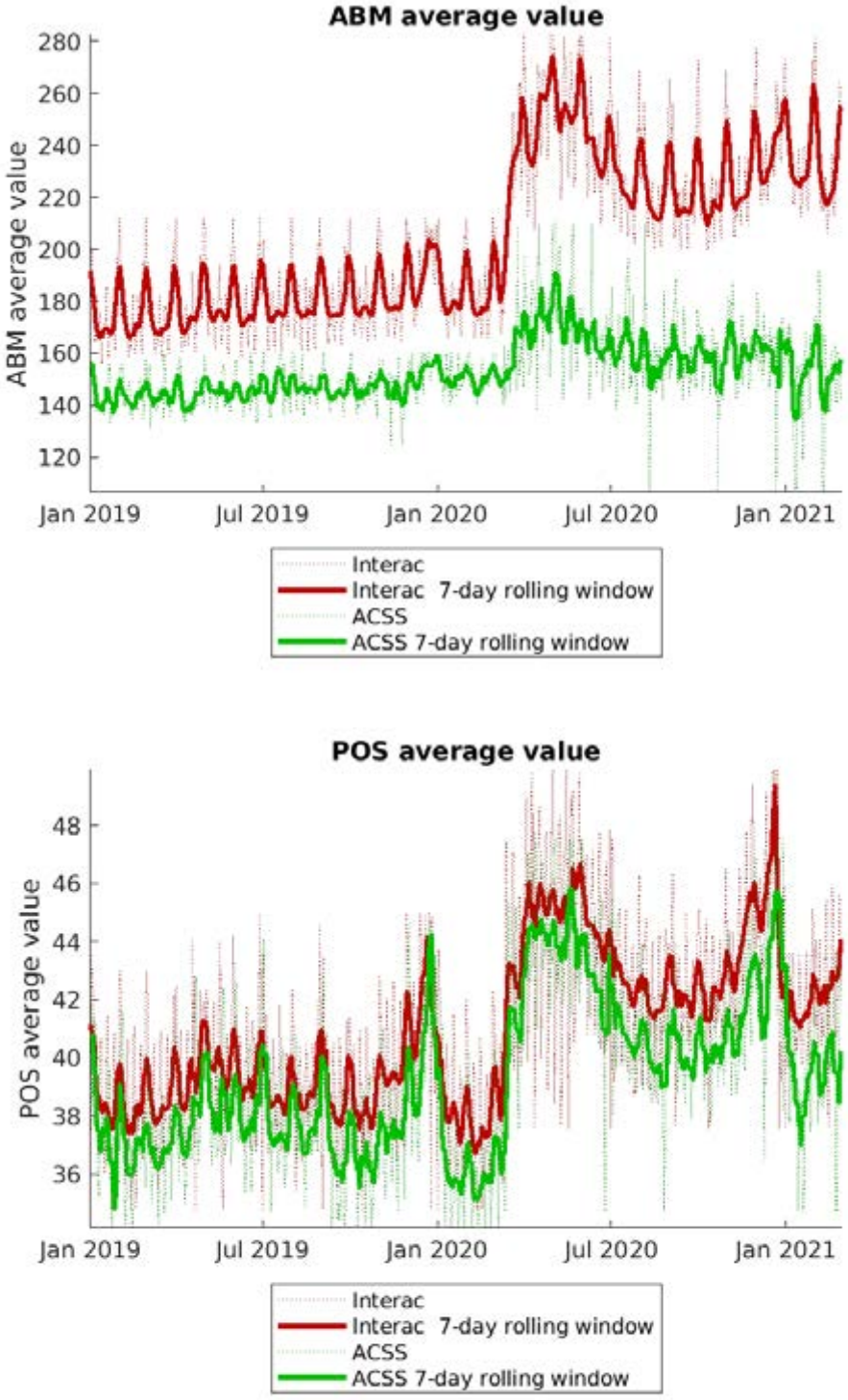


Figure 3: *Cash-Card Ratios*



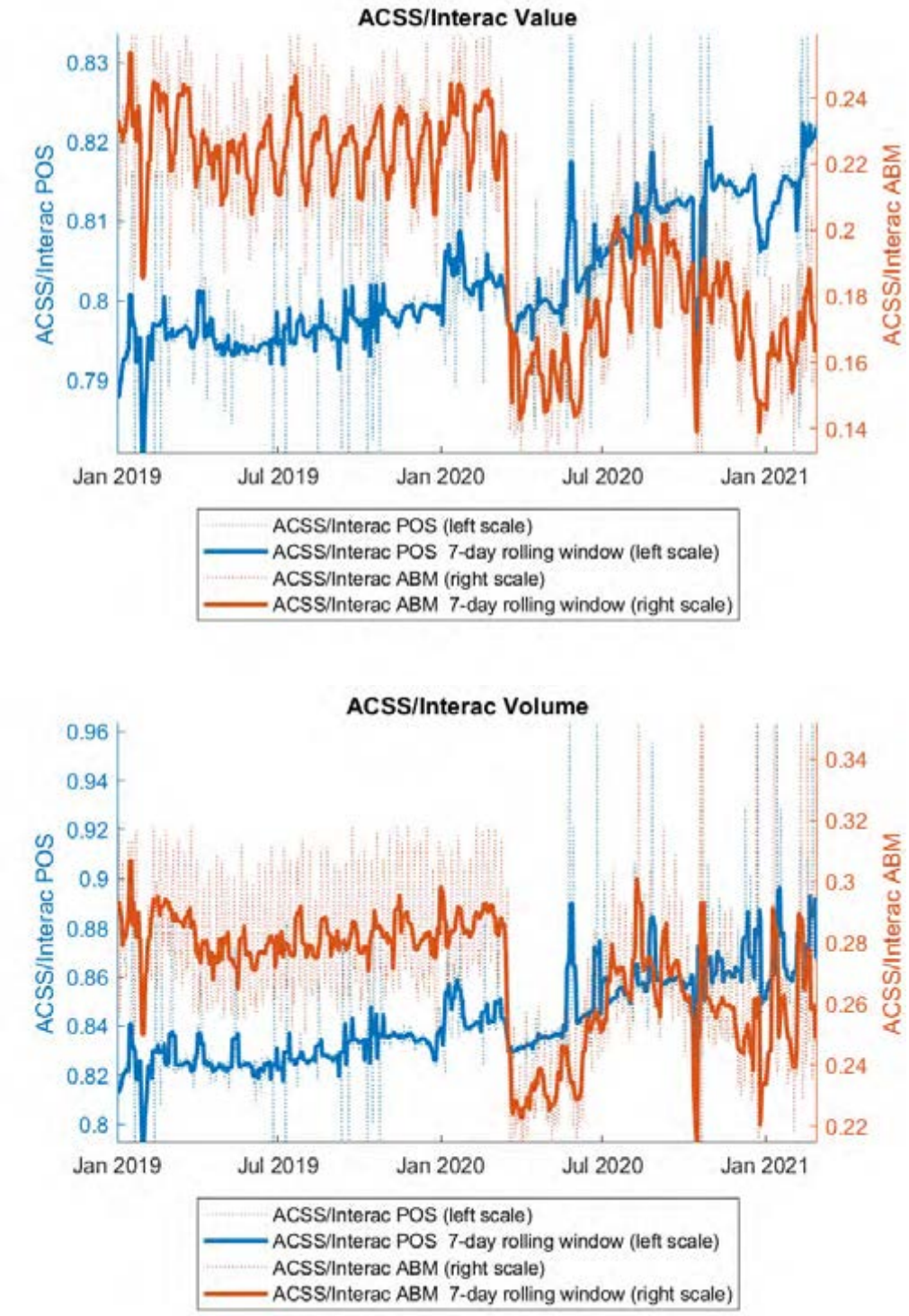
Notes: Daily values are winsorized at the 2.5 and 97.5 percentiles for presentation purposes. Percentiles are computed separately for the periods before and during the pandemic.

Figure 4: *Average Transaction Values*



Notes: Daily values are winsorized at the 2.5 and 97.5 percentiles for presentation purposes. Percentiles are computed separately for the periods before and during the pandemic.

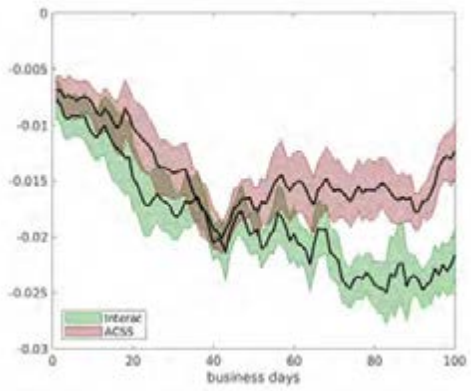
Figure 5: *ACSS Transactions as a Share of Interac Transactions*



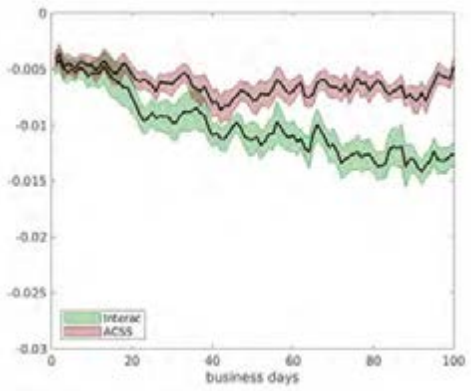
Notes: Daily values are winsorized at the 2.5 and 97.5 percentiles for presentation purposes. Percentiles are computed separately for the periods before and during the pandemic.

Figure 6: *Impulse Responses of Cash-Card Value and Volume Ratios for Interac and ACSS Transactions*

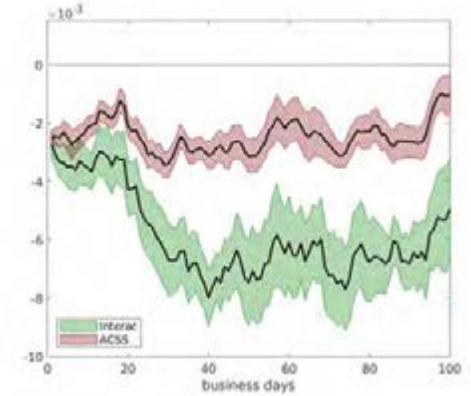
(a) *Cash-Card Ratio: Value Stringency*



(b) *Cash-Card Ratio: Value Cases*



(c) *Cash-Card Ratio: Volume Stringency*



(d) *Cash-Card Ratio: Volume Cases*

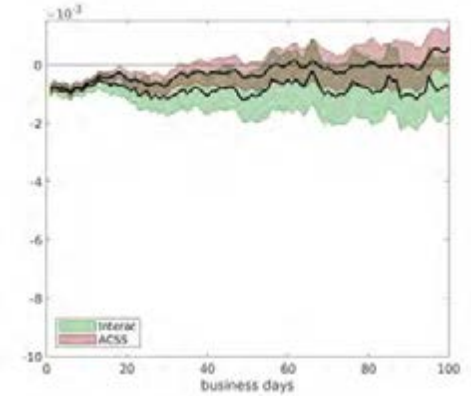
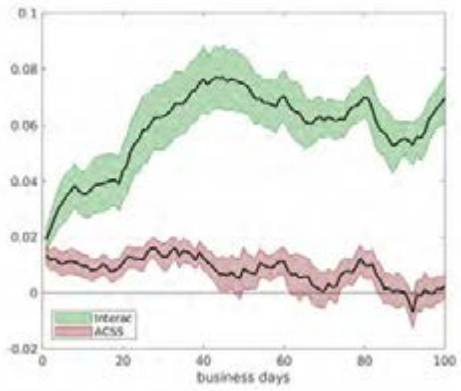
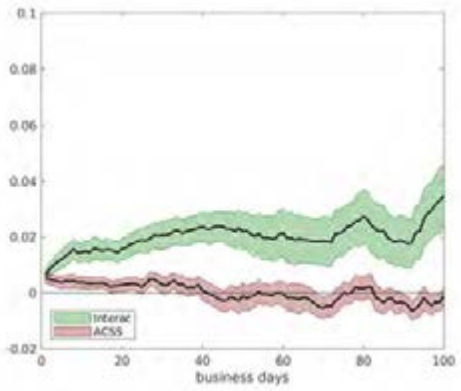


Figure 7: *Impulse Responses of ABM and POS Average Transaction Values for Interac and ACSS Transactions*

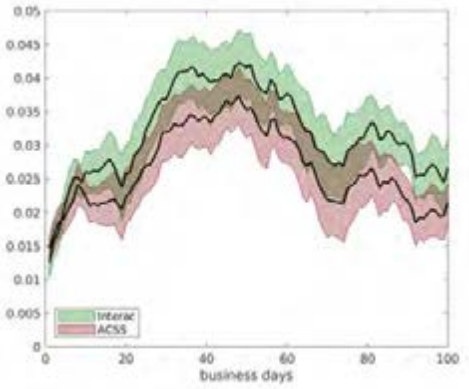
(a) *ABM Average Value: Stringency*



(b) *ABM Average Value: Cases*



(c) *POS Average Value: Stringency*



(d) *POS Average Value: Cases*

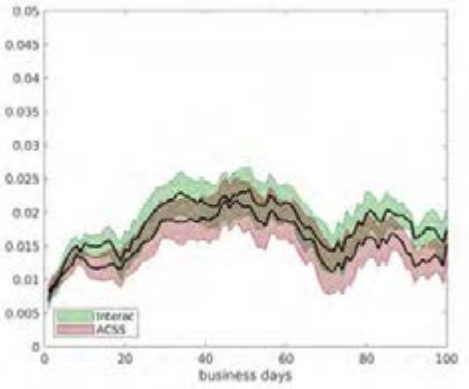
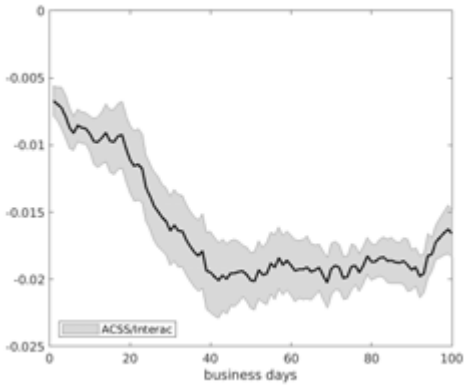
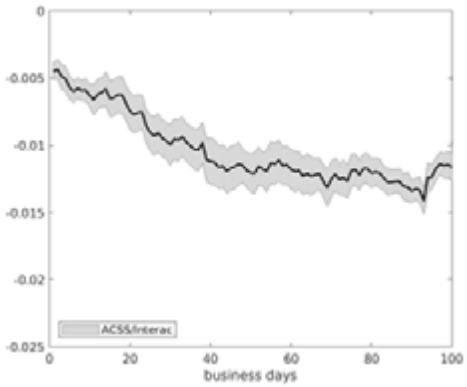


Figure 8: *Impulse Responses of ACSS-Interac Value Ratios for ABM and POS Transactions: Cases*

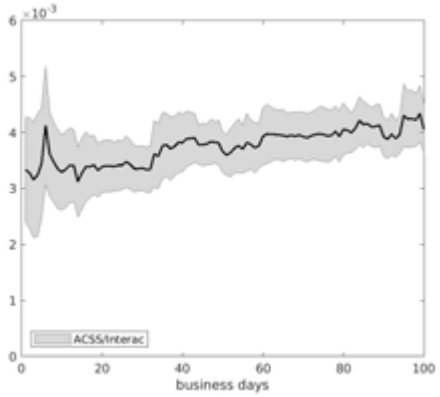
(a) *ACSS-Interac ABM Value: Stringency*



(b) *ACSS-Interac ABM Value: Cases*



(c) *ACSS-Interac POS Value: Stringency*



(d) *ACSS-Interac POS Value: Cases*

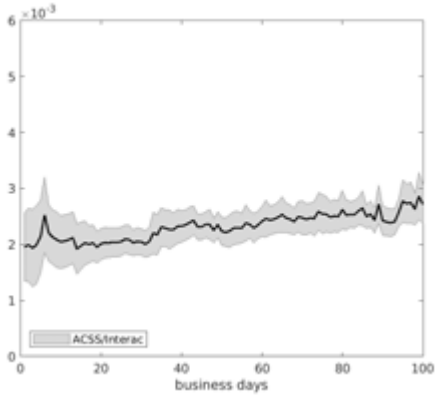
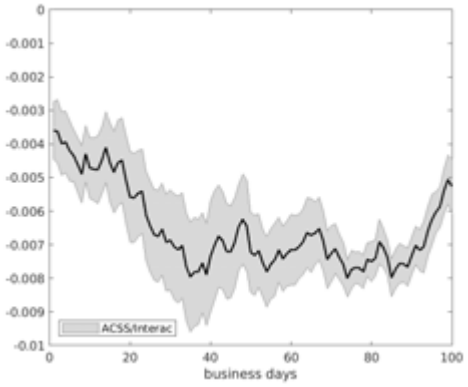
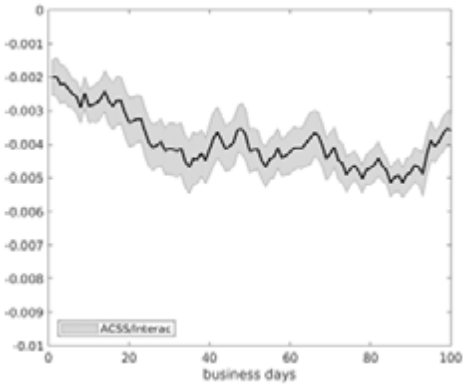


Figure 9: *Impulse Responses of ACSS-Interac Volume Ratios for ABM and POS Transactions: Cases*

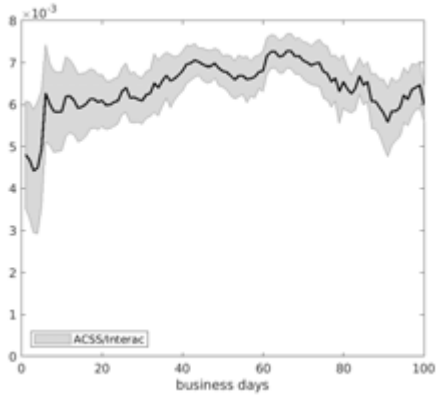
(a) *ACSS-Interac ABM Volume: Stringency*



(b) *ACSS-Interac ABM Volume: Cases*



(c) *ACSS-Interac POS Volume: Stringency*



(d) *ACSS-Interac POS Volume: Cases*

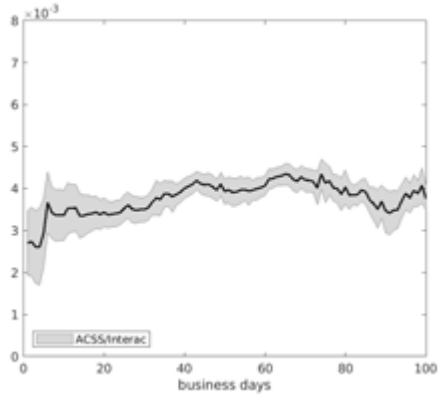
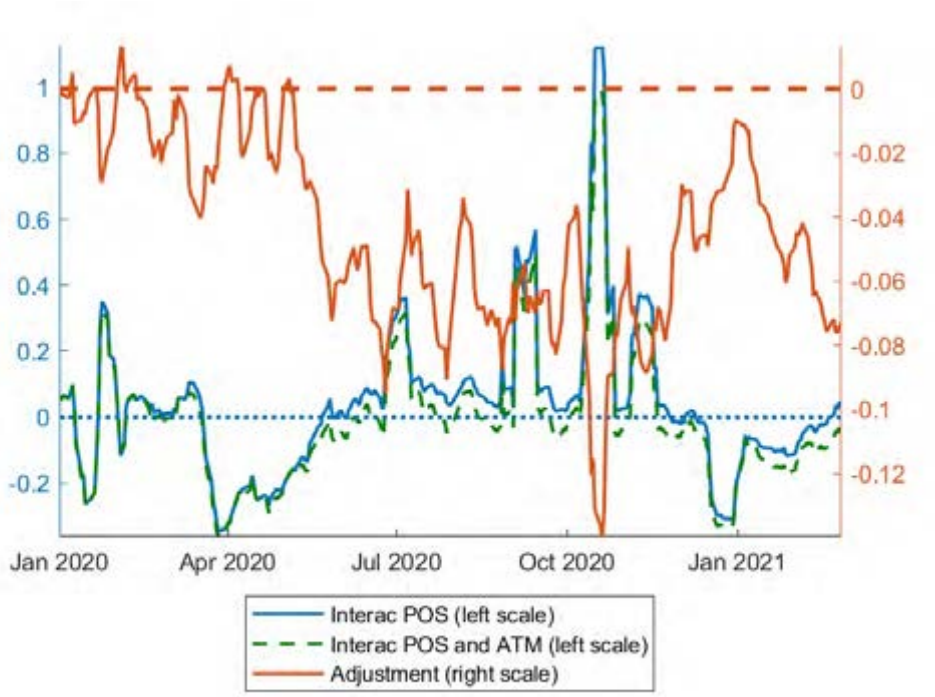
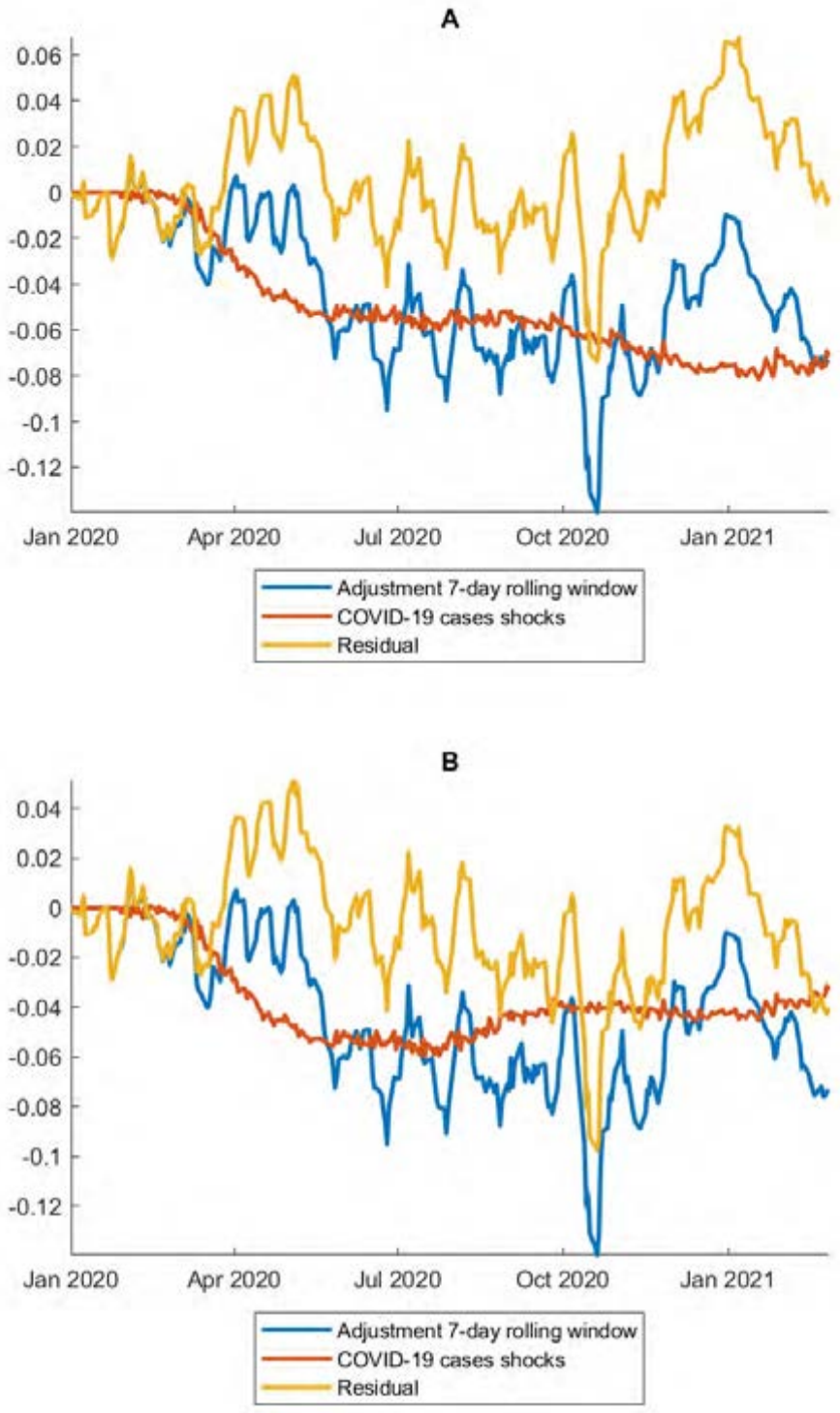


Figure 10: *Historical Nominal y/y Growth and Adjustments*



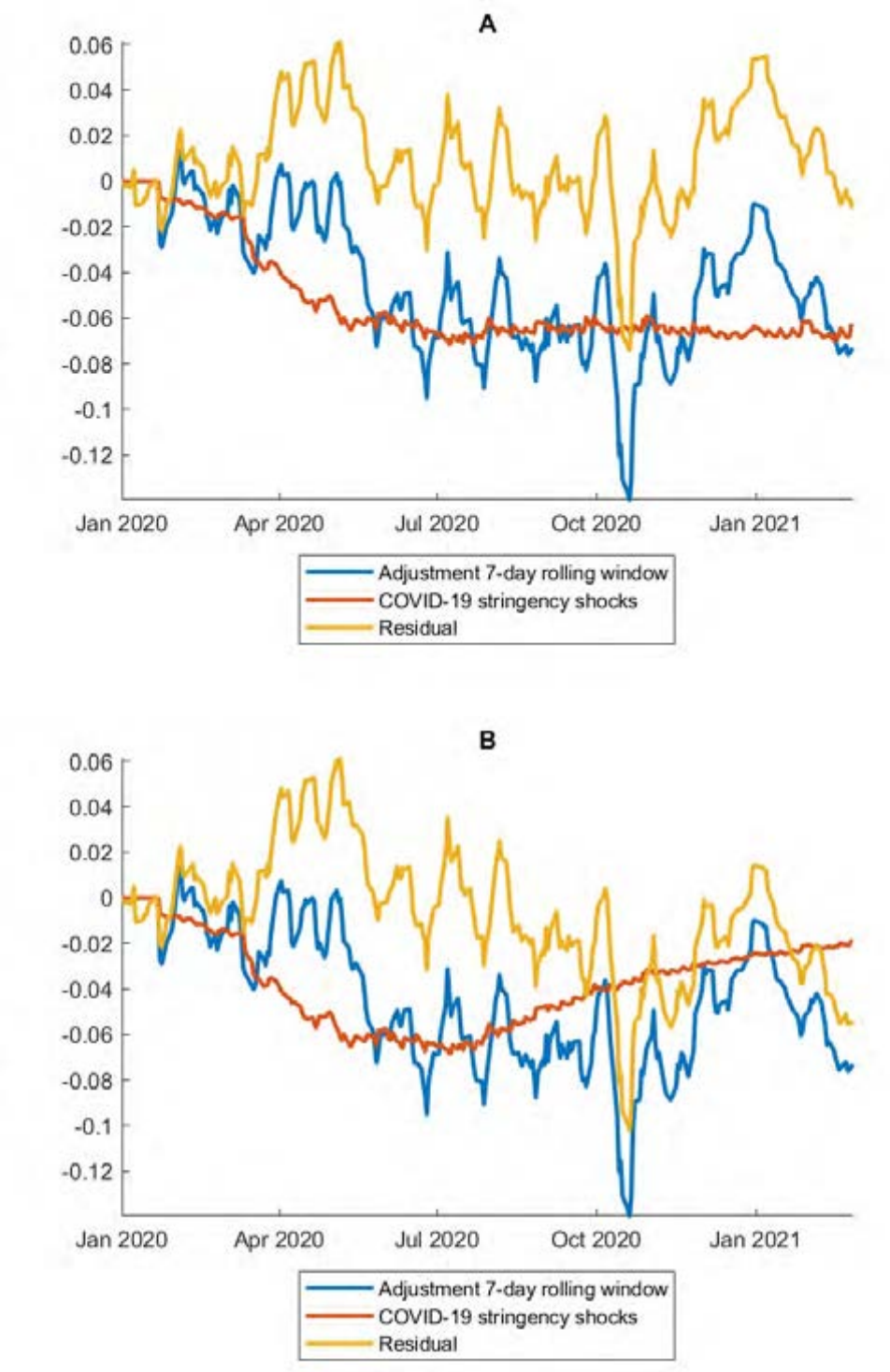
Note: The chart displays growth rates and adjustments, computed on a daily basis from the Interac data for POS debit and ATM transactions. Adjusted for the timing of New Year's Eve, New Year's Day, Canada Day, Thanksgiving Day, Remembrance Day, Christmas Eve, and Christmas Day.

Figure 11: *Cash Adjustment Decomposition for Newly Reported Cases*



Note: Based on Interac data. Y/Y growth rate adjustments, COVID-19 component and residual. Adjusted for the timing of New Year's Eve, New Year's Day, Canada Day, Thanksgiving Day, Remembrance Day, Christmas Eve, and Christmas Day.

Figure 12: *Cash Adjustment Decomposition for Stringency*



Note: Based on Interac data. Y/Y growth rate adjustments, COVID-19 component and residual. Adjusted for the timing of New Year's Eve, New Year's Day, Canada Day, Thanksgiving Day, Remembrance Day, Christmas Eve, and Christmas Day.

Did the Covid-19 Pandemic Change Payment Habits?

Evidence form High-frequency Transaction Data[1]

Tatjana Dahlhaus and Angelika Welte

Bank of Canada

15.7.2023

[1]The views expressed here are those of the authors and do not necessarily represent the views of the Bank of Canada.

Introduction

- Central banks use more and more alternative, high-frequency data available in real time.
- Transaction data (card providers, payment processors, settlement system) comes at high frequency and informs us about consumer behavior: How much do they spend? When and how often? How do they pay?

$\rightarrow$ Real-time high-frequency assessment of

1. cash usage and demand
2. economy (consumption)

This Paper

Research Question:

- How do shocks restricting cash use affect consumers' payment habits?

This Paper:

- Employs Canadian data on Interac and ACSS ABM and point-of-sale (POS) transactions.
- Obtains high-frequency and real-time measures of payment habits: share of cash transactions, average transactions values, share of on-us transactions.
- Uses the Covid-19 pandemic to understand how shocks to cash use can change payment habits.
- Shows how shifts in payment habits following shocks may influence the reliability of transaction data for economic monitoring.

Related Literature

Cash usage and demand following shocks:

- Greek government's imposition of daily ATM withdrawals: European Payments Council (2019), Esselink and van Gijsel (2017).
- Indian demonetization: Lahiri (2020), Chodorow-Reich et al. (2019), Bhandari (2020).
- Covid-19 pandemic: Chen et al. (2020), Ardizzi et al. (2020), etc.

Macroeconomic forecasting using high-frequency transaction data:

- Galbraith and Tkacz (2018), Chapman and Desai (2020, 2022) for Canada, Aastveit et al. (2020) for Norway, Cavallo (2020) for the US, and Aprigliano et al. (2019) for Italy.

Transactions data (Interac and ACSS)

Cash transactions:

- Withdrawals (and adjustments) at ATMs in Canada, volume & value.
- ATM can be on premise of a financial institution, off premise bank-owned, or white-label.

Card transactions

- Debit card payments (and refunds) at brick and mortar point-of-sales in Canada, volume & value.
- Chip&Pin, contactless transactions (tap).

Terminology

- Issuing bank: Bank of the cardholder,
- Acquiring Bank: Bank of the ATM owner, or merchant.
- Issuing bank = acquiring bank: the transaction is called "on-us."

Transactions data (Interac and ACSS)

Interac

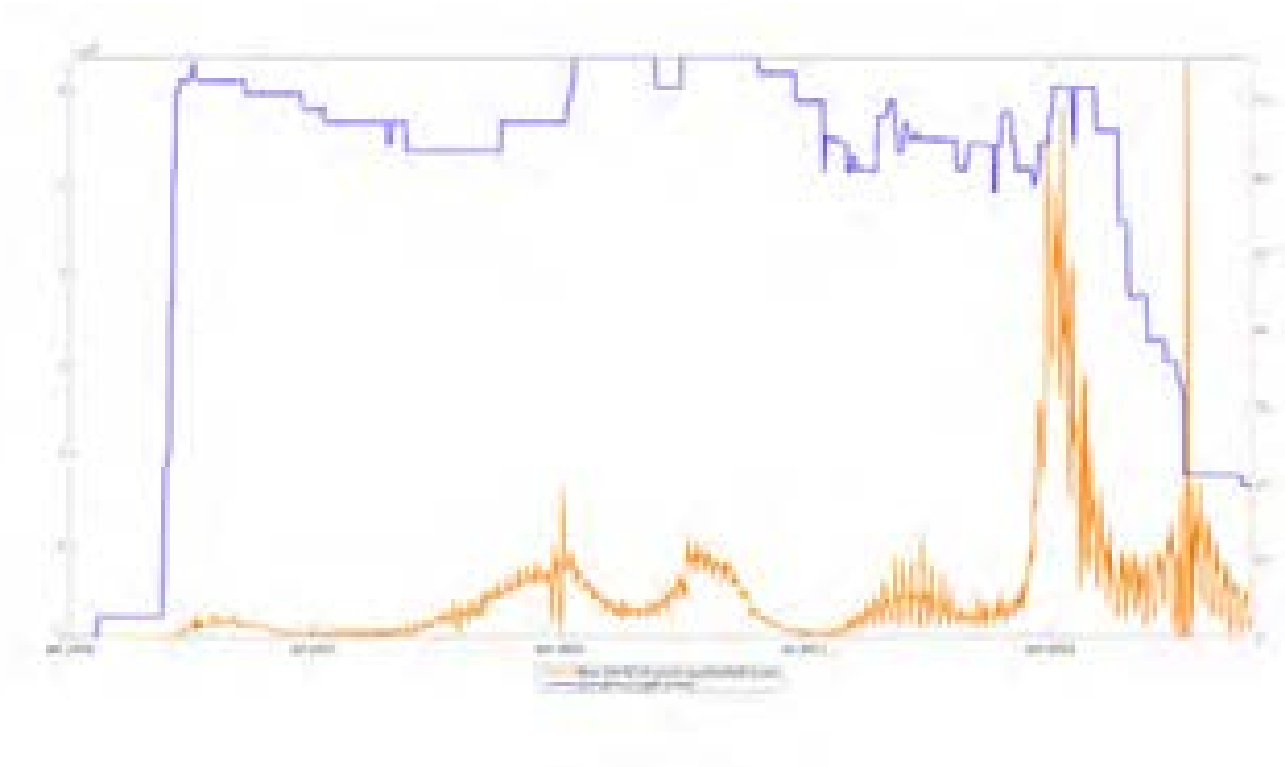
- Operates the main debit card network in Canada, Inter-Member Network (IMN).
- All ABM withdrawals and debit purchases, on-us and non-on-us, are processed.
- Data is available for every day incl. weekends and holidays.

ACSS

- ACSS operated by Payments Canada further clears the non-on-us transactions that are processed by the IMN.
- Data is available for business days and ACSS clears with a delay.
- ACSS was replaced by Lynx effective September 1, 2021

Shocks to Cash Use: COVID-19 Data

- Indicator variables for lock-down and reopening periods.
- Oxford Covid-19 Government stringency index combines nine indicators (closures, restrictions, controls, communication)
- Confirmed cases from Centre of Systems Science and Engineering(JHU).



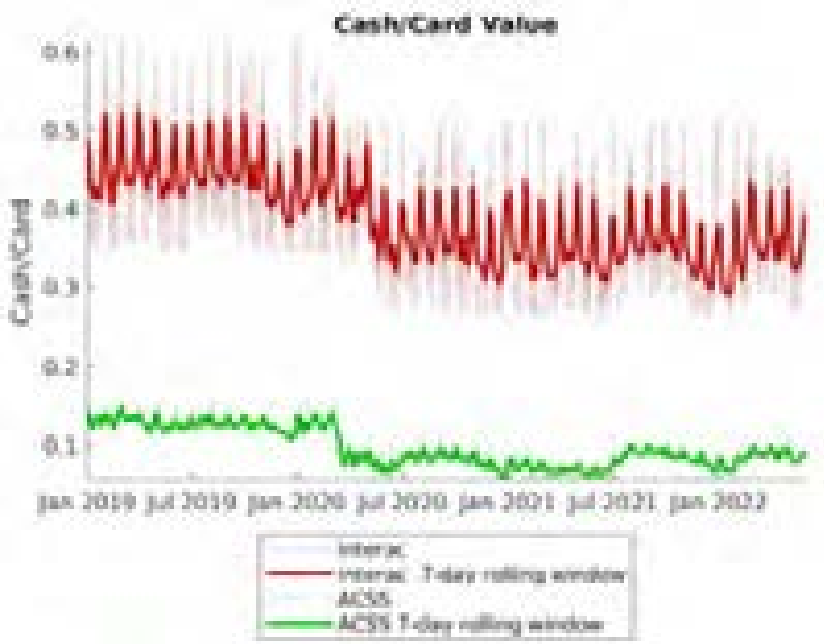
Real-Time Measures of Payment Habits

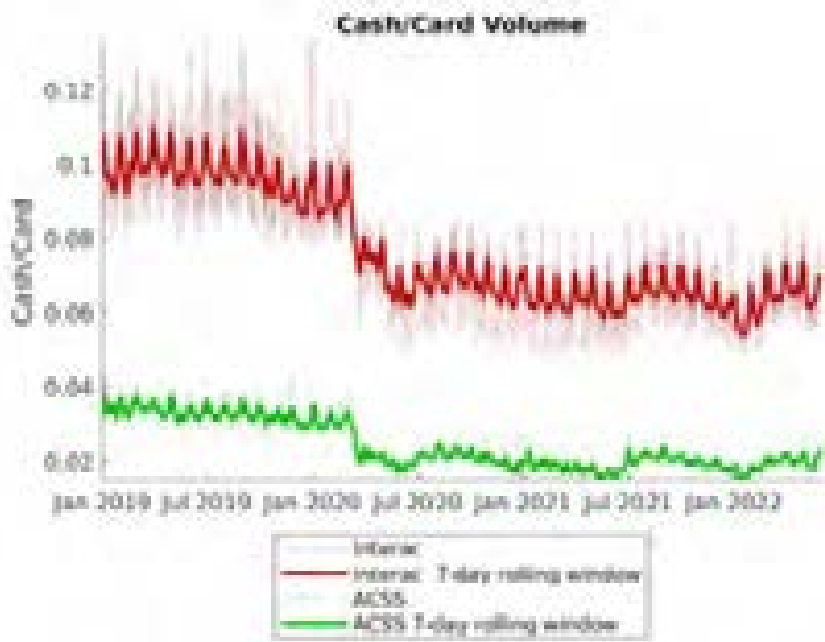
- Cash/card ratio: cash withdrawals divided by card transactions,
- Average transactions value: value of transactions divided by volume of transactions,
- ACSS/Interac: ACSS transactions (non-on-us) divided by Interac transactions (non-on-us and on-us).

Harmonization of Interac and ACSS:

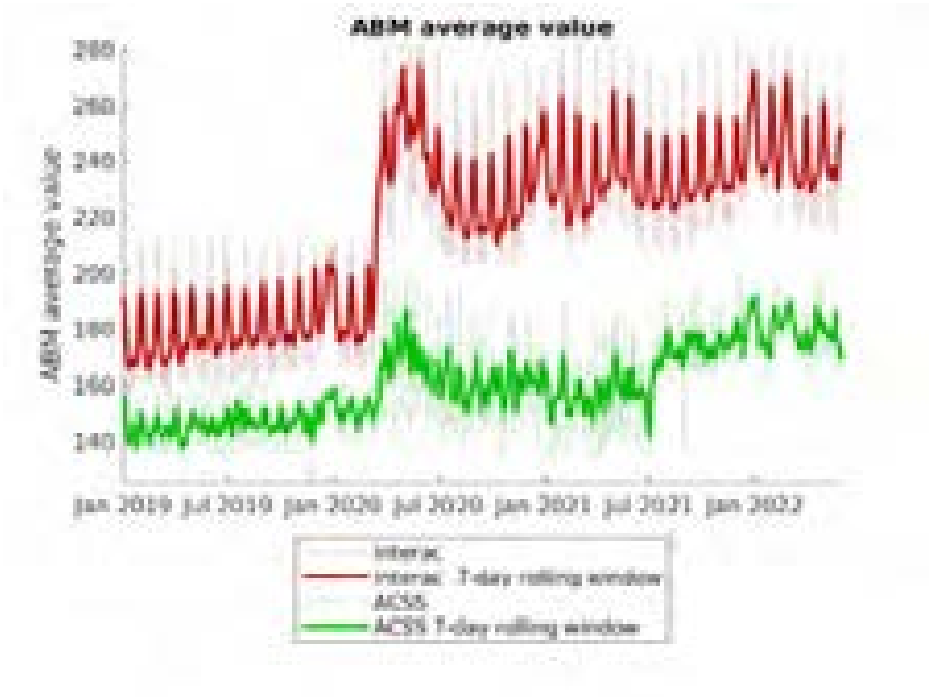
- Aggregate Interac transactions.
- Next business day if ACSS is open on authorization date; second next business data if ACSS is closed on authorization day.
- Example: Tuesday aggregates Saturday, Sunday, and Monday, etc.

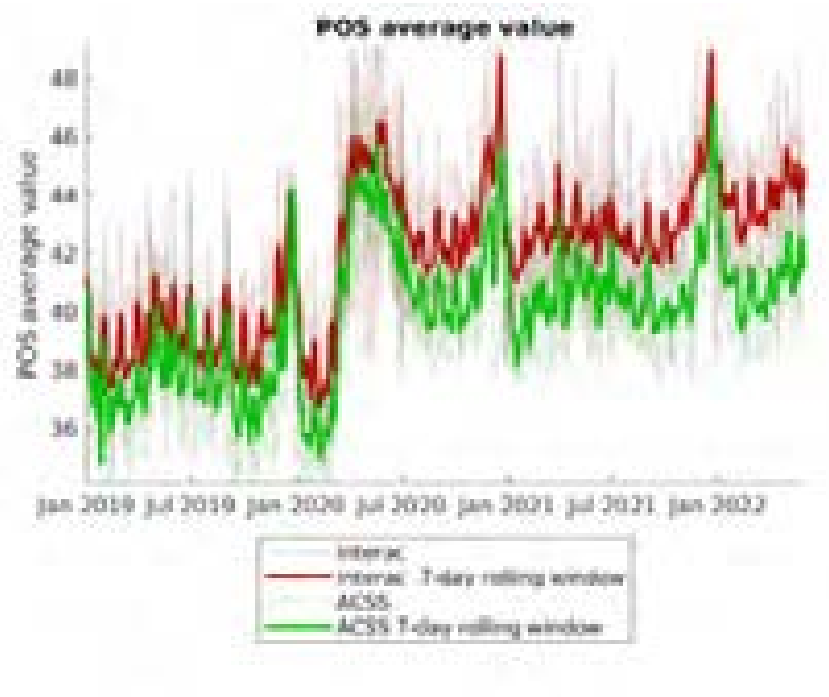
Cash/Card Ratios



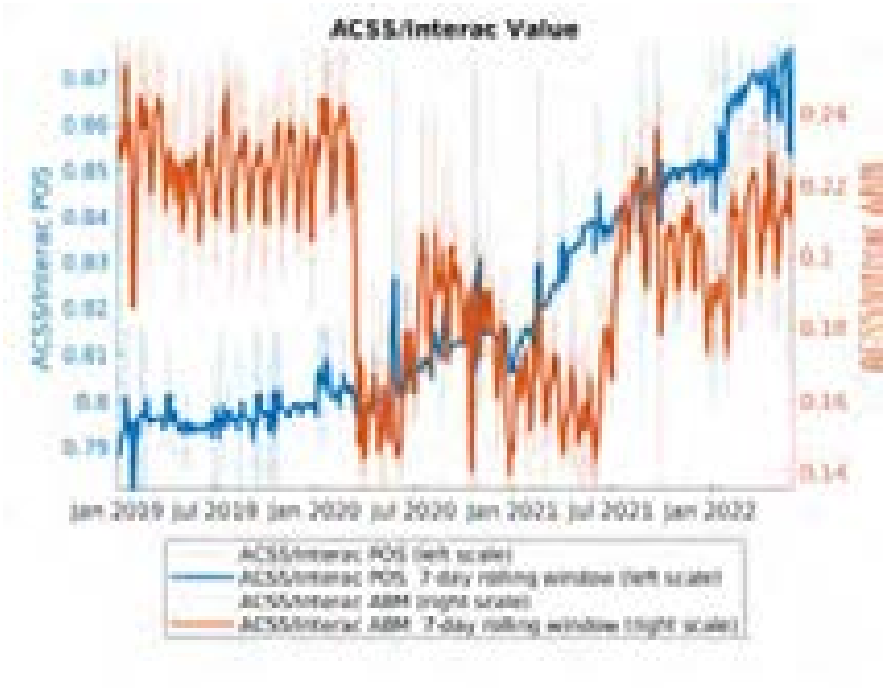


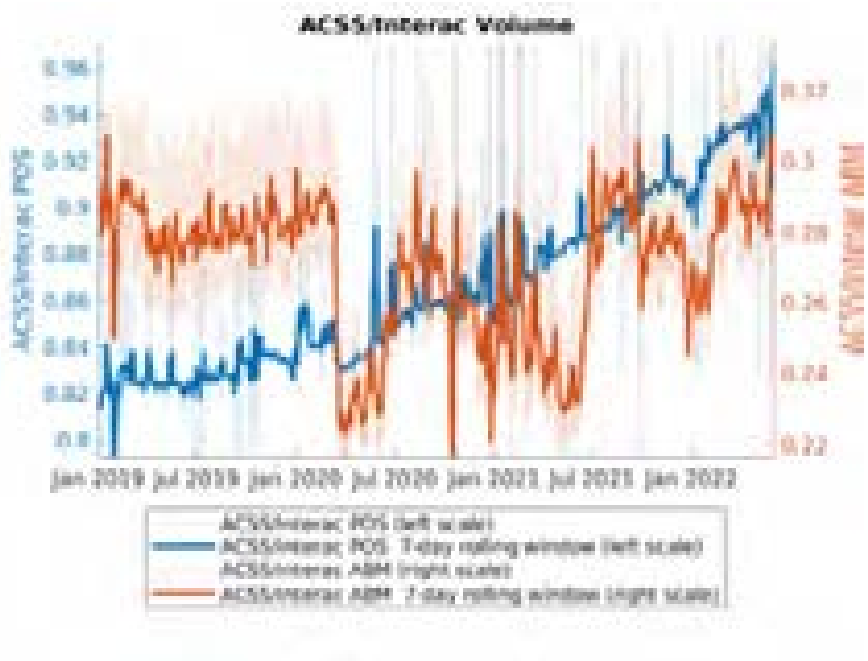
Average Transactions Values





ACSS/Interac Ratios for ABM and POS





Static Model: Dummy Regressions

- Regress measures of payment habits on pandemic-phase dummies.
- Dummies for pandemic phases: lockdown 1 (March-April 2020), reopening 1 (May-October 2020), lockdown 2 (November 2020-January 2021), reopening 2 (February 2021), lockdown 3 (April 2021), reopening 3 (June 2021).
- We control for seasonal patterns and trends by including day-of-week, week-of-month, and month-of-year dummies.
- Sample: January 1st, 2019 – May 31st, 2022.

Cash/Card Ratio Lower During Pandemic, Transaction Values Higher

		Cash-Card Value Ratio	Cash-Card Volume Ratio	log(ABM Average Value)	log(POS Average Value)
Lockdown 1	Interac	-0.036*	-0.022*	0.294*	0.126*
	ACSS	-0.041*	-0.011*	0.142*	0.120*
Reopening 1	Interac	-0.090*	-0.029*	0.238*	0.106*
	ACSS	-0.046*	-0.012*	0.093*	0.087*
Lockdown 2	Interac	-0.073*	-0.027*	0.247*	0.098*
	ACSS	-0.050*	-0.010*	-0.004	0.065*
Reopening 2	Interac	-0.114*	-0.034*	0.264*	0.122*
	ACSS	-0.065*	-0.014*	0.003	0.101*
Lockdown 3	Interac	-0.099*	-0.034*	0.254*	0.080*
	ACSS	-0.058*	-0.015*	0.069*	0.063*
Reopening 3	Interac	-0.094*	-0.032*	0.283**	0.113*
	ACSS	-0.043*	-0.013*	0.180*	0.087*
# of observations		849	849	849	849

Notes: * denotes significance at the 1% level based on standard errors robust against heteroskedasticity and autocorrelation (HAC) in the residuals. Daily sample from January 1st, 2019 –May 31th, 2022.

Non-On-Us Share Decreased for Cash During Pandemic, Increased for Cards

	ABM Value	ABM Volume	POS Value	POS Volume
Lockdown 1	-0.060*	-0.044*	-0.002	0.003
Reopening 1	-0.042*	-0.017	0.017	0.034*
Lockdown 2	-0.058*	0.014	0.019	0.052*
Reopening 2	-0.071*	-0.014	0.024	0.044*
Lockdown 3	-0.061*	-0.036*	0.038*	0.054*
Reopening 3	-0.023*	-0.002	0.060*	0.086*
# of observations	849	849	849	849

Notes: * denotes significance at the 1% level based on standard errors robust against heteroskedasticity and autocorrelation (HAC) in the residuals. Daily sample from January 1st, 2019–May 31st, 2022.

Dynamic Model: Local Projections

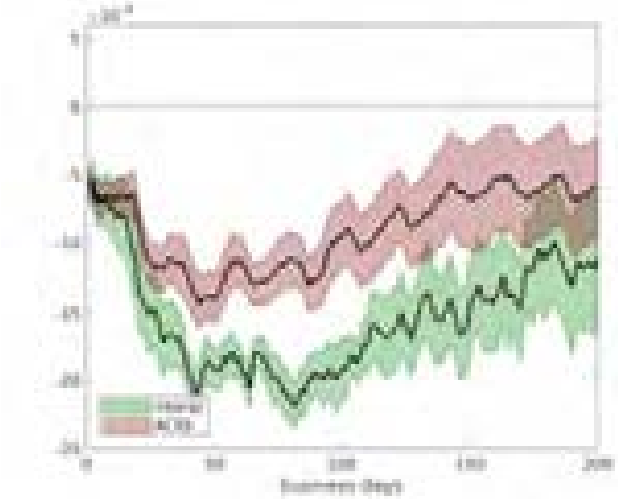
- To evaluate the dynamic effects of the pandemic and their persistence, we use local projections:

$$y_{t+h} = a_h + \sum_{i=1}^{p} b_{i,h} y_{t-1} + c_h x_t + e_{t+h}, \quad h = 1, ..., 100,$$

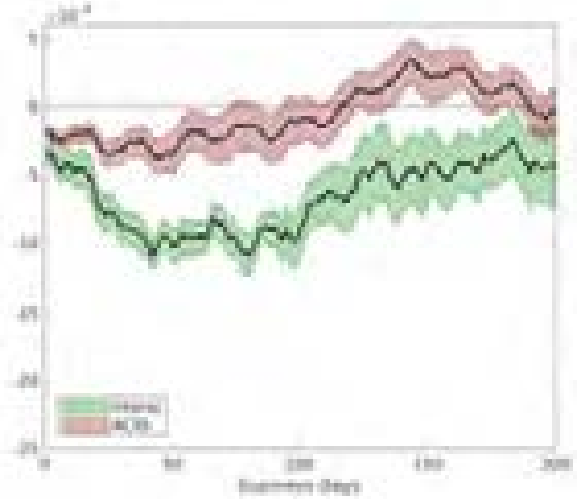
- y_{t+h} is a measure of payment habits.
- x_t is log(Covid-19 cases) or log(stringency index).
- $p = 10$.
- Includes day-of-week, day-of-month, week-of-month, and month-of-year dummies.
- To render the responses of ACSS and Interac variables comparable, we use Interac data harmonized to business-day frequency.

Responses of Cash/Card Ratios

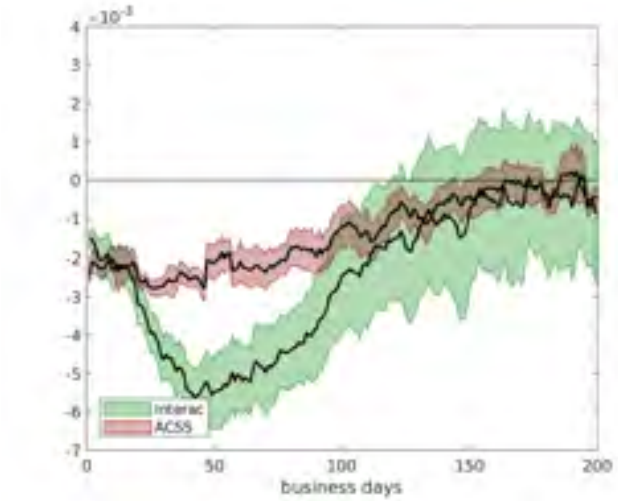
Cash/Card Value Ratio: Stringency



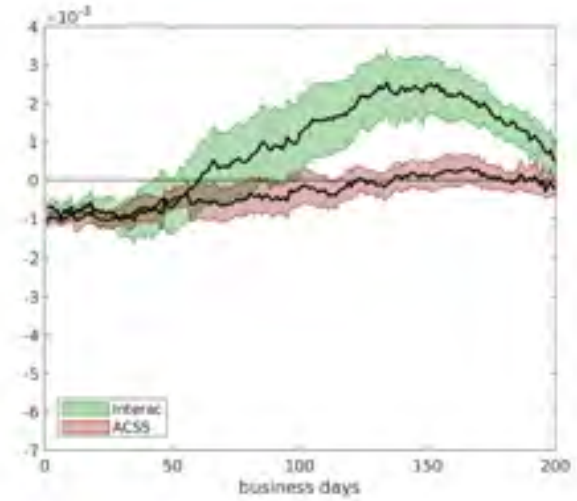
Cash/Card Value Ratio: Cases



Cash/Card Volume Ratio: Stringency

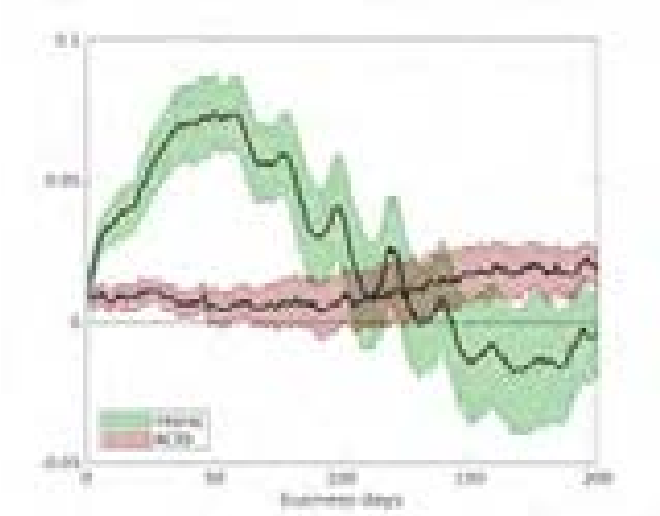


Cash/Card Volume Ratio:Cases

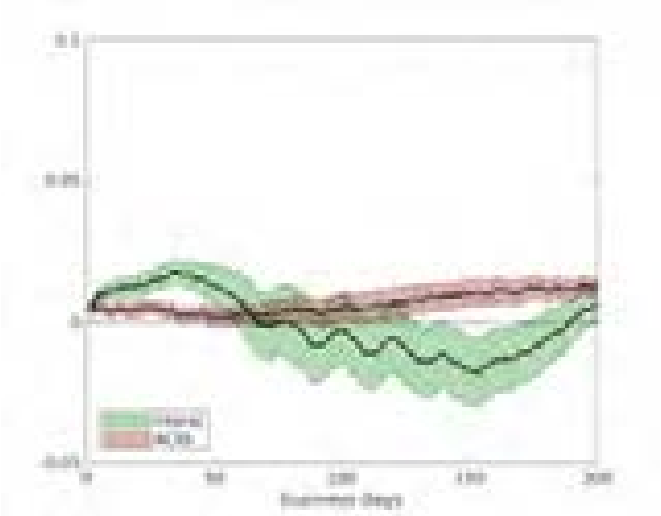


Responses of Average Transaction Values

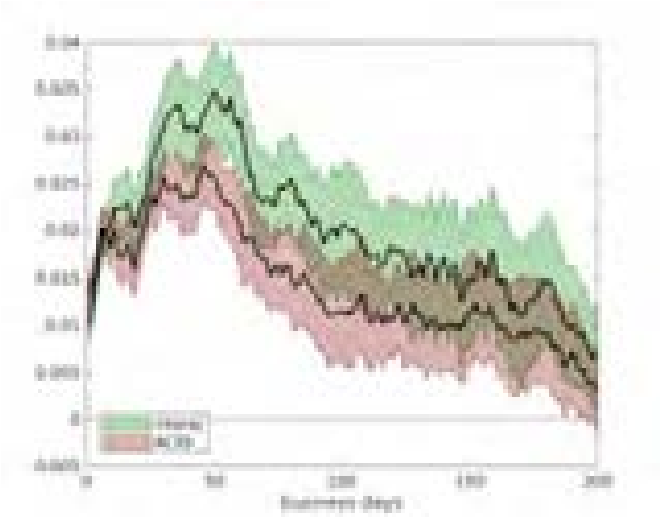
ABM Average Value: Stringency



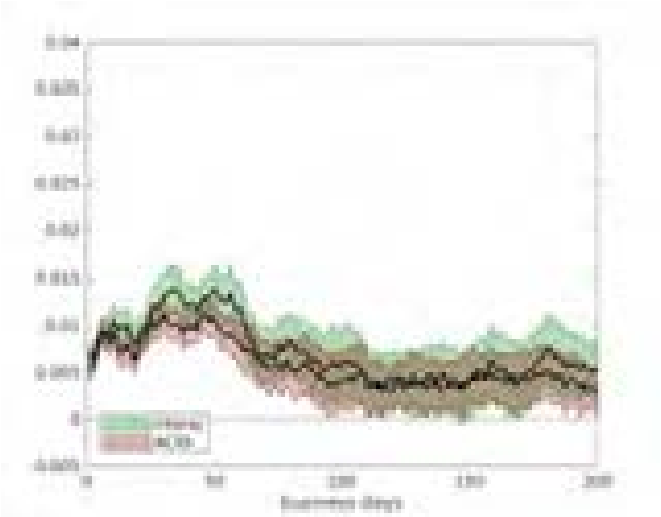
ABM Average Value: Cases



POS Average Value: Stringency

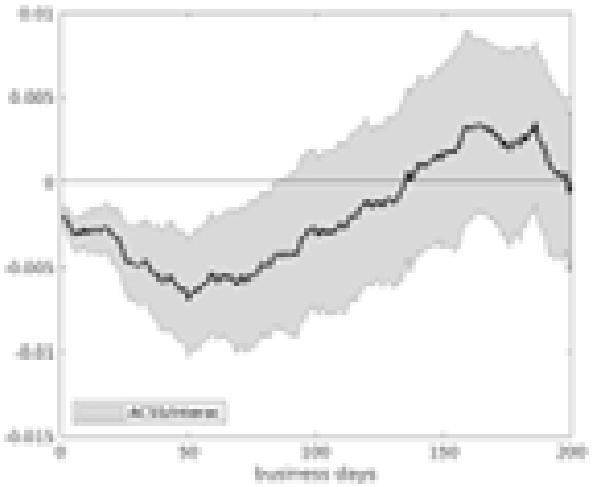


POS Average Value: Cases

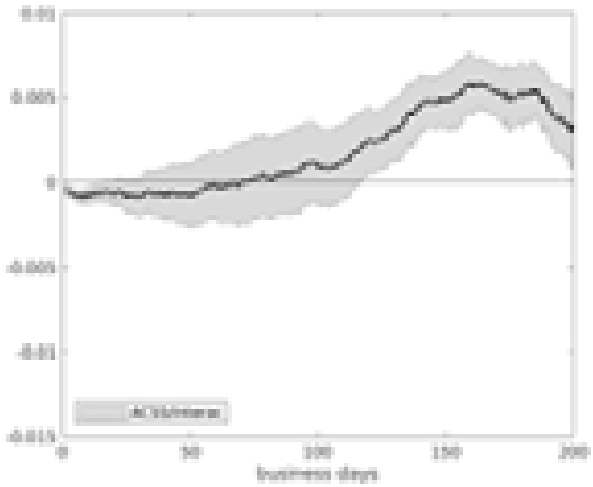


Responses of ACSS/Interac Ratios

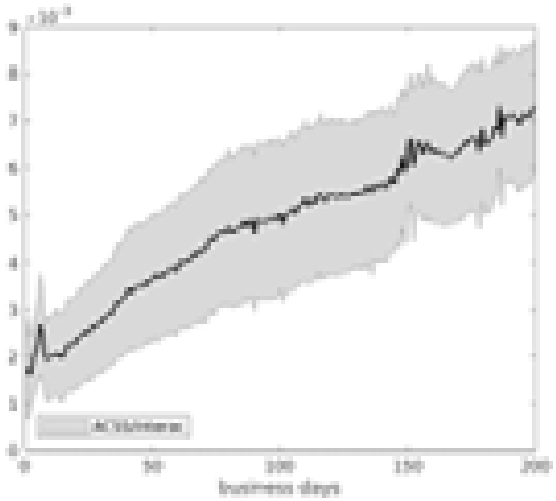
ACSS/Interac ABM Value: Stringency



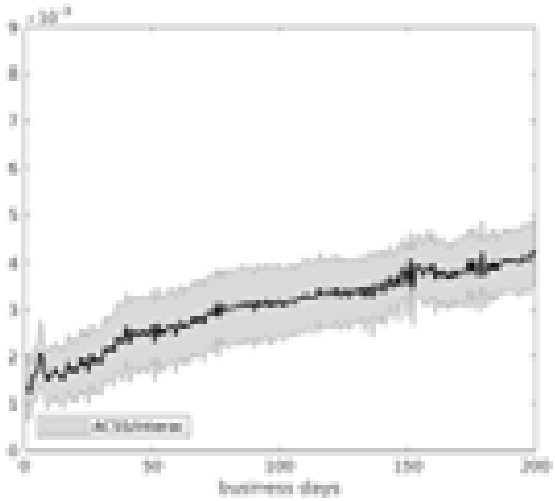
ACSS/Interac ABM Value: Cases



ACSS/Interac POS Value: Stringency



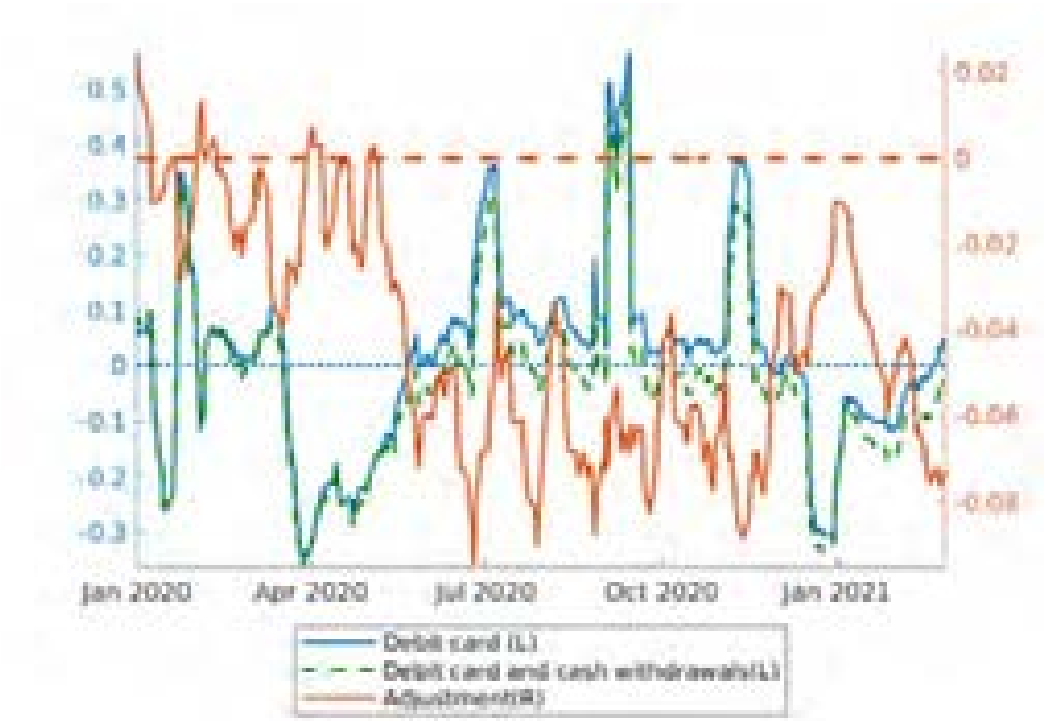
ACSS/Interac POS Value: Cases



Reliability of Card Transactions to Proxy Consumption

- Estimates of consumption based on card data alone may be biased if one abstracts from shifts in payment habits.
- Illustration: Two estimates of consumption growth:
  1. based on POS debit card transactions only
  2. based on card and cash transactions

Historical Daily y/y Growth of Consumption Estimates



Conclusion

Consumer behaviour adjusted in response to the pandemic:

- Less frequent trips for cash withdrawals and point-of-sale purchases reflected in higher transactions amounts.
- Decline in cash-card ratio may suggest reduced use of cash for point-of-sale transactions.
- Consumers made more withdrawals from ATMs that are linked to their financial institution (on-us transactions).

Shifts in payment habits matter for economic monitoring:

- Estimates of consumption growth based on card data alone are biased during the pandemic.
- Interac debit card transactions might have overstated consumption growth by up to 7 pp.

Questions? Comments?
Thank you! Merci!

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Classifying Job Postings into NAICS Codes[1]

Tatjana Dahlhaus, Reinhard Ellwanger, Gabriela Galassi and Pierre-Yves Yanni,
Bank of Canada

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Classifying Job Postings into NAICS Codes

Tatjana Dahlhaus, Reinhard Ellwanger,
Gabriela Galassi, and Pierre-Yves Yanni

Bank of Canada

BoC-IFC satellite seminar
July 15, 2023

The views expressed in this paper are those of the authors and no responsibility for them should be attributed to the Bank of Canada

Motivation

- Compared to JVWS data, Indeed job postings data are
  - more timely and frequent
  - more granular
- But they have their own sectoral classification, different from NAICS
- By classifying these job postings into NAICS, we can
  - track vacancies by industries at higher frequency and ahead of official data
  - backcast the missing official data
  - further exploit the granularity of the data
- Methodology to match datasets without common keys

Literature

- **Job posting-company matching:** Van Dijcke, Buckmann, Turrell, and Key (2023)

  Match firm-level balance sheets to analyze effect of leverage on employment; match on name only

- **Technology adoption during recessions:**

  - Hershbein and Kahn (2018), Jaimovich and Siu (2020)
  - Soh, Oikonomu, Pizzinelli, Shibata, and Tavares (2023), Pierri and Timmer (2020), Bellatin, Chu and Galassi (2023)

  Do not consider different types of firms; implications for how permanent are changes.

Roadmap

Data

Matching to NAICS

- Methodology
- Validation

Applications

- Backcasting missing data
- Digital occupations and tech firms

Takeaways

Data

Indeed job postings: largest job board in Canada; posting from companies and fetched from other websites (2018-present)

Advan: visits to points-of-interests from cellphone location data (2019-present)

Indeed job postings
job key
company name
city
province
first date
last date

Advan
company name
city
province
NAICS (4-6 digits)

Matching to NAICS

1. Preprocess data

2. Transform company names into numerical vectors

3. Match company names at 3 geographical levels: city, province, national

4. Pick cutoff on matching scores

Matching to NAICS

Preprocess | Embeddings | Matching | Cutoff

- For company names and cities, remove
  - capital letters
  - accents
  - spaces and special characters
- Remove stopwords from company names:
  inc, limited, group, . . .
- Map cities in Indeed to cities in Advan

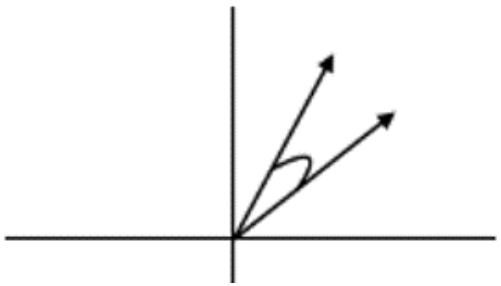
Matching to NAICS

- We create embeddings for company names using tf-idf:
  tf-idf = term frequency (tf) $\times$ inverse document freq. (idf)
  idf = log(number of documents / document frequency)
- In our case:
  - *document* = company name
  - *term* = trigram
- Example: *pepsico, pepsicola and pepper*

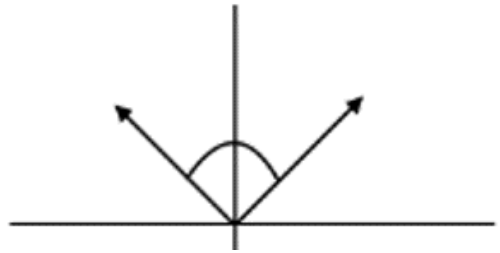
	Term (trigram) frequency									
	pep	eps	psi	sic	ico	col	ola	epp	ppe	per
pepsico	1	1	1	1	1	0	0	0	0	0
pepsicola	1	1	1	1	1	1	1	0	0	0
pepper	1	0	0	0	0	0	0	1	1	1
Document (company name) frequency	3	2	2	2	2	1	1	1	1	1

Matching to NAICS

- Cosine similarity (2 dimensions):



similar trigrams:
$\cos \to 1$



different trigrams:
$\cos \to 0$

- In our example:
  - vectors of dimension 10
  - cos(pepsico, pepsicola) = 0.76, cos(pepsico, pepper) = 0.12
- Our data:
  - 670k company names in Indeed and 640k in Advan
  - vectors of dimension 26k

Matching to NAICS

Preprocess | Embeddings | **Matching** | Cutoff

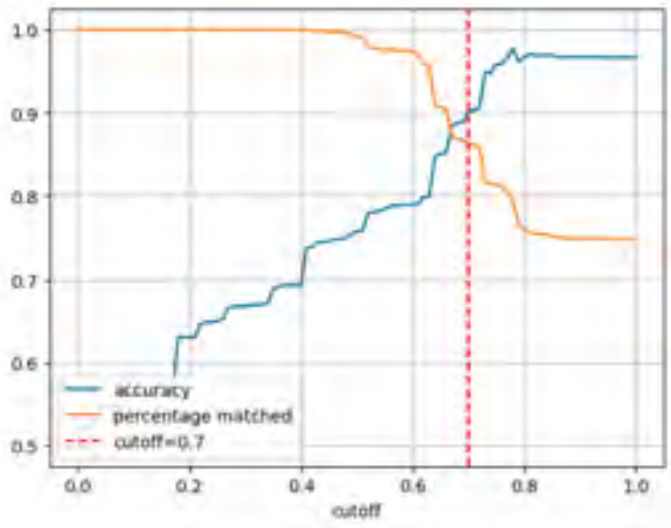
- Find best matches at city, province, national levels based on cosine similarity

- For given cutoffs on cosine similarity,
  1. keep matches above cutoff at **city** level
  2. among companies unmatched at city level,
     keep matches above cutoff at **province** level
  3. among companies unmatched at city and province level,
     keep matches above cutoff at **national** level

Matching to NAICS

Preprocess | Embeddings | Matching | Cutoff

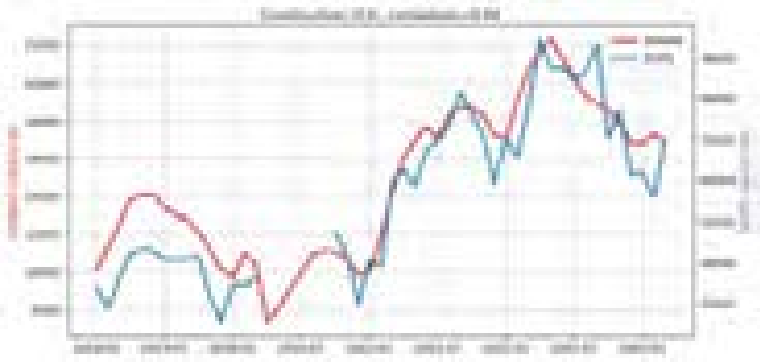
- Vary cutoff on labelled sample of firms matched:

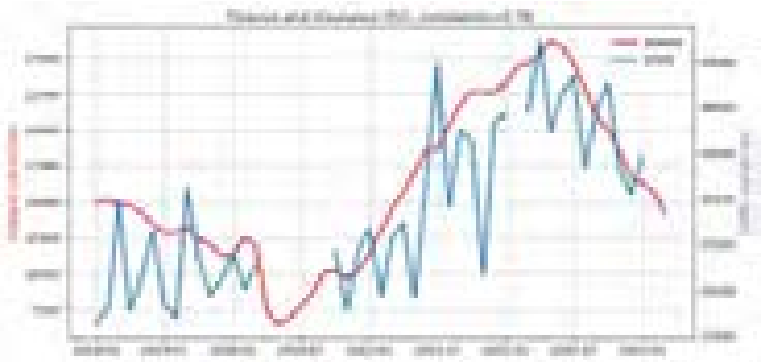


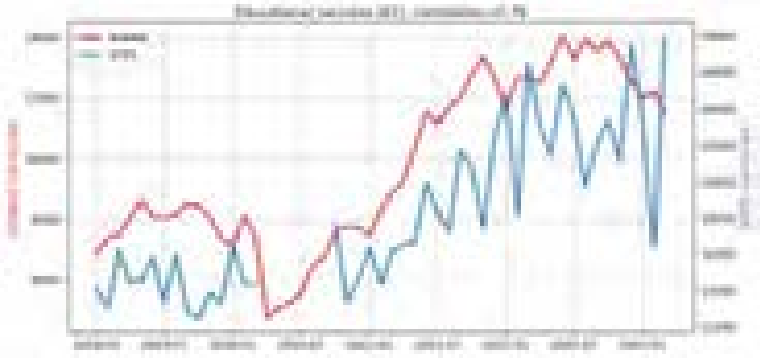
- Cutoff = 0.7 chosen to get 90% weighted accuracy
- Matching results:

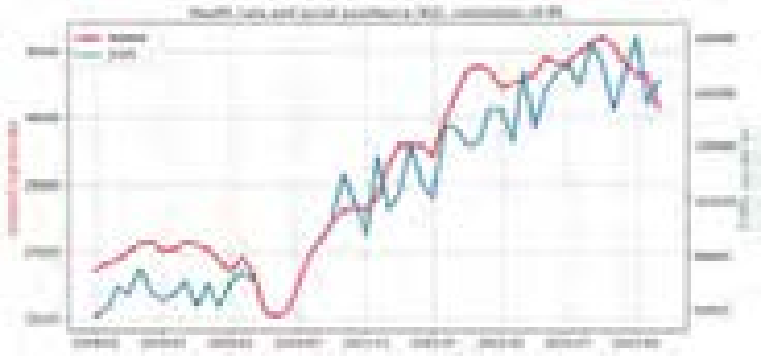
Indeed company name	Advan company name	City	Province	Score	NAICS
Pizza Munno	Pizza Munno	Oshawa	ON	1.0	722511 Full-service restaurants
David's Tea	Les Thés David's Tea	Sherbrooke	QC	0.72	722515 Snack and Nonalcoholic Beverage Bars
Canadian Pacific	Canadian Tire	Cochrane	AB	0.42	unmatched

Matching to NAICS: Validation with JVWS Data







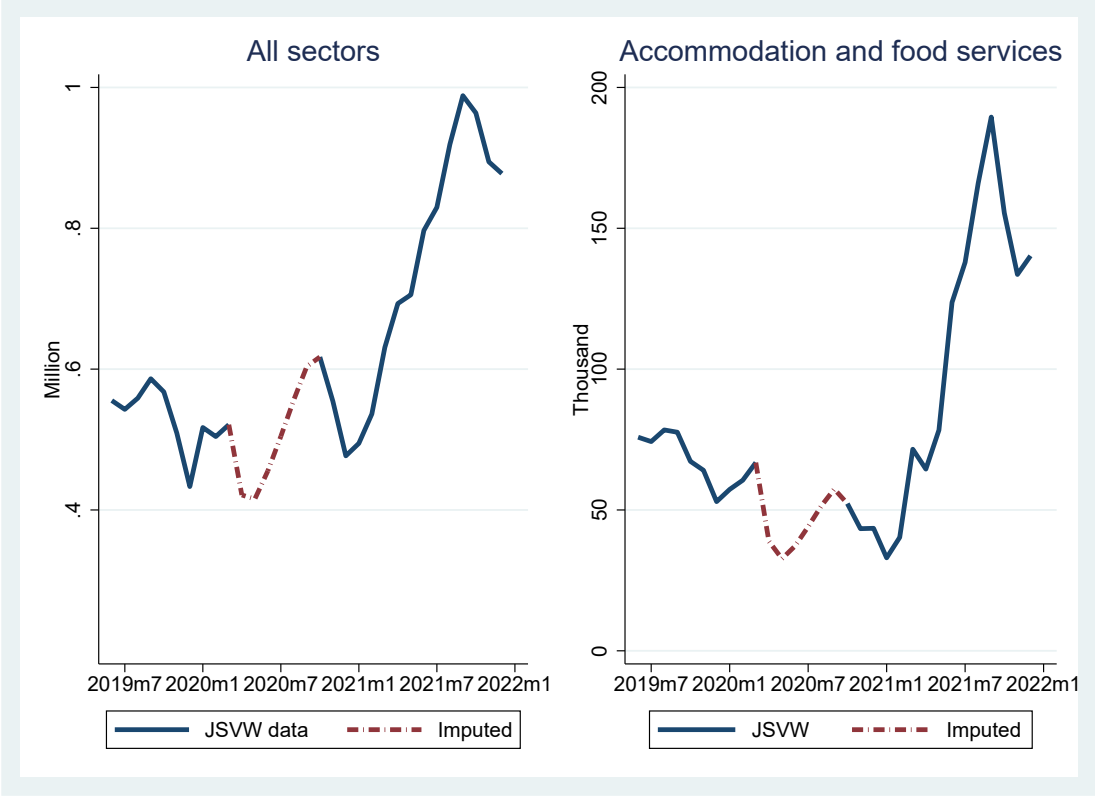


Application 1: Backcasting Missing Data

- JVWS did not collect data from April to September 2020
- Estimate relationship:

$$\Delta v_{t,j}^{jvws} = \alpha + \beta_0 \Delta v_{t-1,j}^{ind} + \beta_1 \Delta v_{t,j}^{ind} + \beta_2 \Delta v_{t+1,j}^{ind} + \varepsilon_{t,j}.$$

- Use the estimated $\widehat{\Delta v_{t,j}^{jvws}}$ to fill in the missing values:



Application 2: Digital occupations and tech firms

- Bellatin, Chu and Galassi (2023) study digital job creation during the COVID-19 pandemic

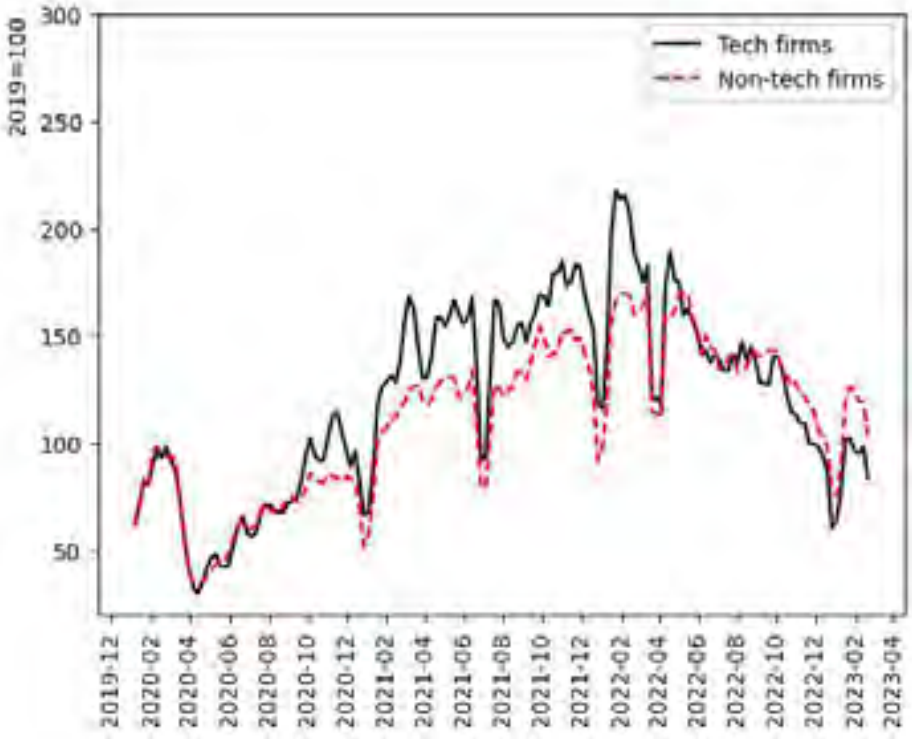
- Adding tech/non-tech firms dimension:

  1. what type of firms drive the creation of digital vacancies?
  2. digital production vs digital adoption:

	Digital Jobs	**Non-Digital Jobs**
Tech Firms	Digital production	Digital-adjacent
Non-Tech Firms	Digital adoption	

Application 2: Digital Jobs and Tech Firms

Digital job postings



Non-digital job postings



- disproportionately more digital jobs vacancies in tech industry during pandemic
- relative to non-tech industry, post-pandemic vacancies decreased mainly in non-digital occupations

Takeaways

- Fuzzy matching to classify job postings into NAICS codes
  - high accuracy in matching
  - tracks well vacancies stocks by industry from official statistics
- Newly created dataset
  - complements official data
  - allows to analyze digital production vs digital adoption
- Next steps:
  - obtain Business Register data
  - reduce lag in job postings stock

Representativeness Indeed Job Postings

name	NAICS	cutoffs	fractions	corr.
Utilities	22	0.90, 0.80, 1.00	0.04	0.617
Construction	23	0.85, 0.65, 0.60	0.38	0.96
Manufacturing	31-33	0.90, 0.80, 0.95	0.06	0.951
Wholesale trade	41/42	0.70, 0.70, 0.60	0.22	0.9
Retail trade	44-45	0.90, 0.80, 0.85	0.67	0.877
Transportation and warehousing	48-49	0.70, 0.80, 1.00	0.15	0.86
Information and cultural industries	51	0.80, 1.00, 0.60	0.31	0.858
Finance and insurance	52	0.95, 0.60, 0.75	0.57	0.791
Real estate and rental and leasing	53	0.80, 0.90, 0.60	0.84	0.803
Prof, scientific and technical services	54	0.70, 0.60, 0.70	0.22	0.938
Management of companies and enterprises	55	0.60, 0.90, 0.70	1.42	0.762
Administrative and support	56	0.85, 0.70, 0.60	0.09	0.811
Educational services	61	0.85, 0.75, 0.60	0.64	0.841
Health care and social assistance	62	1.00, 1.00, 1.00	0.16	0.96
Arts	71	0.70, 0.60, 1.00	0.68	0.815
Accommodation and food services	72	0.70, 0.70, 0.60	0.32	0.862
Other services (except public administration)	81	0.60, 0.60, 0.60	0.61	0.939
Public administration	91/92	0.95, 0.65, 0.60	0.19	0.794

Representativeness Indeed Job Postings

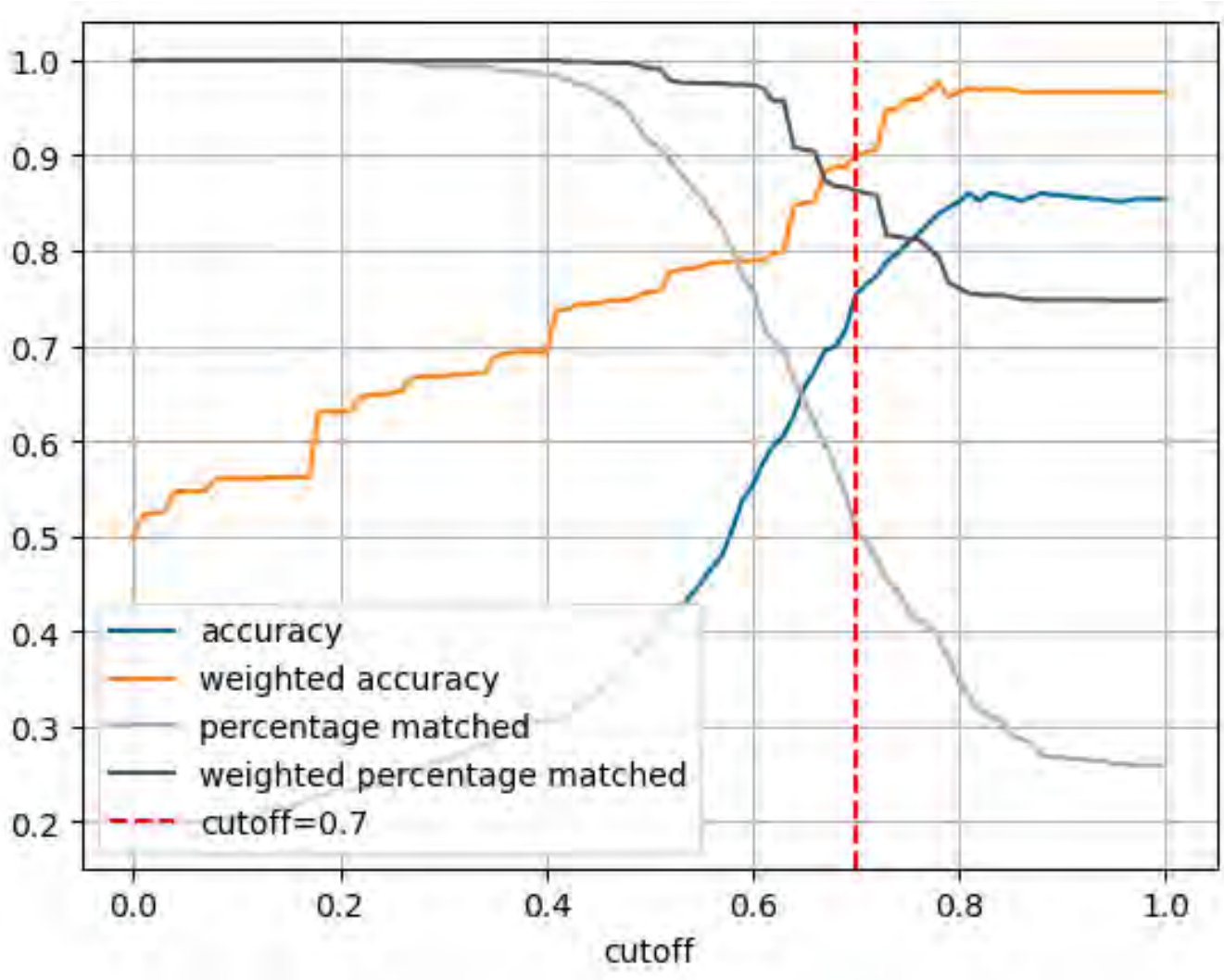
Ratio companies in Advan / Employer businesses from StatCan:

NAICS	11	21	22	23	31-33	41/42	44-45	48-49	51
ratio Advan / StatCan	0	0	0.31	0.27	0.31	0.23	1.15	1.39	0.37

NAICS	52	53	54	55	56	71	72	81
ratio Advan / StatCan	0.54	0.33	0.12	0.04	0.15	1.25	0.97	0.78

source: Business Dynamics measures, Statistiques Canada

Unweighted vs Weighted Accuracy



Application 1: Backcasting Missing Data

- JVWS did not collect data in Q2 and Q3 2020
- Solution:
  - estimate relationship:

$$\Delta v_{t,j}^{jvws} = \alpha + \beta_0 \Delta v_{t-1,j}^{ind} + \beta_1 \Delta v_{t,j}^{ind} + \beta_2 \Delta v_{t+1,j}^{ind} + \varepsilon_{t,j}.$$

  - estimate difference in trend during the period with missing observations

$$\Delta T = (v_{Oct20,j}^{jvws} - v_{Mar20,j}^{jvws}) - \sum_{i=1}^{7} \widehat{\Delta v_{Mar20+i,j}^{jvws}}$$

  - use the estimated $\widehat{\Delta v_{t,j}^{jvws}}$ to fill in the missing values k periods after March 2020:

$$\hat{v}_{Mar20+k,j} = v_{Mar20,j} + \sum_{i=1}^{k} \widehat{\Delta v_{Mar20+i,j}^{jvws}} + k \cdot \Delta T/7,$$

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Narrative monetary policy uncertainty[1]

Charles Martineau, Zissis Poulos, Yuntao Wu, Cameron Thompson, Jun Yuan and John Hull, Joseph L. Rotman School of Management, University of Toronto,

and

Maryam Haghighi,
Bank of Canada

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Narrative Monetary Policy Uncertainty

Charles Martineau* Zissis Poulos* Yuntao Wu* Cameron Thompson*
Maryam Haghighi† Jun Yuan* John Hull*

September 1, 2023

Abstract

We develop a novel method to capture monetary policy uncertainty reflected in the media. The proposed uncertainty measure is constructed using advanced natural language processing techniques and can be computed daily, even when news coverage is low. We find an average increase in uncertainty on days leading to scheduled central bank announcements and a significant decrease on days after announcements. However, following negative news on announcement dates, uncertainty increases. At lower frequency, we show that the uncertainty measure responds to changes in macroeconomic fundamentals such as unemployment, housing prices, and inflation.

Keywords: monetary policy uncertainty, risk premia, central bank communication, machine learning, natural language processing

*Joseph L. Rotman School of Management, University of Toronto. We thank Dr. Andreas Veneris from the Department of Electrical and Computer Engineering, University of Toronto, for his valuable comments.

†Bank of Canada. Staff research is work in-progress and produced independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

1 Introduction

Monetary policy announcements are instrumental in shaping the expectations of economic agents and maintaining the stability of the financial system. Policy changes' influence on equity prices and long-term real interest rates is well-studied and empirically validated.[1] A more recent thread of studies goes beyond policy changes and focuses on the role of policy uncertainty and how it is transmitted to financial markets. However, identifying this effect remains challenging considering that there is no one-size-fits-all method to quantify monetary policy uncertainty.

One approach to measuring uncertainty is by eliciting its impact on financial markets. Bauer et al. [2021] do so by looking into the prices of Eurodollar options and futures in the cycle between FOMC announcements. The uncertainty measure proposed is related to the implied conditional volatility of future LIBOR, which is considered a benchmark short-term interest rate. An alternative approach involves analyzing the narrative around monetary policy found in newspapers and the general media. The assumption is that if the readership cares about policy changes and forward guidance, and if uncertainty about those is significant, then the level of uncertainty will be reflected in the tone used by the media to describe the state of the economy. Along these lines, several text-based policy uncertainty measures have recently been introduced. The seminal work of Baker et al. [2016] gave rise to the widely-used Economic Policy Uncertainty (EPU) and Monetary Policy Uncertainty (MPU) indices, while Husted et al. [2020] constructed an alternative monetary policy uncertainty index (referred to as Husted-Rogers-Sun MPU, or HRS-MPU) based on the same methodology. These measures mostly rely on the occurrence of particular keywords in articles. The methodology lacks the fine granularity required to capture nuance in language; it can only identify the presence of uncertainty in an article but not its level. Being keyword-based also means that a large supply of articles is necessary in order to aggregate data meaningfully. This limits the frequency with which uncertainty can be measured with confidence. Moreover, it is not yet established in the literature whether text-based and market-based policy uncertainties are directionally aligned, especially considering that several macroeconomic variables may influence market-based proxies of uncertainty.

In this paper, we propose a novel soft information measure that quantifies the level of language uncertainty for any topic, ranging from news about monetary policy to news about

[1] See, for example, Kuttner [2001], Bernanke and Kuttner [2005], Swanson [2021], Savor and Wilson [2013], Lucca and Moench [2015], Nakamura and Steinsson [2018], Cieslak et al. [2019], Boguth et al. [2019], among others.

the housing market and the energy sector.[2] We apply a state-of-the-art language model[3] to the textual content of news articles and obtain article-level uncertainty scores that can be aggregated within various time periods. The model takes language context into account, making it possible to obtain trustworthy scores even from a handful of articles (our methodology does not require but can benefit from a large article count). By forming a measure of the level, rather than only the presence, of uncertainty in a given news article, more informative results can be obtained with fewer articles. This is important when analyzing uncertainty toward topics with low coverage and becomes essential in event studies concerned with how financial markets and macroeconomic fundamentals respond to monetary policy announcements.[4]

Our method, like the EPU and MPU, relies on narratives and text analysis to address a wide range of uncertainties in the economy, not just limited to market-based monetary policy uncertainty (such as using Eurodollar option volatility as a proxy). This becomes particularly relevant as many central banks in developed countries have resorted to two alternative measures after exhausting conventional interest rate cuts. These measures involve implementing large-scale quantitative easing and providing explicit forward guidance, both of which have been frequently discussed in the news since the global financial crisis of 2008/2009.

We focus on monetary policy and show that the proposed uncertainty measure, which we refer to as Narrative Monetary Policy Uncertainty (NMPU), positively correlates with the EPU and MPU indices at daily and monthly frequencies. Considering the US and Canadian economies separately, we find that FOMC meetings and Bank of Canada interest rate announcements lower media uncertainty on average, while uncertainty builds up in the days preceding the announcements. In general, equity returns are negatively correlated with the level of uncertainty after announcements and with the change in uncertainty between the days leading to the announcement and the days following. However, bad news around monetary policy, which is proxied by negative equity returns on the announcement day, correlate with an increase in uncertainty. In Canada, we expose a difference between newswires in terms of how uncertainty around monetary policy is reported. Finally, we show that the proposed uncertainty measure responds to macroeconomic fundamentals in the US, such as the unemployment rate, Consumer Price Index, the housing index (Case-Shiller), and the federal funds effective rate. However, the sensitivity depends on the source used to measure uncertainty since periodicals in the US focus on macroeconomic issues in varying degrees.

[2]Keyword search can be used to obtain articles from a specific database, but the proposed uncertainty measure does not rely on the keywords used and is universally applicable to any topic.

[3]The model is based on the transformer architecture proposed by Devlin et al. [2019a]

[4]Applications of contextual uncertainty include corporate-level measurements and analysis of uncertainty in countries with lower coverage (*e.g.*, Canada).

2 NMPU based on Language Models

In our methodology we interpret language uncertainty as any instance in which writers intentionally convey the fact that they do not have all the information on a matter, because the situation is inherently unpredictable, or because they're speculating about hypothetical future scenarios. Uncertainty can be expressed in many ways, including questions, tentative statements, expressions of doubt, or with the use of "hedging" words. It is evident that to quantify the level of uncertainty in an article it does not suffice to simply identify whether the word "uncertain" or derivatives and synonyms thereof are present. This typical approach ignores the space of possibilities when it comes to expressing uncertainty. The problem calls for symbolic or data-driven solutions. We focus on the latter and particularly on natural language processing (NLP) techniques that have now reached maturity when it comes to semantic analysis and language understanding.

The field of NLP recently experienced a major breakthrough stemming from a new class of language models referred to as *transformers*, and specifically an architecture called Bidirectional Encoder Representations from Transformers (BERT). To this day transformer models are considered the state-of-the-art NLP technique, having superseded all other language models.[5] BERT models involve a process, referred to as pre-training, where deep neural networks are trained using unlabeled text in order to predict missing words in sentences based on the "right" and "left" context of each missing word (Devlin et al. [2019b]). The premise is that pre-trained models must learn the basic structure, rules and organization of language to be able to fulfill the prediction task accurately. Pre-trained models can then be fine-tuned (*e.g.*, by supervised learning) with only minor modifications to produce cutting-edge models for various downstream tasks. The ability to take a publicly available pre-trained BERT model and fine-tune it for a downstream task such as uncertainty analysis is broadly known as *transfer learning*. Fine-tuning only requires a small set of examples (*e.g.,* examples of words/sentences that express uncertainty) in order to achieve robust performance, because the fine-tuned model does not need to re-learn language properties

2.1 Data

In the ProQuest Database we search for articles related to monetary policy in major newspapers for both the US and Canada. The keywords used to identify relavent articles are listed in Table A1. For the US, we search for news articles in "The Wall Street Journal" (WSJ) and "New York Times" (NYT) data sources. We sample from newspapers only and exclude other source types such as magazines under the same publication. The sample period begins

[5]Indicatively, GPT models, such as OpenAI's ChatGPT are also based on the transformer paradigm.

on Jan 2, 1984 and ends on Mar 22, 2023 for both publications. This gives a total of 70,356 articles from WSJ and 21,101 articles from NYT. For Canada, we search for news articles in the National Post. The sample period begins on Oct 27, 1998 and ends on Mar 22, 2023. This gives a total of 10,758 articles. We remove all special characters, numbers and multi-spaces from each news article, lemmatize all the words and split the article into sentences to reduce the dimensionality of the problem.

2.2 NMPU Index Construction

We apply a publicly available BERT model[6] on each article, which is fine-tuned to the task of word-level uncertainty detection. The model returns a list of all words that express uncertainty in each article and we compute the article's uncertainty score as the number of words expressing uncertainty over the total number of words in the article. An alternative approach is to use a fixed dictionary of uncertainty words commonly used in finance and perform the same calculations as described above. One such dictionary is the popular lexicon proposed by Loughran and McDonald [2011]. Apart from our method we also apply this dictionary-based technique and treat it as context-agnostic baseline. We anticipate that the BERT-based method will prove to be superior for two main reasons: (a) dictionary-based methods are *static*, that is, words that imply uncertainty are always the same across various time periods and periodicals and, (b), they ignore context, where in reality context gives rise to the meaning of a word and the implications of using it.[7]

2.3 Daily and Monthly NMPU

With the raw uncertainty score assigned to each available monetary policy article from a single publication, we take a daily average to get a raw uncertainty score for day t, denoted UNC_t. If there are no articles on day t, UNC_t is assumed to be 0 in the following calculation. We keep only the trading day scores, and subtract the 252-day trailing average to obtain the

[6]The model is available at https://huggingface.co/jeniakim/hedgehog

[7]Consider for example a phrase taken from a WSJ article on March 16, 2023: "Last week's jobs report told a similar story: The US economy added 311,000 jobs in February, more than economists expected but down sharply from January's 517,000." A dictionary-based method will tag the word "expected" since it is considered an uncertainty-related word in the Loughran and McDonald [2011] dictionary. A language model, such as BERT, takes the entire sentence into account, recognizes that "expected" refers to uncertainty in the past, and does not classify it as a word that implies uncertainty about current or future outcomes. The word will, correctly, not contribute to the uncertainty score of that article. Similarly, cases of negation, co-reference and ambiguity, all of which may affect uncertainty levels are correctly resolved by BERT models.

NMPU on day t, denoted NMPU_t:

$$\text{NMPU}_t = \text{UNC}_t - \frac{1}{252} \sum_{k=t-251}^{t} \text{UNC}_k. \tag{1}$$

This calculation is done on news from Wall Street Journal and on news from New York Times to get the publication specific scores, $\text{NMPU}_t^{\text{WSJ}}$ for Wall Street Journal and $\text{NMPU}_t^{\text{NYT}}$ for New York Times. To get the combined score on day t, we calculate the average of $\text{NMPU}_t^{\text{WSJ}}$ and $\text{NMPU}_t^{\text{NYT}}$. If one of the publications is not available on day t, then the combined score will be the NMPU_t of the other publication. For the publication specific monthly scores $\text{NMPU}_t^{\text{WSJ}}$ and $\text{NMPU}_t^{\text{NYT}}$ of month t, we start with the raw daily scores UNC_t, keep only the trading day values, then average over the month. Then NMPU_t of a month t is defined to be the average of $\text{NMPU}_t^{\text{WSJ}}$ and $\text{NMPU}_t^{\text{NYT}}$. For Canadian data, we keep daily $\text{NMPU}_t = \text{UNC}_t$ whenever UNC_t is available and calculate monthly NMPU_t by a simple average over the month, as the news count and coverage is much lower and setting 0s to days without news will make 0 the dominant value in the timeseries.

2.4 NMPU Properties

We first demonstrate how NMPU correlates with existing narrative-based and market-based uncertainty measures. The main narrative-based measure we consider is the EPU by Baker et al. [2016].[8] Moreover, we consider indices such as the CBOE Volatility Index (VIX), the Merrill Lynch Option Volatility Estimate (MOVE) and the LIBOR-based uncertainty measure by Bauer et al. [2021] (denoted MPU-MKT). Figure 1 illustrates the monthly NMPU paired with the above narrative-based and market-based indices. Table 1 shows correlations at the daily and monthly frequency between NMPU and the aforementioned measures that are available at both frequencies, focusing on the US economy. Since MPU and HRS MPU are not available at daily frequency, we include them in Table A2 found in the Appendix.

At the daily frequency, we find that NMPU obtained from New York Times positively correlates with marked-based measures and even more so with narrative-based measures. NMPU based on Wall Street Journal is directionally aligned with narrative-based measures. Interestingly, its correlation to market-based measures is negative. Similar conclusions are drawn at the monthly frequency. The LIBOR-based MPU-MKT by Bauer et al. [2021] positively correlates with VIX and MOVE. At the monthly frequency NMPU obtained from WSJ is inversely correlated to MPU-MKT. Finally, it is worth noting that EPU, MPU,

[8] Results for MPU by Baker et al. [2016] and HRS-MPU by Husted et al. [2020] can be found in the Appendix, Figure A1.

HRS-MPU move with VIX but have a weaker and often negative statistical relationship to the MOVE index. The above results are an indication that different uncertainty measures may have varying interpretations and applicability; some may serve as proxies for market volatility, while others are better at capturing uncertainty about interest rates. With our methodology which allows us to obtain reliable scores even from a single news source, we find that uncertainty reported in different periodicals may also correspond to these alternative interpretations. When it comes to Canadian newspapers, we find that NMPU is positively correlated with EPU (the correlation is 0.324). The latter is the only other uncertainty index we can reliably obtain for Canada.

3 NMPU and Policy Announcements

In this section we focus our attention to the behavior of NMPU around monetary policy announcements. We consider FOMC meetings starting on Feb 04, 1994, and fixed announcement date (FAD) press releases by the Bank of Canada starting on Aug 28, 2001.

To perform the event study, for each announcement day τ we define $\text{NMPU}_{\tau,[a,b]}$ to be the mean NMPU in the window $[\tau + a, \tau + b]$ $(a < b)$. Specifically:

$$\text{NMPU}_{\tau,[a,b]} = \frac{1}{|T|} \sum_{t \in T} \text{NMPU}_t, \text{ where } T = \{t \in [\tau + a, \ldots, \tau + b] : \text{NMPU}_t \text{ is defined}\}. \quad (2)$$

We also define $\Delta\text{NMPU}_{\tau,[a,b]}$ $(a < 0,\ b > 0)$ to be the change of average NMPU before and after announcement. Specifically:

$$\Delta\text{NMPU}_{\tau,[a,b]} = \text{NMPU}_{\tau,[1,b]} - \text{NMPU}_{\tau,[a,-1]}. \quad (3)$$

Figure 2 shows that the average NMPU starts building up in the 5 days prior to the FOMC meeting and peaks on the day following the meeting when newspapers discuss the event with complete information. Uncertainty drops sharply immediately after and is gradually resolved within the first 5 post-announcement days, on average. Both in the US and Canada, we observe weekly seasonality in uncertainty before and after announcements. Finally, in the US, when uncertainty is sourced from different newspapers (WSJ and NYT) its peaks appear shifted by a day. NMPU^{WSJ} peaks on the day before the FOMC meeting, while NMPU^{NYT} peaks on the day after the FOMC meeting. Similar conclusions are drawn when examining NMPU's movements around BoC FAD press releases, as shown in Figure A3 in

the Appendix.[9]

Next, we study how NMPU and changes in NMPU respond to announcement-day returns. Specifically, we regress $\text{NMPU}_{\tau,[1,3]}$ and $\Delta\text{NMPU}_{\tau,[-3,3]}$ on the returns of the S&P500 index observed on the day of the FOMC meeting (day τ). We control for absolute announcement-day returns. Orthogonally, to examine the effects of bad (*e.g.,* "hawkish") monetary policy news we regress $\text{NMPU}_{\tau,[1,3]}$ and $\Delta\text{NMPU}_{\tau,[-3,3]}$ on the dummy indicator variable $\mathbb{1}_{\text{Ret}_\tau<0}$. In Panel A of Table 2 we report results. We also replicate the regression models but with EPU, MPU-MKT and VIX as the response variables, and we report results in Panel B. We find a statistically significant relationship where positive market returns on announcement days are associated with reduced NMPU (negative β coefficient) over the following 3 days. This behavior is also seen when it comes to the market-based index, MPU-MKT. Instead, the daily EPU index by Baker et al. [2016] yields a positive β coefficient. In scenarios where news are "bad", we observe that poor (negative) returns on FOMC meeting days are associated with an increase in uncertainty sourced from WSJ over the following 3 days after the announcement. This increase is consistent over time but results are mixed when applied to a less business focused paper such as NYT. MPU-MKT and VIX react similarly to bad news, but EPU responds in the opposite manner.

Going beyond the short-window increase (decrease) in NMPU following negative (positive) announcement-day returns, we are also interested in how uncertainty evolves over longer time windows following FOMC announcements. To this end, we calculate the average cumulative NMPU for the 30 days following FOMC announcements. To illustrate how the polarity of announcement-day news affects uncertainty over longer horizons, we calculate cumulative NMPU conditional on announcement-day returns (positive vs. negative). Figure 3 illustrates that narrative uncertainty captured by WSJ continues growing following "hawkish" news for the entire window, whereas uncertainty drops and is continuously resolved over time in the case of good news. When it comes to NYT, uncertainty remains elevated when FOMC news are bad, and abruptly drops and remains supressed in the opposite scenario. In both cases, an uptick in uncertainty is observed towards the end of the 30-day window, which coincides with the days leading up to the next FOMC meeting; uncertainty starts building up. The same figure also shows that the gap between uncertainty levels in the "hawkish" and "dovish" scenarios continues growing over the 30-day period.

To showcase that the advanced NLP method is superior to a context-agnostic method, we repeat a similar experiment but use words from the dictionary of Loughran and McDonald [2011] to calculate article-level uncertainty. We refer to the measure as NMPU-LM and we

[9]We find that the build-up and resolution window is narrower. Uncertainty rises sharply within 2 days prior to the event. Then, the press release rapidly lowers uncertainty within the span of 2 days.

regress NMPU-LM and changes in NMPU-LM on FOMC announcement returns without controlling for absolute returns this time. Table A3 in the Appendix includes results. For reference we also include results of the same regression for the NLP-based NMPU (referred to as NMPU-ML in that context) and other existing measures (EPU, MPU-MKT and VIX). Despite the fact that the coefficient sign is the one expected (negative) we find that the response of NMPU-LM is not statistically significant. Finally, we repeat the same analysis focusing on how NMPU derived from Canadian newspapers responds to BoC FAD press releases. Conclusions are similar to those drawn from the US study. Regression results are summarized in Table A5 in the Appendix.

4 NMPU and Macroeconomic Fundamentals

Studies by Bansal and Shaliastovich [2011], Kacperczyk et al. [2016] have established a link between endogenous attention and uncertainty. Specifically, it is shown that voluntary investor attention rises with economic uncertainty and risk aversion. Subsequent studies by Ai and Bansal [2018], Fisher et al. [2022] expose a relationship between attention, macroeconomic fundamentals and announcement risk premia. This thread of work motivates us to empirically study how low-frequency NMPU responds to macroeconomic fundamentals, and to understand whether macroeconomic drivers of NMPU are different from those driving market-based measures of uncertainty. To this end we consider fundamentals such as unemployment, the housing index (Case-Shiller U.S. National Home Price Index), Consumer Price Index, Fed Fund rates and the 10-year T-bill yield.

Let F_t^M denote a fundamental available at a calendar month frequency, where t indexes months. Let F_t^Q, F_t^Y, and F_t^{4Y} be the moving averages over windows of size 3, 12, and 48 months, ending at t (*i.e.*, $F_t^Q = \frac{1}{3}\sum_{k=0}^{2} F_{t-k}^M$). We then define a decomposition of F_t^M as follows:

$$\begin{aligned} F_t^M &= \left(F_t^M - F_t^Q\right) + \left(F_t^Q - F_t^Y\right) + \left(F_t^Y - F_t^{4Y}\right) + F_t^{4Y} \\ &= F_t^{M-Q} + F_t^{Q-Y} + F_t^{Y-4Y} + F_t^{4Y} \end{aligned} \tag{4}$$

where each term is a detrended moving average over the appropriate calendar interval (month, quarter, year, four years). Similar types of expansions are used to capture variability in persistence of the fundamentals and long-cycle dependencies by Fisher et al. [2022], Ortu et al. [2013], Calvet and Fisher [2007].

We regress NMPU on the detrended moving averages of fundamentals and the absolute values of the detrended moving averages:

$$\text{NMPU}_t = \alpha + \beta_1 F_t^{M-Q} + \beta_2 F_t^{Q-Y} + \beta_3 F_t^{Y-4Y} + \beta_4 |F_t^{M-Q}| + \beta_5 |F_t^{Q-Y}| + \beta_6 |F_t^{Y-4Y}| + \epsilon_t, \tag{5}$$

We expect that large changes in macroeconomic variables irrespective of direction should lead to increased levels of monetary uncertainty. This is captured by the coefficients of the absolute value terms in Eq. 5. On the other hand, asymmetries in how uncertainty responds (*i.e.,* bigger changes for "bad" changes in fundamentals) are expected to be shown by the coefficients of the remaining terms.

Results for the NMPU measure are in Panel A of Table 3. The adjusted R^2 reaches up to 19.1% across all regressions, and all of them have at least one statistically significant coefficient, with the Federal fund rate being an exception. We find that large changes in the absolute values for housing index, CPI, unemployment rates and 10 year treasury bond yield leads to more uncertainty about the monetary policy as measured by NMPU. However, we find that shocks in CPI at the yearly frequency are associated with a drop in uncertainty. Panel B in the same table shows results when we regress MPU-MKT on the same macro variables. The regressions expose significance in the effect of macro variable changes. However, the direction (sign of coefficients) does not always agree with what is observed in NMPU regressions. For instance, a large change in unemployment leads to a drop in MPU-MKT, but NMPU rises. On the other hand, MPU-MKT has a stronger response to Fed fund rate changes, which is expected since LIBOR closely tracks the key interest rates by the Federal Reserve.

5 Conclusion

We showed how advanced language models can be used to measure uncertainty about monetary policy in news articles. Using a sample of about 70 thousand news articles from 1984 to 2023, we construct a monetary policy uncertainty index for the US. The index can be reliably computed daily since the proposed method can generate meaningful uncertainty levels even with a few articles available. We show that the new uncertainty measure: (a) positively correlates with existing narrative-based measures that are typically restricted to monthly frequencies, (b) on average it builds up before FOMC announcements and is lowered in the days following, (c) has a statistically significant response to"hawkish" announcements, appearing elevated after bad news, and (d) responds to large changes in macroeconomic fundamentals such as CPI, unemployment and housing prices. We repeat the empirical study around Bank

of Canada FAD press releases and observe a similar behavior when uncertainty is measured using Canadian news sources.

The primary focus of this paper is on monetary policy, but the measure can be applied to any topic, including sectors such as housing, energy, and other macroeconomic fundamentals such as inflation. We plan to extend the study and empirically evaluate the measure's ability to forecast changes in macro variables, and sector-level returns and volatility. Finally, we plan to build upon existing frameworks that link endogenous attention -which is partly driven by uncertainty- to announcement risk premia and evaluate whether the proposed uncertainty measure can serve as an instrument for future risk premia following central bank announcements.

Figure 1: US NMPU, compared with exisiting uncertainty measures

This figure shows the monthly US NMPU with EPU, Market-based MPU (MPU-MKT), VIX and MOVE from Jan, 1984. All the scores are smoothed by a 12 month moving average. The gray vertical bars are NBER recessions.

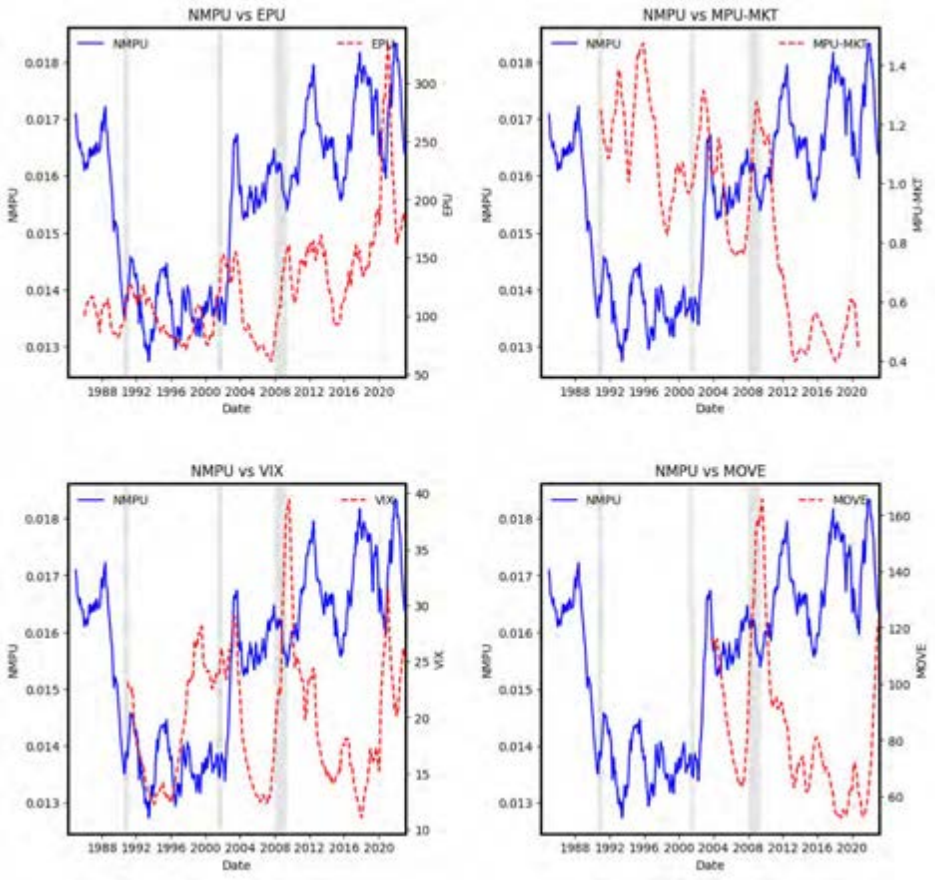


Figure 2: US NMPU around FOMC meetings

This figure shows the average US NMPU from WSJ (NMPU^{WSJ}), NYT (NMPU^{NYT}) and combined (NMPU), using both Loughran McDonald (LM) and Machine Learning (ML) scores, around FOMC meetings, starting at Feb 04, 1994.

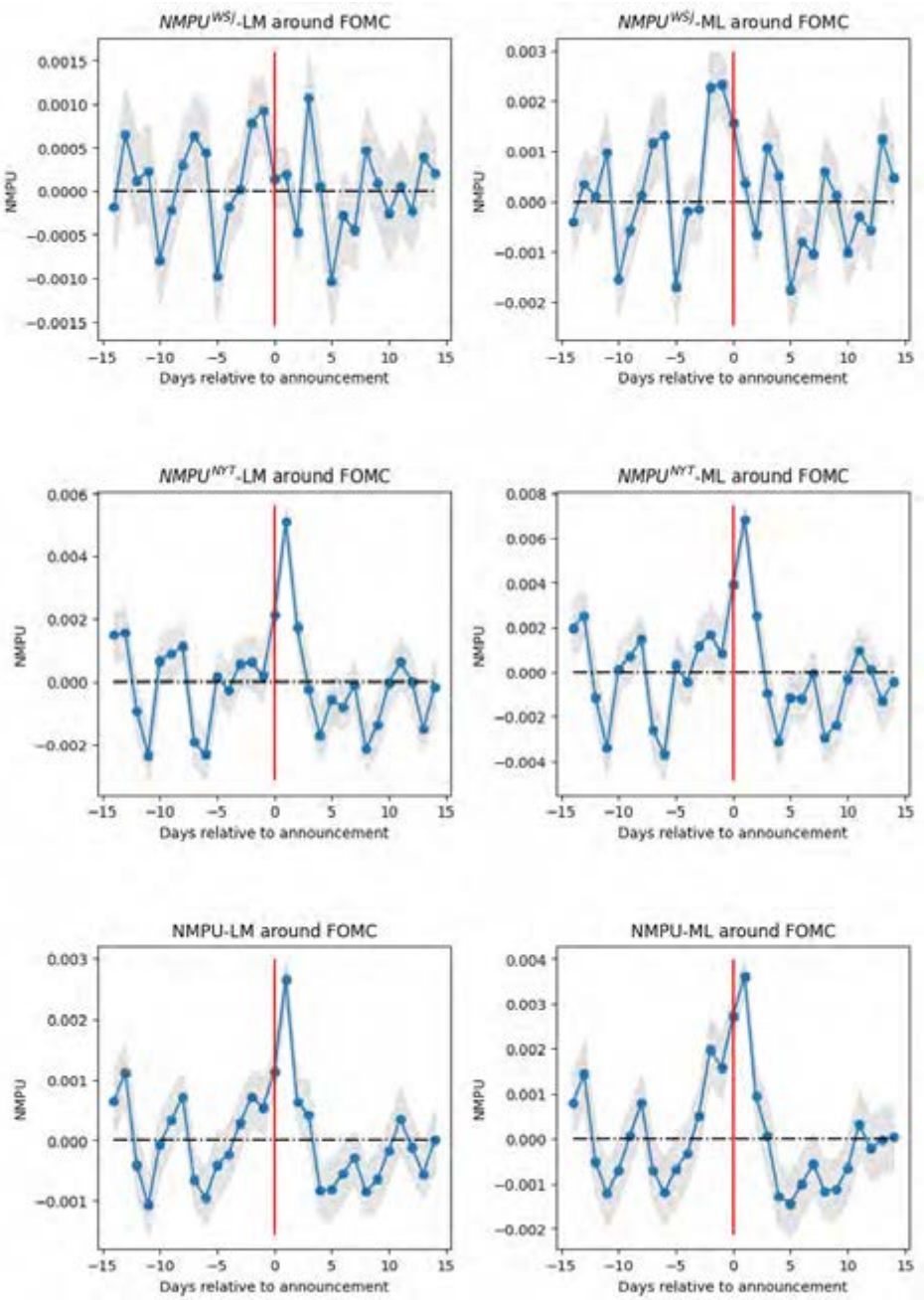


Figure 3: US cumulative NMPU after FOMC meetings

The top two figures show the average US cumulative NMPU from WSJ and NYT after FOMC meetings, detrended by the NMPU on the day before the FOMC announcement, starting at Feb 04, 1994. The red line shows the evolution of NMPU following bad news on the announcement day, where Ret < 0. The blue line shows the evolution of NMPU following good news on the announcement day, where Ret > 0. The bottom two figures show the difference between the blue and red line, with 95% confidence intervals.

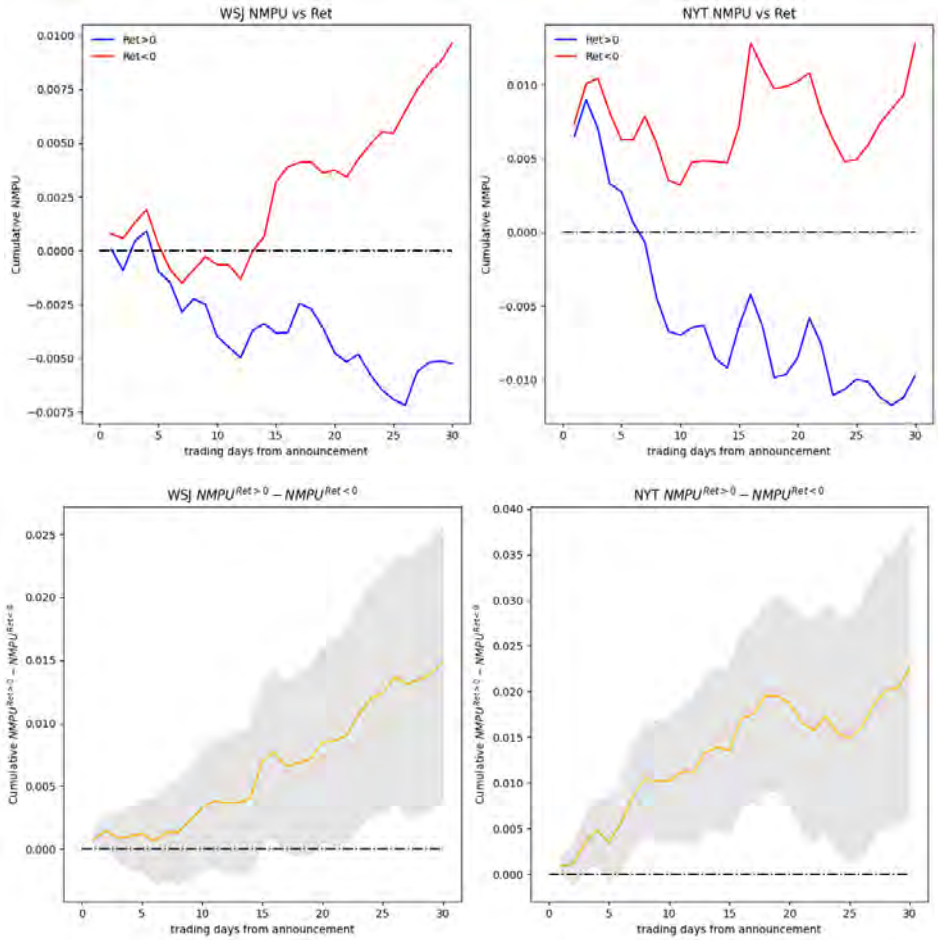


Table 1: Correlation

Panel A of this table reports the daily correlation between NMPU obtained from Wall Street Journal ($NMPU^{WSJ}$), NMPU obtained from New York Times ($NMPU^{NYT}$), and the EPU, Market-Based MPU (MPU-MKT), VIX, MOVE indices. Panel B reports the monthly correlation. The indices are all detrended by 252 day average. The monthly indices $NMPU^{WSJ}$, $NMPU^{NYT}$ and MPU-MKT are averaged across trading days within a month. The monthly VIX and MOVE are taken as the last available index in the month.

Panel A. Daily

	$NMPU^{WSJ}$	$NMPU^{NYT}$	EPU	MPU-MKT	VIX	MOVE
$NMPU^{WSJ}$	1.000	0.128	0.031	-0.007	-0.021	-0.017
$NMPU^{NYT}$	0.128	1.000	0.042	0.006	0.017	0.018
EPU	0.031	0.042	1.000	0.004	0.372	0.101
MPU-MKT	-0.007	0.006	0.004	1.000	0.308	0.124
VIX	-0.021	0.017	0.372	0.308	1.000	0.321
MOVE	-0.017	0.018	0.101	0.124	0.321	1.000

Panel B. Monthly

	$NMPU^{WSJ}$	$NMPU^{NYT}$	EPU	MPU-MKT	VIX	MOVE
$NMPU^{WSJ}$	1.000	0.110	0.286	-0.552	-0.023	-0.219
$NMPU^{NYT}$	0.110	1.000	0.252	-0.082	0.026	-0.120
EPU	0.286	0.252	1.000	-0.316	0.425	-0.046
MPU-MKT	-0.552	-0.082	-0.316	1.000	0.183	0.787
VIX	-0.023	0.026	0.425	0.183	1.000	0.623
MOVE	-0.219	-0.120	-0.046	0.787	0.623	1.000

Table 2: Uncertainty changes around announcements

This table reports the results of the following regressions:

$$\begin{aligned}
\text{NMPU}_{\tau,[1,3]} =& \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau, \\
\text{NMPU}_{\tau,[1,3]} =& \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau, \\
\Delta\text{NMPU}_{\tau,[-3,3]} =& \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau \text{ and} \\
\Delta\text{NMPU}_{\tau,[-3,3]} =& \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau,
\end{aligned}$$

where $\mathbb{1}_{\text{Ret}_\tau<0}$ is 1 if $\text{Ret}_\tau < 0$ and 0 otherwise. We look at 223 FOMC announcements starting at Feb 4, 1994. The daily EPU, MPU-MKT and VIX are also trading day only and detrended by 252 day average. The definitions of EPU, ΔEPU, MPU and ΔMPU are the same as NMPU and ΔNMPU, following the variable definitions in the formula. For VIX, the variable definitions are $\text{VIX}_{\tau,[1,3]} = \text{VIX}_{\tau+3} - \text{VIX}_{\tau+1}$, $\Delta\text{VIX}_{\tau,[-3,3]} = \text{VIX}_{\tau+3} - \text{VIX}_{\tau-3}$. The robust standard errors HC3 are reported in parenthesis. *, **, *** denote the statistical significance at the 10%, 5%, 1% levels, respectively.

Panel A. NMPU

	WSJ				NYT				Combined			
	NMPU		ΔNMPU		NMPU		ΔNMPU		NMPU		ΔNMPU	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\mathbb{1}_{\text{Ret}_\tau<0}$	0.0003		0.0007		0.0011*		0.0000		0.0007*		0.0004	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
Ret_τ		-0.0392**		-0.0642**		-0.0802**		-0.0654		-0.0597***		-0.0648**
		(0.02)		(0.03)		(0.03)		(0.04)		(0.02)		(0.03)
$\lvert\text{Ret}_\tau\rvert$		-0.0145		0.0145		0.0419		0.0178		0.0137		0.0162
		(0.03)		(0.04)		(0.05)		(0.06)		(0.03)		(0.04)
Intercept	0.0001	0.0005*	-0.0015***	-0.0012***	0.0023***	0.0027***	0.0016***	0.0016**	0.0012***	0.0016***	0.0000	0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	-0.002	0.023	0.004	0.025	0.009	0.019	-0.005	0.002	0.009	0.034	-0.002	0.019
N	223	223	223	223	223	223	223	223	223	223	223	223

Panel B. Existing Measures

	EPU				MPU-MKT				VIX			
	EPU		ΔEPU		MPU		ΔMPU		VIX		ΔVIX	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\mathbb{1}_{\text{Ret}_\tau<0}$	-12.7573*		-7.2418		0.0122		0.0169***		-0.6314*		0.6390	
	(7.21)		(6.32)		(0.02)		(0.01)		(0.33)		(0.55)	
Ret_τ		606.1612		295.7977		-1.6229		-1.1311***		23.8529		-64.3195
		(461.19)		(419.98)		(1.12)		(0.35)		(18.51)		(39.26)
$\lvert\text{Ret}_\tau\rvert$		1017.0459*		-46.9111		-0.3716		-0.4200		-75.0089***		-11.3744
		(617.92)		(549.83)		(1.56)		(0.59)		(28.48)		(53.21)
Intercept	15.1899***	-0.2255	10.2654**	6.7469	-0.0353***	-0.0227	-0.0244***	-0.0106**	0.3260	0.6042**	-0.3357	0.2030
	(5.64)	(4.77)	(4.42)	(4.41)	(0.01)	(0.01)	(0.00)	(0.00)	(0.25)	(0.28)	(0.33)	(0.39)
R^2	0.008	0.043	0.001	-0.005	-0.003	0.009	0.028	0.088	0.011	0.038	0.002	0.028
N	223	223	223	223	223	223	223	223	223	223	223	223

Table 3: Uncertainty with Fundamentals

This table reports the results of the following regression:

$$y_t = \alpha + \beta_1 F_t^{M-Q} + \beta_2 F_t^{Q-Y} + \beta_3 F_t^{Y-4Y} + \beta_4 |F_t^{M-Q}| + \beta_5 |F_t^{Q-Y}| + \beta_6 |F_t^{Y-4Y}| + \epsilon_t,$$

where $y_t = \text{NMPU}_t$ in Panel A and $y_t = \text{MPU-MKT}_t$ in Panel B. The MPU-MKT is also the monthly average of trading day values. The definitions of F can be found in Section 4. The standard errors are calculated using Newy-west (HAC) with max lags set to $N^{\frac{1}{4}}$. The fundamental indices are scaled down by a factor of 100 in both regressions.

Panel A. NMPU

	Housing	CPI	Unemployment	Fed Fund	Treasury Bond (10y)
	(1)	(2)	(3)	(4)	(5)
F_t^{M-Q}	0.0239	-0.0090	0.0403	0.0275	-0.0372
	(0.02)	(0.02)	(0.03)	(0.08)	(0.05)
F_t^{Q-Y}	-0.0009	0.0044	-0.1025***	0.0263	0.0516*
	(0.01)	(0.00)	(0.04)	(0.03)	(0.03)
F_t^{Y-4Y}	-0.0039**	-0.0132***	-0.0279*	-0.0153	0.0429
	(0.00)	(0.00)	(0.02)	(0.01)	(0.03)
$\lvert F_t^{M-Q}\rvert$	0.0135	0.0310	-0.0734	-0.0728	-0.0442
	(0.02)	(0.02)	(0.05)	(0.09)	(0.07)
$\lvert F_t^{Q-Y}\rvert$	0.0046	0.0528***	0.1835***	-0.0395	0.0153
	(0.01)	(0.01)	(0.05)	(0.04)	(0.04)
$\lvert F_t^{Y-4Y}\rvert$	0.0093***	-0.0132***	0.0823***	-0.0224	0.0861**
	(0.00)	(0.00)	(0.02)	(0.02)	(0.04)
Intercept	0.0143***	0.0161***	0.0142***	0.0159***	0.0153***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.191	0.089	0.095	0.025	0.018
N	387	464	440	441	466

Panel B. MPU-MKT

	Housing	CPI	Unemployment	Fed Fund	Treasury Bond (10y)
	(1)	(2)	(3)	(4)	(5)
F_t^{M-Q}	-2.9593	1.1934	13.6536**	19.7738	-2.6874
	(3.28)	(5.80)	(6.10)	(28.45)	(9.69)
F_t^{Q-Y}	-2.1240*	1.2559	34.4241**	-2.4955	8.8908
	(1.22)	(1.23)	(14.68)	(11.35)	(8.60)
F_t^{Y-4Y}	0.0973	2.7754***	14.1548***	-1.8300	-8.7707
	(0.42)	(0.95)	(3.71)	(2.83)	(12.93)
$\lvert F_t^{M-Q}\rvert$	-4.0780	-11.3145	-25.3535***	55.0558**	59.1808***
	(4.18)	(7.63)	(7.87)	(26.87)	(15.85)
$\lvert F_t^{Q-Y}\rvert$	-0.2229	-1.0564	-36.8630**	21.0834*	33.4872**
	(1.26)	(3.29)	(16.05)	(11.39)	(13.90)
$\lvert F_t^{Y-4Y}\rvert$	-1.2753**	2.7754***	-17.3927***	16.9889***	-0.2458
	(0.52)	(0.95)	(5.05)	(3.42)	(15.59)
Intercept	1.1080***	0.5891***	1.1979***	0.5851***	0.6843***
	(0.06)	(0.12)	(0.04)	(0.05)	(0.07)
R^2	0.201	0.064	0.342	0.392	0.112
N	358	369	369	369	369

Appendix

Table A1: Search words

This table reports search words to select articles related to monetary policy and national banks in the Wall Street Journal (WSJ), the New York Times (NYT) and National Post to construct the narrative monetary policy uncertainty indices (NMPU) for the US and CA. We retrieve the articles for which the search words appear either in the headline or in the article.

Country	Newspapers search words
US	("monetary policy" OR "monetary policies" OR "interest rate" OR "interest rates" OR "Federal fund rate" OR "Federal funds rate" OR "Fed fund rate" OR "Fed funds rate") AND ("Federal Reserve" OR "the Fed" OR "Federal Open Market Committee" OR "FOMC")
CA	("monetary policy" OR "monetary policies" OR "interest rate" OR "interest rates") AND ("Bank of Canada" OR "BoC")

Figure A1: US NMPU, compared with MPU and HRS MPU

This figure shows the monthly US NMPU with MPU, HRS MPU from Jan, 1984. All the scores are smoothed by a 12 month moving average. The gray vertical bars are NBER recessions.

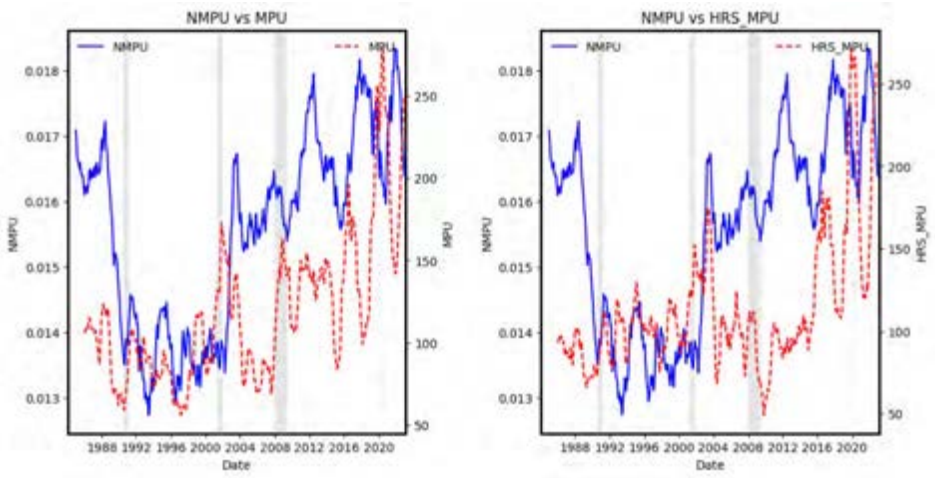


Table A2: US monthly correlation with MPU

This table adds monthly MPU and HRS MPU as compared to Table 1.

	$NMPU^{WSJ}$	$NMPU^{NYT}$	EPU	MPU-MKT	VIX	MOVE	MPU	HRS MPU
$NMPU^{WSJ}$	1.000	0.110	0.286	-0.552	-0.023	-0.219	0.322	0.200
$NMPU^{NYT}$	0.110	1.000	0.252	-0.082	0.026	-0.120	0.172	0.184
EPU	0.286	0.252	1.000	-0.316	0.425	-0.046	0.790	0.488
MPU-MKT	-0.552	-0.082	-0.316	1.000	0.183	0.787	-0.301	-0.175
VIX	-0.023	0.026	0.425	0.183	1.000	0.623	0.369	0.106
MOVE	-0.219	-0.120	-0.046	0.787	0.623	1.000	0.098	-0.077
MPU	0.322	0.172	0.790	-0.301	0.369	0.098	1.000	0.712
HRS MPU	0.200	0.184	0.488	-0.175	0.106	-0.077	0.712	1.000

Table A3: Uncertainty changes around announcements

This table reports the results of the following regressions:

$$\begin{aligned}\text{NMPU}_{\tau,[1,3]} =&\alpha + \beta_1 \text{Ret}_\tau + \epsilon_\tau \text{ and}\\ \Delta\text{NMPU}_{\tau,[-3,3]} =&\alpha + \beta_1 \text{Ret}_\tau + \epsilon_\tau.\end{aligned}$$

The Ret variable is calculated as the log return of S&P500 on the announcement day. We look at 223 FOMC announcements starting at Feb 4, 1994. The daily EPU, MPU-MKT and VIX are also trading day only and detrended by 252 day average. The definitions of EPU, ΔEPU, MPU and ΔMPU are the same as NMPU and ΔNMPU, following the variable definitions in the formula. For VIX, the variable definitions are $\text{VIX}_{\tau,[1,3]} = \text{VIX}_{\tau+3} - \text{VIX}_{\tau+1}$, $\Delta\text{VIX}_{\tau,[-3,3]} = \text{VIX}_{\tau+3} - \text{VIX}_{\tau-3}$. The robust standard errors HC3 are reported in parenthesis. *, **, *** denote the statistical significance at the 10%, 5%, 1% levels, respectively. We see that LM is not showing significance for any of the regression, thus it is not included in our main analysis.

Panel A. NMPU-LM

	WSJ		NYT		Combined	
	NMPU	ΔNMPU	NMPU	ΔNMPU	NMPU	ΔNMPU
	(1)	(2)	(3)	(4)	(5)	(6)
Ret_τ	-0.0109	-0.0100	-0.0263	-0.0364	-0.0186	-0.0232
	(0.01)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)
Intercept	0.0003**	-0.0003	0.0023***	0.0018***	0.0013***	0.0008***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	-0.000	-0.003	0.002	0.003	0.004	0.003
N	223	223	223	223	223	223

Panel B. NMPU-ML

	WSJ		NYT		Combined	
	NMPU	ΔNMPU	NMPU	ΔNMPU	NMPU	ΔNMPU
	(1)	(2)	(3)	(4)	(5)	(6)
Ret_τ	-0.0436***	-0.0597***	-0.0672**	-0.0599	-0.0554***	-0.0598**
	(0.02)	(0.02)	(0.03)	(0.04)	(0.02)	(0.03)
Intercept	0.0004**	-0.0011***	0.0030***	0.0017***	0.0017***	0.0003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.026	0.028	0.020	0.006	0.037	0.023
N	223	223	223	223	223	223

Panel C. Existing measures

	EPU		MPU-MKT		VIX	
	EPU	ΔEPU	MPU	ΔMPU	VIX	ΔVIX
	(1)	(2)	(3)	(4)	(5)	(6)
Ret_τ	921.6324**	281.2467	-1.7382*	-1.2613***	0.5864	-67.8477**
	(441.31)	(328.18)	(1.00)	(0.44)	(20.91)	(32.89)
Intercept	7.2536**	6.4019*	-0.0254**	-0.0137***	0.0526	0.1193
	(3.61)	(3.31)	(0.01)	(0.00)	(0.18)	(0.30)
R^2	0.030	-0.000	0.013	0.088	-0.005	0.032
N	223	223	223	223	223	223

Table A4: Uncertainty with Fundamentals Shifted

This table reports the results of the following regression:

$$y_t = \alpha + \beta_1 F_{t-1}^{M-Q} + \beta_2 F_{t-1}^{Q-Y} + \beta_3 F_{t-1}^{Y-4Y} + \beta_4 |F_{t-1}^{M-Q}| + \beta_5 |F_{t-1}^{Q-Y}| + \beta_6 |F_{t-1}^{Y-4Y}| + \epsilon_t,$$

where $y_t = \text{NMPU}_t$ in Panel A and $y_t = \text{MPU-MKT}_t$ in Panel B. This is the same regression as in Table 3, but we shift the fundamentals by 1 month to account for the announcement delay (F_{t-1} is reported in month t).

Panel A. NMPU

	Housing	CPI	Unemployment	Fed Fund	Treasury Bond (10y)
	(1)	(2)	(3)	(4)	(5)
F_{t-1}^{M-Q}	0.0402***	-0.0151	0.0284	0.0890	-0.0138
	(0.02)	(0.03)	(0.03)	(0.08)	(0.05)
F_{t-1}^{Q-Y}	-0.0073	0.0055	-0.1410***	0.0092	0.0512
	(0.01)	(0.01)	(0.05)	(0.03)	(0.03)
F_{t-1}^{Y-4Y}	-0.0031	-0.0150***	-0.0184	-0.0139	0.0431
	(0.00)	(0.00)	(0.02)	(0.01)	(0.04)
$\lvert F_{t-1}^{M-Q}\rvert$	0.0240	0.0902***	0.0769**	0.0453	-0.0498
	(0.02)	(0.03)	(0.04)	(0.09)	(0.06)
$\lvert F_{t-1}^{Q-Y}\rvert$	0.0024	0.0434***	0.1489***	-0.0587	0.0041
	(0.01)	(0.01)	(0.05)	(0.04)	(0.05)
$\lvert F_{t-1}^{Y-4Y}\rvert$	0.0096***	-0.0150***	0.0841***	-0.0206	0.0902**
	(0.00)	(0.00)	(0.02)	(0.02)	(0.04)
Intercept	0.0143***	0.0162***	0.0141***	0.0159***	0.0154***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.218	0.109	0.111	0.020	0.018
N	386	463	439	440	465

Panel B. MPU-MKT

	Housing	CPI	Unemployment	Fed Fund	Treasury Bond (10y)
	(1)	(2)	(3)	(4)	(5)
F_{t-1}^{M-Q}	-1.9253	3.6503	13.4730**	21.1984	6.7612
	(3.24)	(5.34)	(5.61)	(29.76)	(9.78)
F_{t-1}^{Q-Y}	-2.4740**	1.1266	37.0900**	-3.5445	7.2327
	(1.21)	(1.22)	(14.77)	(11.22)	(8.53)
F_{t-1}^{Y-4Y}	0.1823	2.9844***	13.5827***	-1.4348	-5.6672
	(0.42)	(0.95)	(3.67)	(2.70)	(12.72)
$\lvert F_{t-1}^{M-Q}\rvert$	-4.6787	-14.6431**	-25.7737***	55.5720*	69.6054***
	(4.07)	(6.84)	(7.38)	(28.93)	(16.04)
$\lvert F_{t-1}^{Q-Y}\rvert$	-0.0615	-0.7754	-38.9876**	21.7277*	33.2328**
	(1.25)	(3.30)	(16.14)	(11.40)	(13.52)
$\lvert F_{t-1}^{Y-4Y}\rvert$	-1.3402**	2.9844***	-17.4062***	16.8733***	3.9768
	(0.53)	(0.95)	(5.16)	(3.36)	(15.45)
Intercept	1.1097***	0.5671***	1.2033***	0.5833***	0.6601***
	(0.06)	(0.12)	(0.04)	(0.05)	(0.07)
R^2	0.201	0.081	0.348	0.394	0.129
N	357	369	369	369	369

Figure A2: CA NMPU, compared with EPU

This figure shows the monthly Canadian NMPU with EPU from Jan, 1984. Note that the NMPU here is a simple monthly average over National Post. Both the scores are smoothed by a 12 month moving average. The gray vertical bars are NBER recessions.

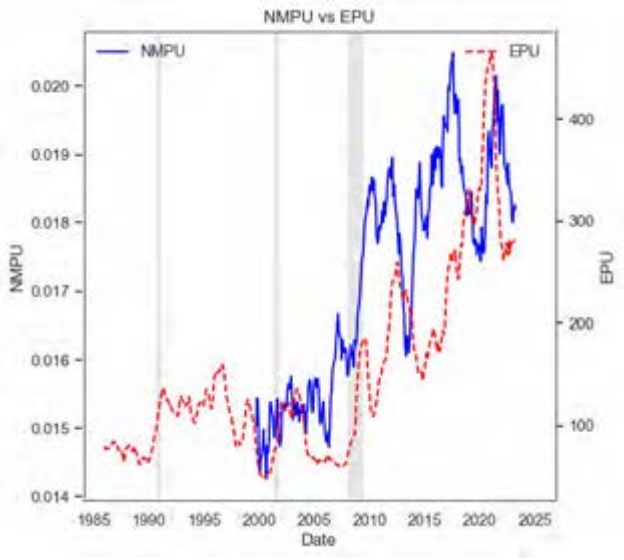


Figure A3: CA NMPU around BoC FAD

This figure shows the average Canadian NMPU from National Post, using Loughran McDonald (LM) and Machine Learning (ML) scores, around BoC FAD (fixed announcement date), starting at Aug 28, 2001. For this plot only, the Canadian NMPU is detrended by a 252 day trailing average.

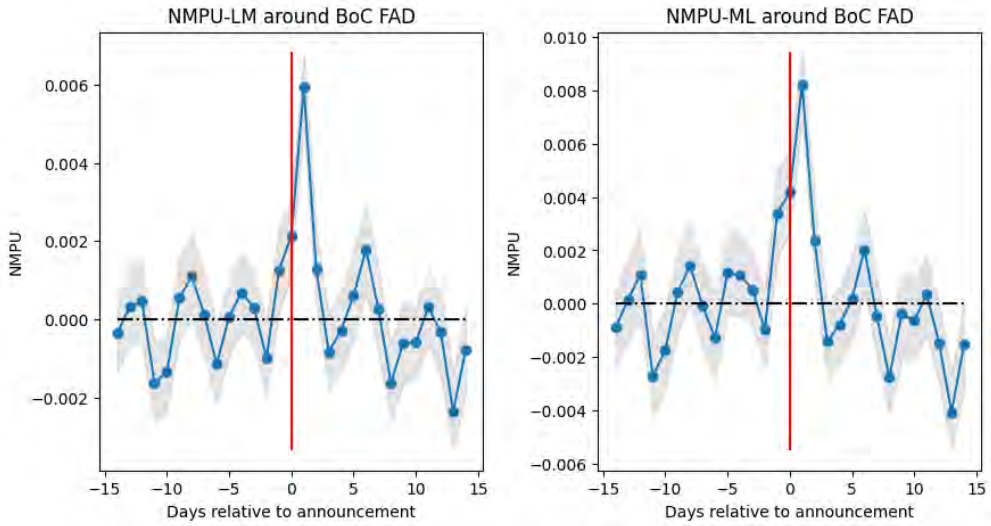


Table A5: Uncertainty changes around announcements (CA)

This table reports the results of the following regressions for Canadian market:

$$\begin{aligned}
\text{NMPU}_{\tau,[1,3]} =& \alpha + \beta_1 \text{Ret}_\tau + \epsilon_\tau, \\
\text{NMPU}_{\tau,[1,3]} =& \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau, \\
\text{NMPU}_{\tau,[1,3]} =& \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau, \\
\Delta\text{NMPU}_{\tau,[-3,3]} =& \alpha + \beta_1 \text{Ret}_\tau + \epsilon_\tau, \\
\Delta\text{NMPU}_{\tau,[-3,3]} =& \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau \text{ and} \\
\Delta\text{NMPU}_{\tau,[-3,3]} =& \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau,
\end{aligned}$$

where $\mathbb{1}_{\text{Ret}_\tau<0}$ is 1 if $\text{Ret}_\tau < 0$ and 0 otherwise. The Ret is calculated as the log return of TSX on the announcement day. We look at BoC announcements starting at Aug 28, 2001. Some datapoints are removed due to lack of news around the announcement. We don't exclude non-trading days nor detrend the Canadian data as the stock return behavior of Canadian market is not yet well studied and the daily coverage of news is lower. The robust standard errors HC3 are reported in parenthesis. *, **, *** denote the statistical significance at the 10%, 5%, 1% levels, respectively.

	NMPU			ΔNMPU		
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}_{\text{Ret}_\tau<0}$		-0.0001			0.0008	
		(0.00)			(0.00)	
Ret_τ	-0.0539		-0.0494	-0.1176**		-0.1096**
	(0.04)		(0.04)	(0.05)		(0.05)
$\lvert\text{Ret}_\tau\rvert$			-0.0526			-0.0727
			(0.05)			(0.05)
Intercept	0.0182***	0.0183***	0.0187***	-0.0003	-0.0007	0.0003
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.009	-0.006	0.013	0.029	-0.005	0.030
N	175	175	175	144	144	144

References

Kenneth N Kuttner. Monetary policy surprises and interest rates: Evidence from the fed funds futures market. *Journal of monetary economics*, 47(3):523–544, 2001.

Ben S Bernanke and Kenneth N Kuttner. What explains the stock market's reaction to federal reserve policy? *The Journal of finance*, 60(3):1221–1257, 2005.

Eric T Swanson. Measuring the effects of federal reserve forward guidance and asset purchases on financial markets. *Journal of Monetary Economics*, 118:32–53, 2021.

Pavel Savor and Mungo Wilson. How much do investors care about macroeconomic risk? evidence from scheduled economic announcements. *Journal of Financial and Quantitative Analysis*, 48(2):343–375, 2013.

David O Lucca and Emanuel Moench. The pre-fomc announcement drift. *The Journal of finance*, 70(1):329–371, 2015.

Emi Nakamura and Jón Steinsson. High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330, 2018.

Anna Cieslak, Adair Morse, and Annette Vissing-Jorgensen. Stock returns over the fomc cycle. *The Journal of Finance*, 74(5):2201–2248, 2019.

Oliver Boguth, Vincent Grégoire, and Charles Martineau. Shaping expectations and coordinating attention: The unintended consequences of fomc press conferences. *Journal of Financial and Quantitative Analysis*, 54(6):2327–2353, 2019.

Michael D Bauer, Aeimit Lakdawala, and Philippe Mueller. Market-Based Monetary Policy Uncertainty. *The Economic Journal*, 132(644):1290–1308, 11 2021. ISSN 0013-0133. doi: 10.1093/ej/ueab086. URL `https://doi.org/10.1093/ej/ueab086`.

Scott R. Baker, Nicholas Bloom, and Steven J. Davis. Measuring Economic Policy Uncertainty*. *The Quarterly Journal of Economics*, 131(4):1593–1636, 07 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw024. URL `https://doi.org/10.1093/qje/qjw024`.

Lucas Husted, John Rogers, and Bo Sun. Monetary policy uncertainty. *Journal of Monetary Economics*, 115:20–36, 2020. ISSN 0304-3932. doi: https://doi.org/10.1016/j.jmoneco.2019.07.009. URL `https://www.sciencedirect.com/science/article/pii/S0304393218301661`.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*, abs/1810.04805, 2019a.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805 [cs.CL]*, 2019b. doi: 10.48550. URL https://doi.org/10.48550/arXiv.1810.04805.

Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 2011. doi: 1540-6261.2010.01625. URL https://doi.org/10.1111/j.1540-6261.2010.01625.x.

Ravi Bansal and Ivan Shaliastovich. Learning and Asset-price Jumps. *The Review of Financial Studies*, 24(8):2738–2780, 04 2011. ISSN 0893-9454. doi: 10.1093/rfs/hhr023. URL https://doi.org/10.1093/rfs/hhr023.

Marcin Kacperczyk, Stijn Van Nieuwerburgh, and Laura Veldkamp. A rational theory of mutual funds' attention allocation. *Econometrica*, 84(2):571–626, 2016. doi: https://doi.org/10.3982/ECTA11412. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11412.

Hengjie Ai and Ravi Bansal. Risk preferences and the macroeconomic announcement premium. *Econometrica*, 86(4):1383–1430, 2018. doi: https://doi.org/10.3982/ECTA14607. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14607.

Adlai Fisher, Charles Martineau, and Jinfei Sheng. Macroeconomic Attention and Announcement Risk Premia. *The Review of Financial Studies*, 35(11):5057–5093, 02 2022. ISSN 0893-9454. doi: 10.1093/rfs/hhac011. URL https://doi.org/10.1093/rfs/hhac011.

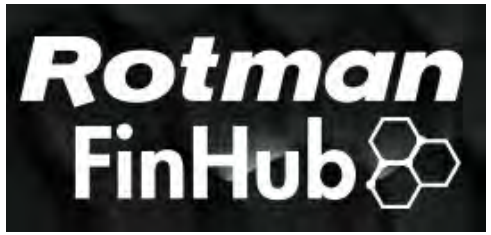
Fulvio Ortu, Andrea Tamoni, and Claudio Tebaldi. Long-Run Risk and the Persistence of Consumption Shocks. *The Review of Financial Studies*, 26(11):2876–2915, 11 2013. ISSN 0893-9454. doi: 10.1093/rfs/hht038. URL https://doi.org/10.1093/rfs/hht038.

Laurent E. Calvet and Adlai J. Fisher. Multifrequency news and stock returns. *Journal of Financial Economics*, 86(1):178–212, 2007. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2006.09.001. URL https://www.sciencedirect.com/science/article/pii/S0304405X07000906.

Narrative Monetary Policy Uncertainty

Charles Martineau, **Zissis Poulos**, Yuntao Wu, Cameron Thompson
Maryam Haghighi, Jun Yuan, John Hull







* "The views in this presentation are those of the authors and do not represent the official views of the Bank of Canada".

News articles, asset prices, and fundamentals

“Do media narratives predict economic fundamentals?”
“How do central bank decisions shape media narratives?”

News articles, asset prices, and fundamentals

"Do media narratives predict economic fundamentals?"
"How do central bank decisions shape media narratives?"

Narrative = Soft Information

- **Sentiment**
  Often good at predicting 1st moment, i.e., changes

- **Uncertainty**
  Can it predict volatility and risk premia?

News articles, asset prices, and fundamentals

"Do media narratives predict economic fundamentals?"
"How do central bank decisions shape media narratives?"

Narrative = Soft Information

- **Sentiment**
  Often good at predicting 1st moment, i.e., changes

- **Uncertainty**
  Can it predict volatility and risk premia?

Why do we care?
Understanding how policy uncertainty transmits to financial markets.

Monetary Policy Uncertainty

Influence of policy change on equity returns and long-term rates is well-established

Effects of uncertainty around monetary policy still underexplored

Monetary Policy Uncertainty

Influence of policy change on equity returns and long-term rates is well-established

Effects of uncertainty around monetary policy still underexplored

The Challenge

It is not obvious how uncertainty can be quantified

Monetary Policy Uncertainty

Influence of policy change on equity returns and long-term rates is well-established

Effects of uncertainty around monetary policy still underexplored

The Challenge

It is not obvious how uncertainty can be quantified

Proxies of uncertainty

Monetary Policy Uncertainty

Influence of policy change on equity returns and long-term rates is well-established

Effects of uncertainty around monetary policy still underexplored

The Challenge

It is not obvious how uncertainty can be quantified

Proxies of uncertainty

- **Market-based** (e.g., prices of interest rate derivatives)

Monetary Policy Uncertainty

Influence of policy change on equity returns and long-term rates is well-established

Effects of uncertainty around monetary policy still underexplored

The Challenge

It is not obvious how uncertainty can be quantified

Proxies of uncertainty

- **Market-based** (e.g., prices of interest rate derivatives)
- **Model-based** (e.g., time-series volatility of policy rule innovations)

Monetary Policy Uncertainty

Influence of policy change on equity returns and long-term rates is well-established

Effects of uncertainty around monetary policy still underexplored

The Challenge

It is not obvious how uncertainty can be quantified

Proxies of uncertainty

- **Market-based** (e.g., prices of interest rate derivatives)
- **Model-based** (e.g., time-series volatility of policy rule innovations)
- **Narrative-based** (e.g., looking at counts of uncertainty words in news)

Monetary Policy Uncertainty

Influence of policy change on equity returns and long-term rates is well-established

Effects of uncertainty around monetary policy still underexplored

The Challenge

It is not obvious how uncertainty can be quantified

Proxies of uncertainty

- **Market-based** (e.g., prices of interest rate derivatives)
- **Model-based** (e.g., time-series volatility of policy rule innovations)
- **Narrative-based** (e.g., looking at counts of uncertainty words in news)
  - Prime examples: Economic/Monetary Policy Uncertainty index (**EPU/MPU**)

Narrative Monetary Policy Uncertainty

A methodology to construct a measure with desirable properties

Narrative Monetary Policy Uncertainty

A methodology to construct a measure with desirable properties

Computational and Semantic Properties

1. Quantifies level of uncertainty per article (not just presence)
2. Can produce reliable uncertainty levels even when data/articles are scarce
3. Can be computed daily
4. Captures uncertainty about future outcomes (context-aware)

Narrative Monetary Policy Uncertainty

A methodology to construct a measure with desirable properties

Computational and Semantic Properties

1. Quantifies level of uncertainty per article (not just presence)
2. Can produce reliable uncertainty levels even when data/articles are scarce
3. Can be computed daily
4. Captures uncertainty about future outcomes (context-aware)

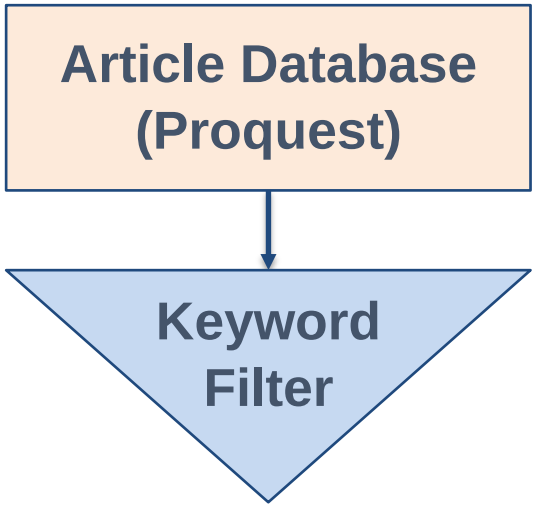
Statistical Properties of Index

1. Correlates with well-established measures of narrative uncertainty
2. Is responsive to monetary policy announcements
3. Responds to shifts in macro fundamentals at lower frequencies (e.g., CPI, unemployment)

US: 1983-2023
Canada: 1998-2023

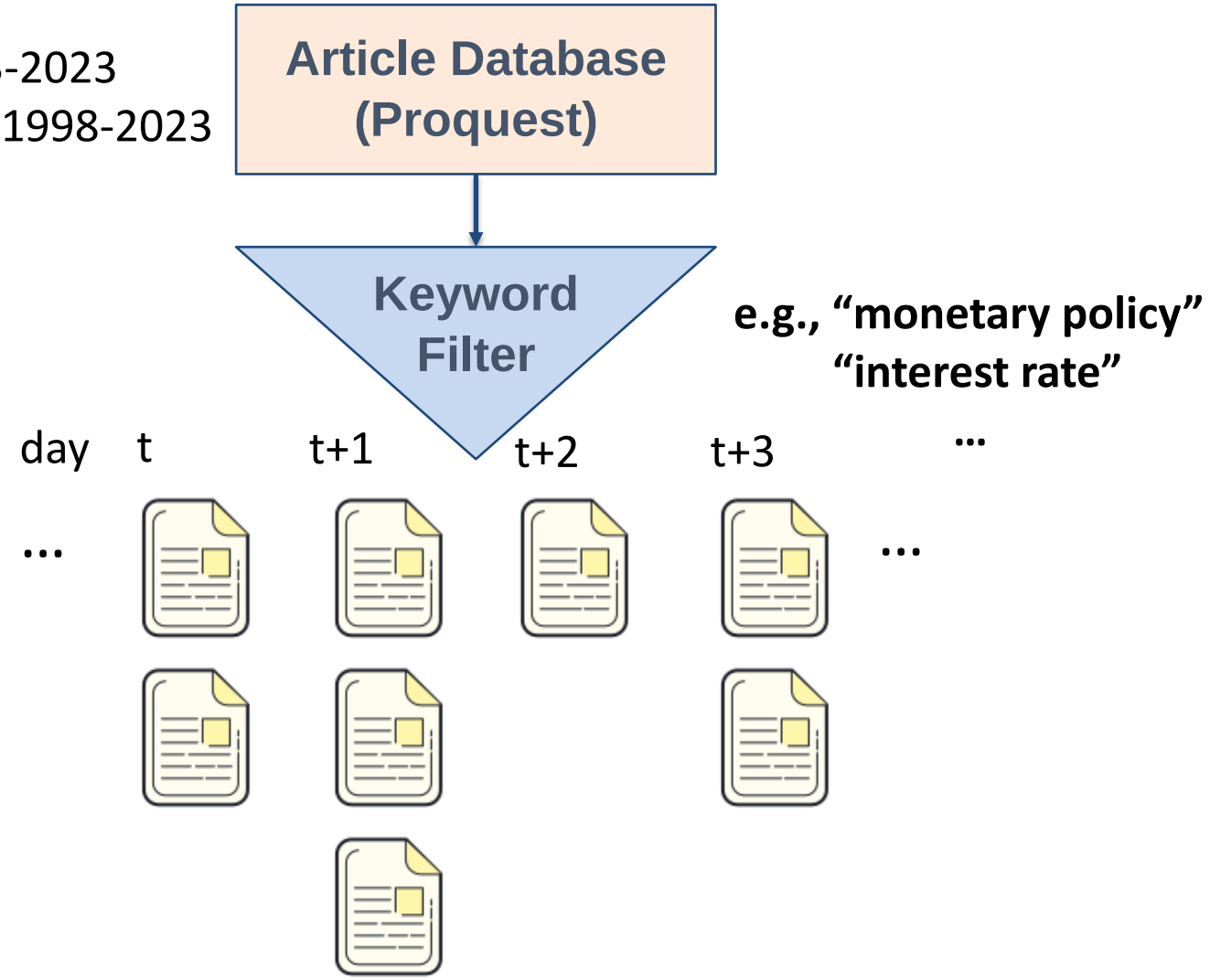
Article Database (Proquest)

US: 1983-2023
Canada: 1998-2023

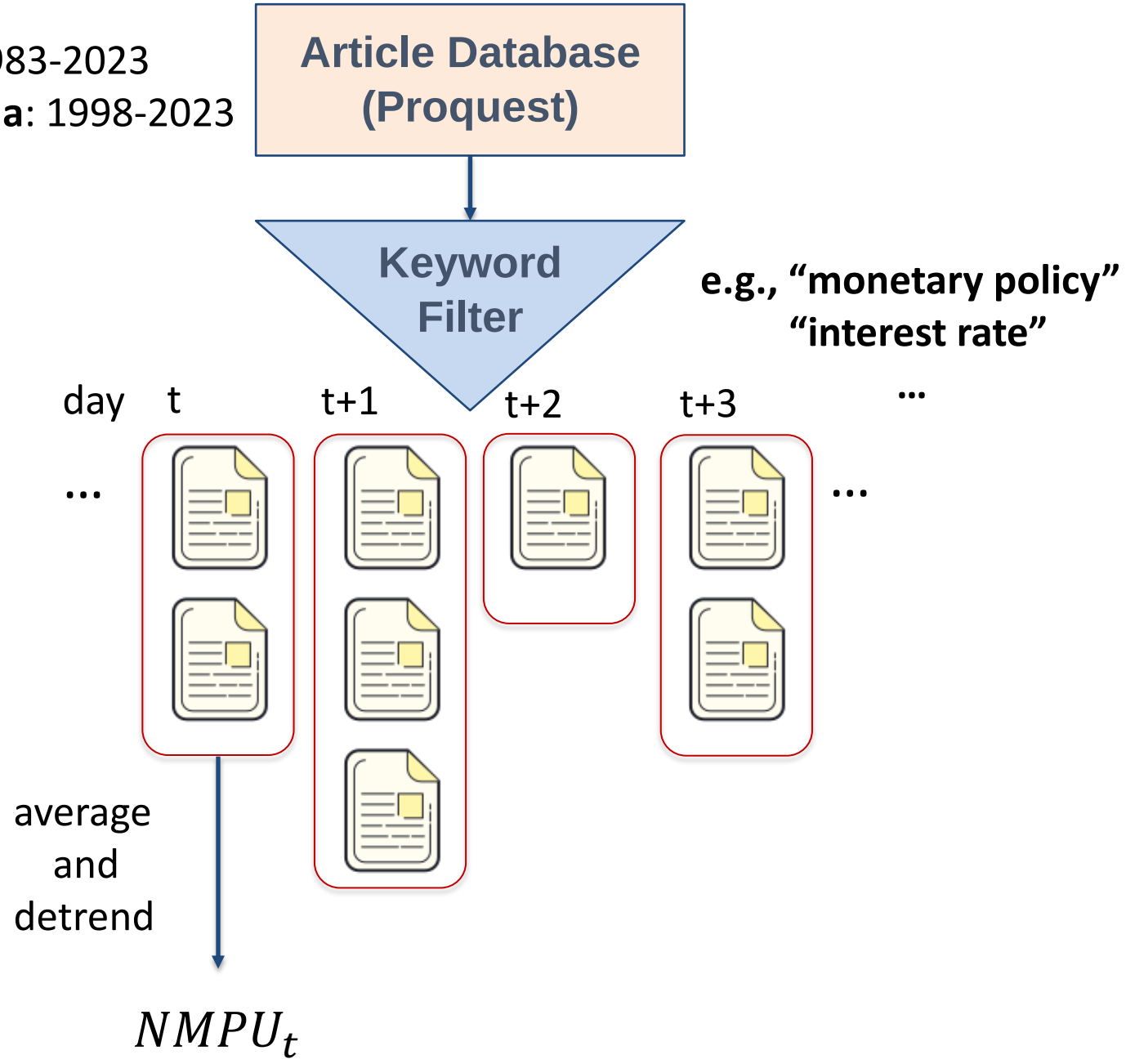


e.g., “monetary policy”
“interest rate”
…

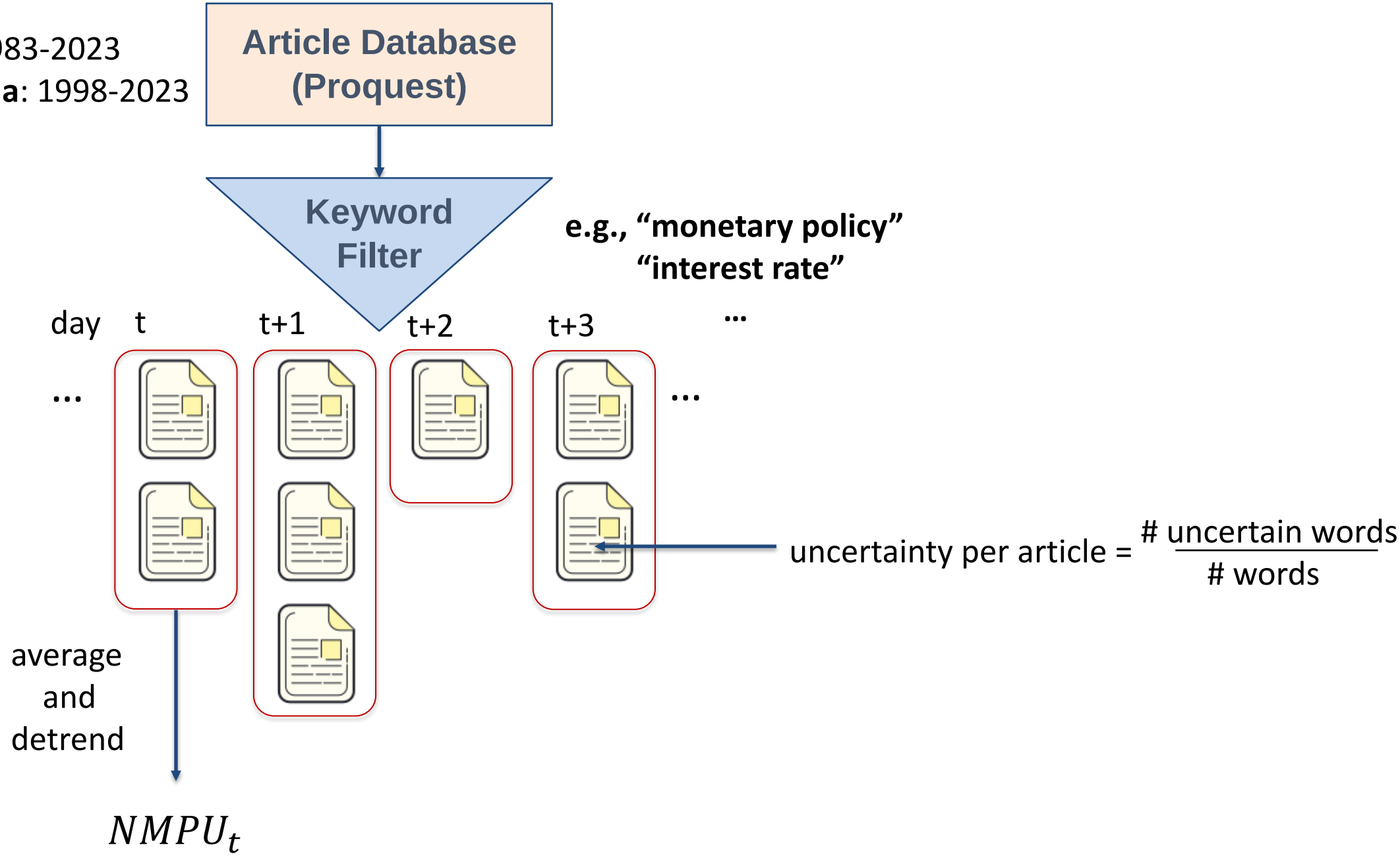
US: 1983-2023
Canada: 1998-2023



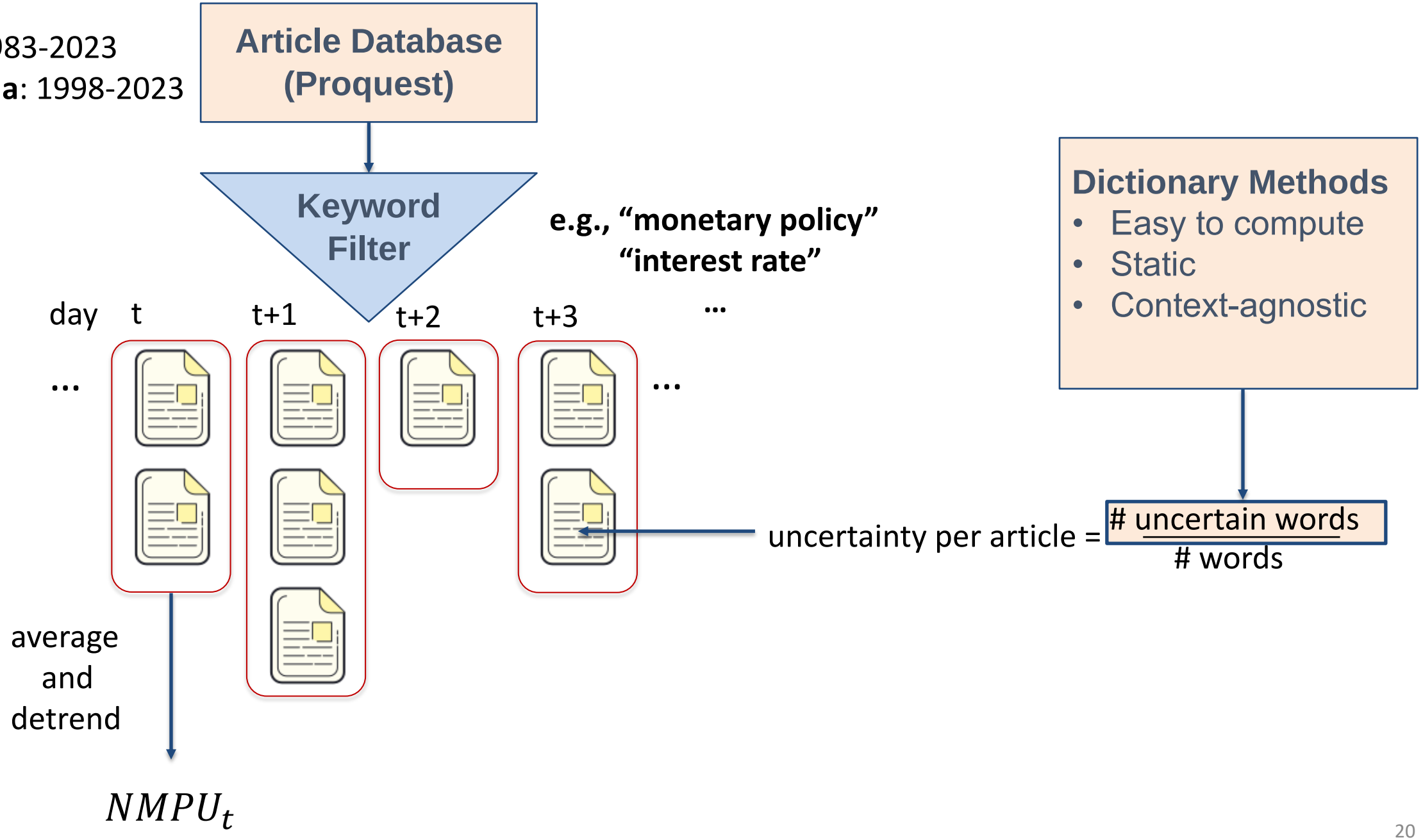
US: 1983-2023
Canada: 1998-2023



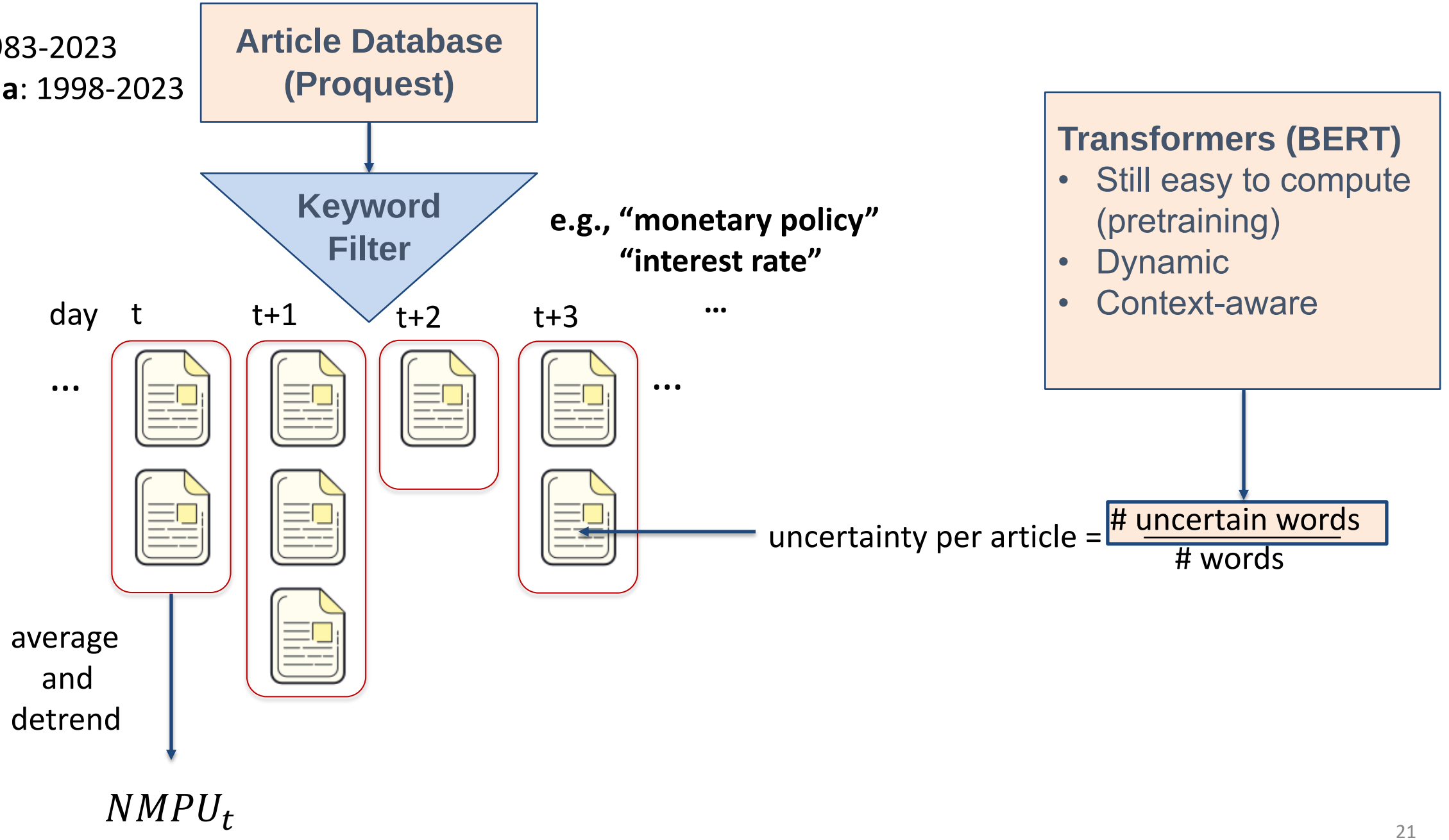
US: 1983-2023
Canada: 1998-2023



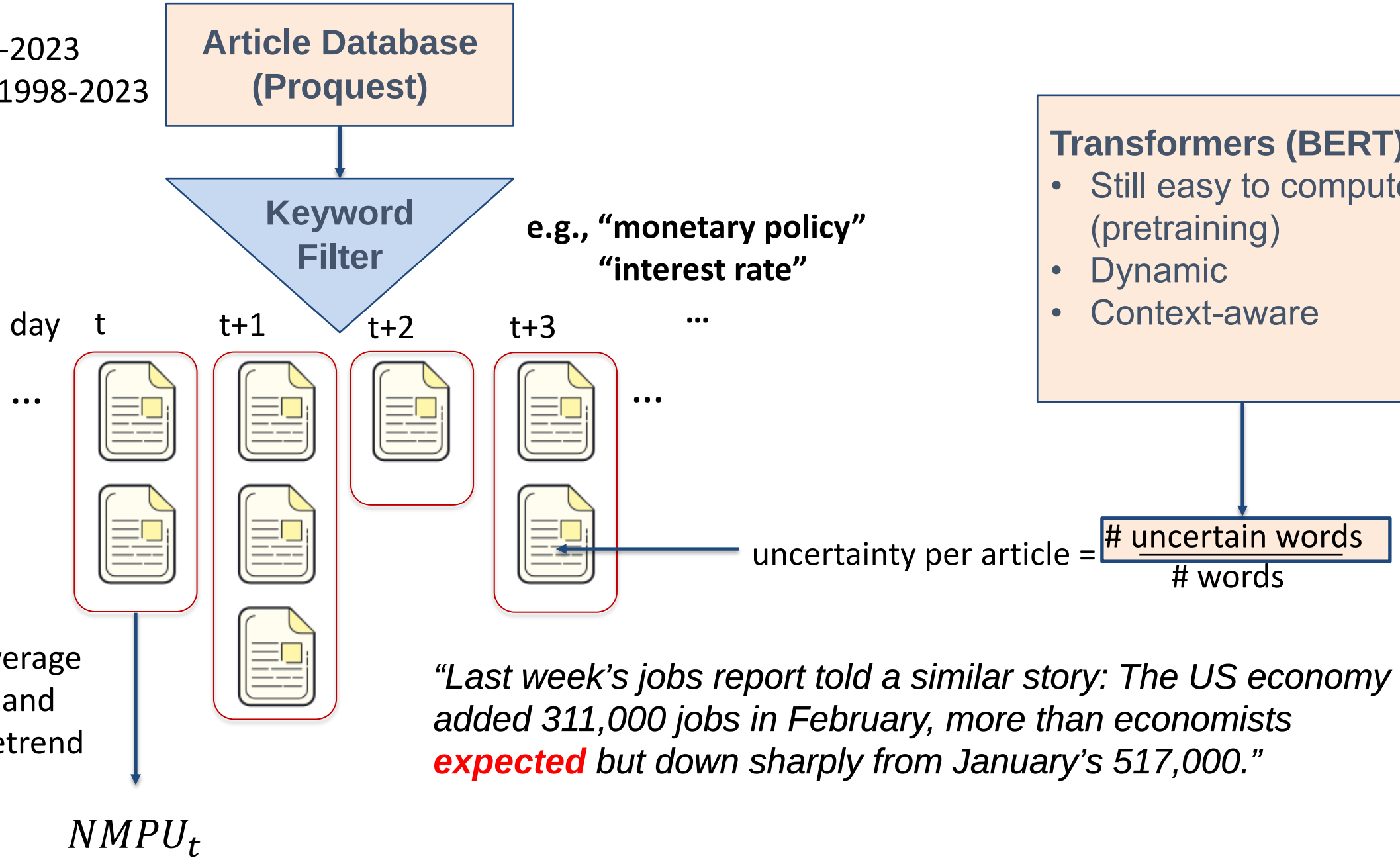
US: 1983-2023
Canada: 1998-2023



US: 1983-2023
Canada: 1998-2023

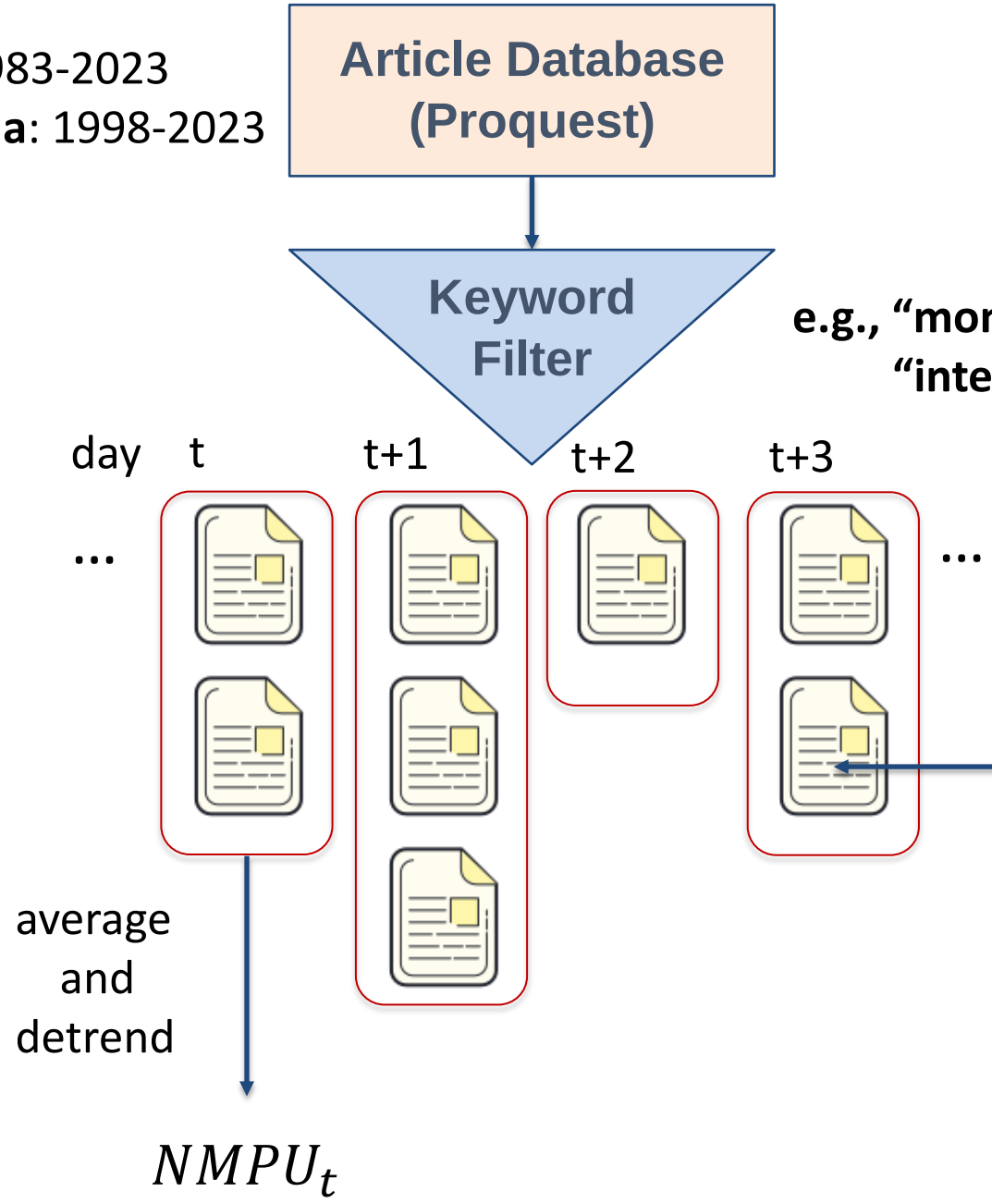


US: 1983-2023
Canada: 1998-2023



"Last week's jobs report told a similar story: The US economy added 311,000 jobs in February, more than economists ***expected*** *but down sharply from January's 517,000."*

US: 1983-2023
Canada: 1998-2023

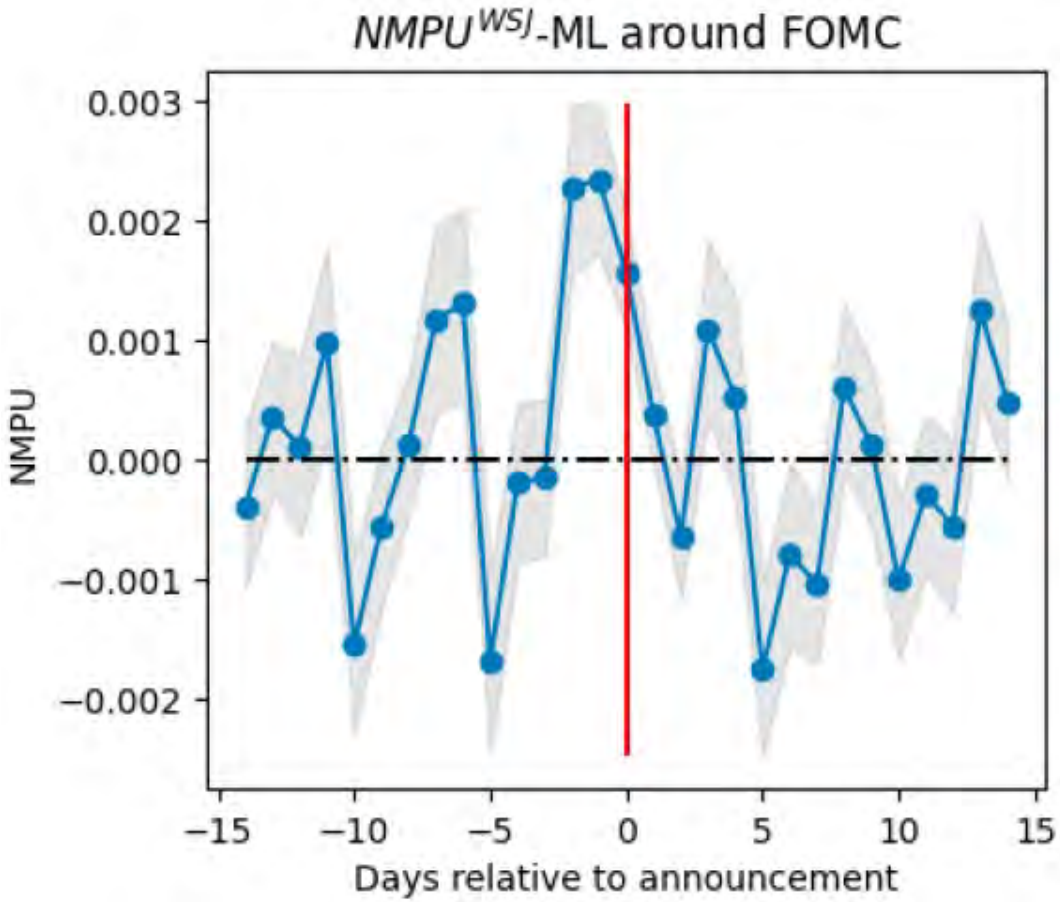


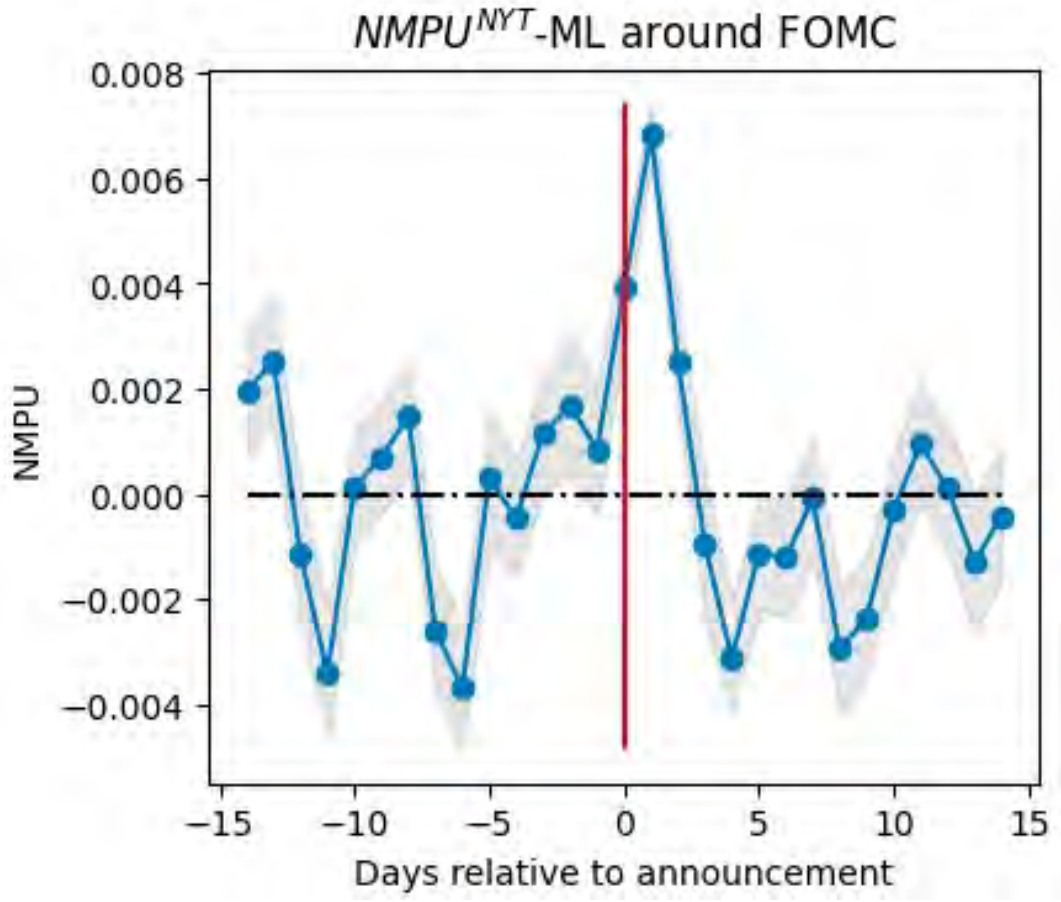
Transformers (BERT)
- Still easy to compute (pretraining)
- Dynamic
- Context-aware

uncertainty per article = $\frac{\text{\# uncertain words}}{\text{\# words}}$

vs. EPU and MPU
- They apply binary logic per article
  - Uncertainty is either present or not
  - Require lots of articles
  - Not easy to obtain at higher frequency

NMPU and FOMC announcements





1. NMPU builds up before FOMC announcements
2. Lowered on the average after announcements
3. Peaks close to announcement date (varies between WSJ and NYT)

NMPU and FOMC announcements

$$\text{NMPU}_{\tau,[1,3]} = \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau,$$
$$\text{NMPU}_{\tau,[1,3]} = \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau$$

$$\Delta\text{NMPU}_{\tau,[-3,3]} = \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau \text{ and}$$
$$\Delta\text{NMPU}_{\tau,[-3,3]} = \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau$$

	WSJ				NYT				Combined			
	NMPU		ΔNMPU		NMPU		ΔNMPU		NMPU		ΔNMPU	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\mathbb{1}_{\text{Ret}_\tau<0}$	0.0003		0.0007		0.0011*		0.0000		0.0007*		0.0004	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
Ret_τ		-0.0392**		-0.0642**		-0.0802**		-0.0654		-0.0597***		-0.0648**
		(0.02)		(0.03)		(0.03)		(0.04)		(0.02)		(0.03)
$\lvert\text{Ret}_\tau\rvert$		-0.0145		0.0145		0.0419		0.0178		0.0137		0.0162
		(0.03)		(0.04)		(0.05)		(0.06)		(0.03)		(0.04)
Intercept	0.0001	0.0005*	-0.0015***	-0.0012***	0.0023***	0.0027***	0.0016***	0.0016**	0.0012***	0.0016***	0.0000	0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	-0.002	0.023	0.004	0.025	0.009	0.019	-0.005	0.002	0.009	0.034	-0.002	0.019
N	223	223	223	223	223	223	223	223	223	223	223	223

NMPU and FOMC announcements

$$\text{NMPU}_{\tau,[1,3]} = \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau,$$
$$\text{NMPU}_{\tau,[1,3]} = \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau$$

$$\Delta\text{NMPU}_{\tau,[-3,3]} = \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau \text{ and}$$
$$\Delta\text{NMPU}_{\tau,[-3,3]} = \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau$$

	WSJ				NYT				Combined			
	NMPU		ΔNMPU		NMPU		ΔNMPU		NMPU		ΔNMPU	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\mathbb{1}_{\text{Ret}_\tau<0}$	0.0003		0.0007		0.0011*		0.0000		0.0007*		0.0004	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
Ret_τ		-0.0392**		-0.0642**		-0.0802**		-0.0654		-0.0597***		-0.0648**
		(0.02)		(0.03)		(0.03)		(0.04)		(0.02)		(0.03)
$\lvert\text{Ret}_\tau\rvert$		-0.0145		0.0145		0.0419		0.0178		0.0137		0.0162
		(0.03)		(0.04)		(0.05)		(0.06)		(0.03)		(0.04)
Intercept	0.0001	0.0005*	-0.0015***	-0.0012***	0.0023***	0.0027***	0.0016***	0.0016**	0.0012***	0.0016***	0.0000	0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	-0.002	0.023	0.004	0.025	0.009	0.019	-0.005	0.002	0.009	0.034	-0.002	0.019
N	223	223	223	223	223	223	223	223	223	223	223	223

1. **Positive announcement-day returns lead to drop in NMPU levels in the days following**

NMPU and FOMC announcements

$$\text{NMPU}_{\tau,[1,3]} = \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau,$$
$$\text{NMPU}_{\tau,[1,3]} = \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau$$

$$\Delta\text{NMPU}_{\tau,[-3,3]} = \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau \text{ and}$$
$$\Delta\text{NMPU}_{\tau,[-3,3]} = \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau$$

	WSJ				NYT				Combined			
	NMPU		ΔNMPU		NMPU		ΔNMPU		NMPU		ΔNMPU	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\mathbb{1}_{\text{Ret}_\tau<0}$	0.0003		0.0007		0.0011*		0.0000		0.0007*		0.0004	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
Ret_τ		-0.0392**		-0.0642**		-0.0802**		-0.0654		-0.0597***		-0.0648**
		(0.02)		(0.03)		(0.03)		(0.04)		(0.02)		(0.03)
$\lvert\text{Ret}_\tau\rvert$		-0.0145		0.0145		0.0419		0.0178		0.0137		0.0162
		(0.03)		(0.04)		(0.05)		(0.06)		(0.03)		(0.04)
Intercept	0.0001	0.0005*	-0.0015***	-0.0012***	0.0023***	0.0027***	0.0016***	0.0016**	0.0012***	0.0016***	0.0000	0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	-0.002	0.023	0.004	0.025	0.009	0.019	-0.005	0.002	0.009	0.034	-0.002	0.019
N	223	223	223	223	223	223	223	223	223	223	223	223

1. **Positive announcement-day returns lead to drop in NMPU levels in the days following**
2. **and greater uncertainty resolution (change in NMPU before and after event)**

NMPU and FOMC announcements

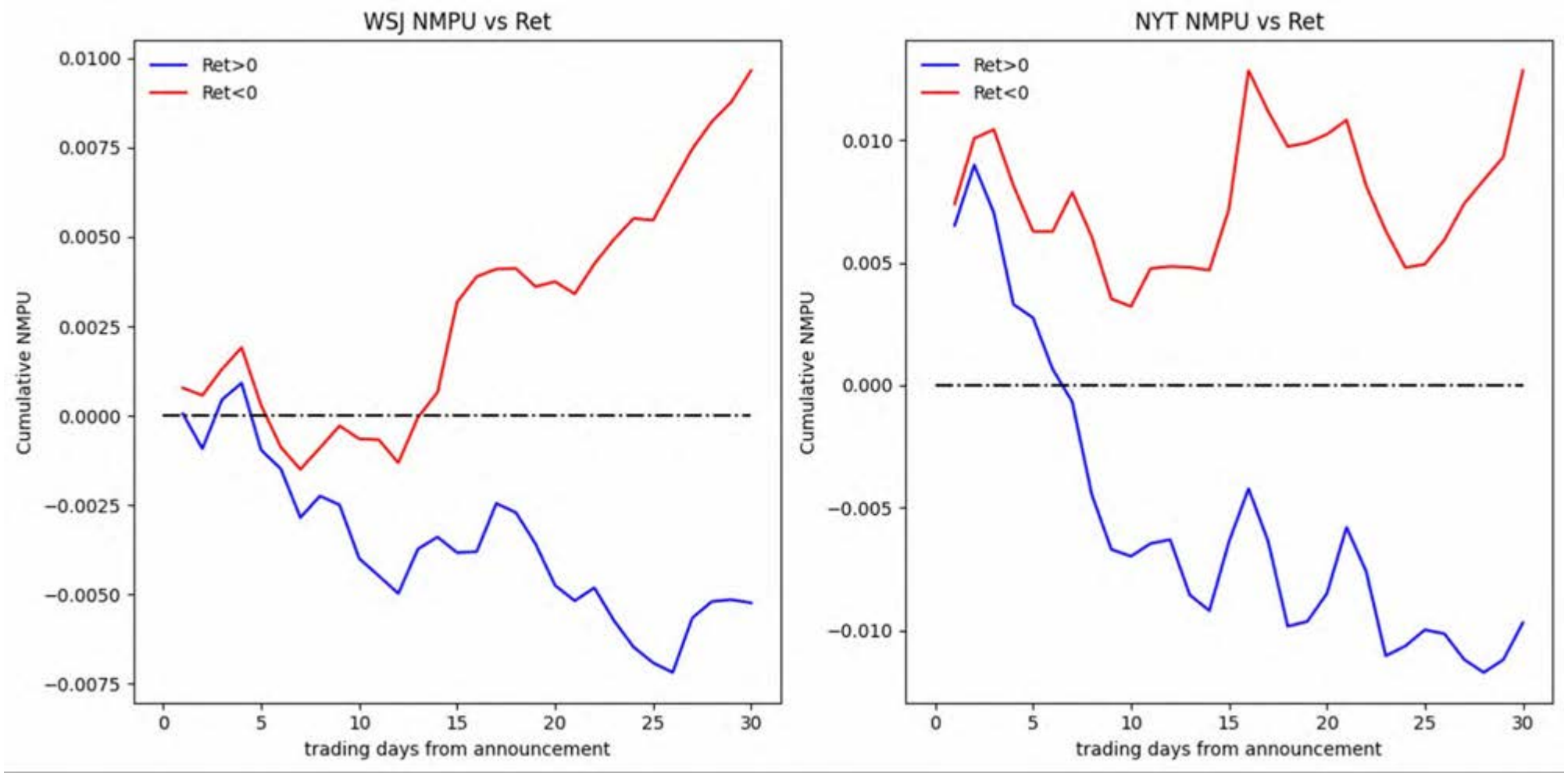
$$\text{NMPU}_{\tau,[1,3]} = \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau,$$
$$\text{NMPU}_{\tau,[1,3]} = \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau$$

$$\Delta\text{NMPU}_{\tau,[-3,3]} = \alpha + \beta_1 \mathbb{1}_{\text{Ret}_\tau<0} + \epsilon_\tau \text{ and}$$
$$\Delta\text{NMPU}_{\tau,[-3,3]} = \alpha + \beta_1 \text{Ret}_\tau + \beta_2 |\text{Ret}_\tau| + \epsilon_\tau$$

	WSJ				NYT				Combined			
	NMPU		ΔNMPU		NMPU		ΔNMPU		NMPU		ΔNMPU	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\mathbb{1}_{\text{Ret}_\tau<0}$	0.0003		0.0007		0.0011*		0.0000		0.0007*		0.0004	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
Ret_τ		-0.0392**		-0.0642**		-0.0802**		-0.0654		-0.0597***		-0.0648**
		(0.02)		(0.03)		(0.03)		(0.04)		(0.02)		(0.03)
$\lvert\text{Ret}_\tau\rvert$		-0.0145		0.0145		0.0419		0.0178		0.0137		0.0162
		(0.03)		(0.04)		(0.05)		(0.06)		(0.03)		(0.04)
Intercept	0.0001	0.0005*	-0.0015***	-0.0012***	0.0023***	0.0027***	0.0016***	0.0016**	0.0012***	0.0016***	0.0000	0.0002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	-0.002	0.023	0.004	0.025	0.009	0.019	-0.005	0.002	0.009	0.034	-0.002	0.019
N	223	223	223	223	223	223	223	223	223	223	223	223

1. Positive announcement-day returns lead to drop in NMPU levels in the days following
2. and greater uncertainty resolution (change in NMPU before and after event)
3. Bad news -> negative returns -> NMPU is elevated

Uncertainty evolution: Dovish vs. Hawkish news



1. Following negative announcement-day returns: uncertainty continues to accumulate
2. Following positive announcement-day returns: uncertainty gradually diminishes
3. Uptick just before next FOMC meeting

NMPU and Fundamentals

	Housing (1)	CPI (2)	Unemployment (3)	Fed Fund (4)	Treasury Bond (10y) (5)
F_t^{M-Q}	0.0239	-0.0090	0.0403	0.0275	-0.0372
	(0.02)	(0.02)	(0.03)	(0.08)	(0.05)
F_t^{Q-Y}	-0.0009	0.0044	-0.1025***	0.0263	0.0516*
	(0.01)	(0.00)	(0.04)	(0.03)	(0.03)
F_t^{Y-4Y}	-0.0039**	-0.0132***	-0.0279*	-0.0153	0.0429
	(0.00)	(0.00)	(0.02)	(0.01)	(0.03)
$\lvert F_t^{M-Q}\rvert$	0.0135	0.0310	-0.0734	-0.0728	-0.0442
	(0.02)	(0.02)	(0.05)	(0.09)	(0.07)
$\lvert F_t^{Q-Y}\rvert$	0.0046	0.0528***	0.1835***	-0.0395	0.0153
	(0.01)	(0.01)	(0.05)	(0.04)	(0.04)
$\lvert F_t^{Y-4Y}\rvert$	0.0093***	-0.0132***	0.0823***	-0.0224	0.0861**
	(0.00)	(0.00)	(0.02)	(0.02)	(0.04)
Intercept	0.0143***	0.0161***	0.0142***	0.0159***	0.0153***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.191	0.089	0.095	0.025	0.018
N	387	464	440	441	466

NMPU and Fundamentals

	Housing (1)	CPI (2)	Unemployment (3)	Fed Fund (4)	Treasury Bond (10y) (5)
F_t^{M-Q}	0.0239	-0.0090	0.0403	0.0275	-0.0372
	(0.02)	(0.02)	(0.03)	(0.08)	(0.05)
F_t^{Q-Y}	-0.0009	0.0044	-0.1025***	0.0263	0.0516*
	(0.01)	(0.00)	(0.04)	(0.03)	(0.03)
F_t^{Y-4Y}	-0.0039**	-0.0132***	-0.0279*	-0.0153	0.0429
	(0.00)	(0.00)	(0.02)	(0.01)	(0.03)
$\lvert F_t^{M-Q}\rvert$	0.0135	0.0310	-0.0734	-0.0728	-0.0442
	(0.02)	(0.02)	(0.05)	(0.09)	(0.07)
$\lvert F_t^{Q-Y}\rvert$	0.0046	0.0528***	0.1835***	-0.0395	0.0153
	(0.01)	(0.01)	(0.05)	(0.04)	(0.04)
$\lvert F_t^{Y-4Y}\rvert$	0.0093***	-0.0132***	0.0823***	-0.0224	0.0861**
	(0.00)	(0.00)	(0.02)	(0.02)	(0.04)
Intercept	0.0143***	0.0161***	0.0142***	0.0159***	0.0153***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.191	0.089	0.095	0.025	0.018
N	387	464	440	441	466

1. **Large changes in absolute value of housing index, CPI and unemployment -> NMPU rises**

NMPU and Fundamentals

	Housing (1)	CPI (2)	Unemployment (3)	Fed Fund (4)	Treasury Bond (10y) (5)
F_t^{M-Q}	0.0239	-0.0090	0.0403	0.0275	-0.0372
	(0.02)	(0.02)	(0.03)	(0.08)	(0.05)
F_t^{Q-Y}	-0.0009	0.0044	-0.1025***	0.0263	0.0516*
	(0.01)	(0.00)	(0.04)	(0.03)	(0.03)
F_t^{Y-4Y}	-0.0039**	-0.0132***	-0.0279*	-0.0153	0.0429
	(0.00)	(0.00)	(0.02)	(0.01)	(0.03)
$\lvert F_t^{M-Q}\rvert$	0.0135	0.0310	-0.0734	-0.0728	-0.0442
	(0.02)	(0.02)	(0.05)	(0.09)	(0.07)
$\lvert F_t^{Q-Y}\rvert$	0.0046	0.0528***	0.1835***	-0.0395	0.0153
	(0.01)	(0.01)	(0.05)	(0.04)	(0.04)
$\lvert F_t^{Y-4Y}\rvert$	0.0093***	-0.0132***	0.0823***	-0.0224	0.0861**
	(0.00)	(0.00)	(0.02)	(0.02)	(0.04)
Intercept	0.0143***	0.0161***	0.0142***	0.0159***	0.0153***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R^2	0.191	0.089	0.095	0.025	0.018
N	387	464	440	441	466

1. **Large changes in absolute value of housing index, CPI and unemployment -> NMPU rises**
2. **R^2 from 10-20% across housing index, CPI, unemployment**

Conclusion

- We present a novel methodology to capture uncertainty in news
- Initial focus on uncertainty around monetary policy (NMPU)
- NMPU is easy to compute even when data is scarce
- NMPU reacts to policy announcements (in US and Canada) and responds to shifts in macro variables

Conclusion

- We present a novel methodology to capture uncertainty in news
- Initial focus on uncertainty around monetary policy (NMPU)
- NMPU is easy to compute even when data is scarce
- NMPU reacts to policy announcements (in US and Canada) and responds to shifts in macro variables

Future Directions

- Quantify uncertainty about inflation, housing market, oil, unemployment
- Predictability study
- Instrument for future risk premia?

Conclusion

- We present a novel methodology to capture uncertainty in news
- Initial focus on uncertainty around monetary policy (NMPU)
- NMPU is easy to compute even when data is scarce
- NMPU reacts to policy announcements (in US and Canada) and responds to shifts in macro variables

THANK YOU!

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Determinants of the price of bitcoin: An analysis with machine learning and interpretability techniques[1]

José Manuel Carbó Martinez and Sergio Gorjón Rivas,
Bank of Spain

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Determinants of the price of bitcoin: An analysis with machine learning and interpretability techniques

José Manuel Carbó y Sergio Gorjón (*)(**)

Abstract

Historically, the price of bitcoin has been subject to large and abrupt fluctuations, as demonstrated once again by its sudden drop following the all-time high of $68,000 in November 2021. Since bitcoin faces significant growth prospects while largely concentrating its activity in unregulated environments, concern raises among authorities all over the world about its potential impact on the integrity of the financial system. As a result, building a sound and proper regulatory and supervisory framework to address these challenges hinges upon achieving a better understanding about both the critical underlying factors that influence the formation of bitcoin price as well as the assessment of their stability over time. In this article we analyse which variables determine the price at which bitcoin is traded on the most relevant exchanges. In order to do that, we use a flexible machine learning model, specifically a Long Short Term Memory (LSTM) neural network, to establish the price of bitcoin as a function of a number of economic, technological and investor attention variables. Our LSTM model replicates reasonable well the behaviour of the price of bitcoin through different periods of time. We then use an interpretability technique called SHAP to understand which are the most important features to the LSTM outcome. We conclude that the importance of the different variables in the formation of the price of bitcoin changes substantially throughout the analysed period. What's more, we also find that not only does its influence vary, but that new explanatory factors seem to appear often over time that, at least for the most part, remain initially unknown.

Keywords: Bitcoin, machine learning, LSTM, interpretability techniques

JEL codes: C40, C45, G12, G15

(*) The authors work in the Financial Innovation Division of Banco de España, and appreciate the comments received from José Manuel Marqués, Ana Fernández and Gabriel Pérez Quirós.

(**) The opinions and analyses expressed in this paper are the responsibility of the authors and, therefore, do not necessarily match with those of the Banco de España or the Eurosystem.

Index

1. Introduction 3

2. Crypto-asset markets: notable characteristics, latest developments and key public policy issues 5

3. Literature review 9

4. Data 11

5. Empirical exercise 13

5.1 LSTM model 14

5.2 LSTM results 18

5.3 SHAP 21

5.4 Determinants of the price of bitcoin according to SHAP 22

6. Conclusions 26

7. References 28

8. Appendix 31

8.1 Definitions 31

8.2 LSMT with lagged price 31

1. Introduction

Crypto-asset markets have been gaining increased attention from both the private and the public sector ever since its early inception in 2009. Despite the fact that growth has been uneven for most its existence, in recent years, developments seem to point in a different direction, thus featuring a consistent upward trend which brought about an expansion unheard of in previous periods. As such, market capitalization rose from merely 15 billion US dollars, in early 2017, to around 300 billion in 2020, right before the pandemic outbreak. It then skyrocketed until a peak of around 3 trillion US dollars was finally reached in November 2022. At the time of writing the market has once again bounced back at around 1,7 trillion US dollars in what yet again appears to be a steady downward race. Taking into account that the volume of bitcoin exhibits a constant growth rate, these fluctuations in capitalization are derived from the large variations in the price of bitcoin.

Notwithstanding the above, crypto-asset markets have in parallel experienced profound transformations, giving rise -among other things- to greater institutional and retail involvement. This was mainly driven by both an increased role of traditional financial institutions in certain segments and the deployment of more sophisticated investment products such as ETFs, futures contracts and other collective investment vehicles. The market has further spread to encompass other applications like Non-Fungible Tokens (NFTs). In addition, it has also supported the emergence of so-called decentralized finance (DeFi): a highly speculative niche that offers significant returns against equally great risks. As result crypto-asset markets are progressively becoming more intertwined with the formal financial and monetary system, thus amplifying their potential to spill their inherent vulnerabilities over to the economy at large.

In this context, volatility arises as a critical subject of study on which a large number of contributions exist to this date. Yet, irrespective of the many different types of available crypto-assets, the bulk of the literature focuses undoubtedly on the particular case of bitcoin. Although its total market share has dropped from 75% in 2017 to 50% in 2021, bitcoin continues to play a critical role to explain overall market trends and to trigger the development of a wealth of other initiatives either as a role model or example of problems/shortcomings that may have to be addressed to promote greater take-up. Bitcoin further echoes the fact that crypto-assets remain largely unregulated, thus it helps identify potential courses of action for regulators and supervisory authorities across jurisdictions. As a result, ascertaining the determinants of bitcoin price formation and assessing their stability over time can shed light and help steer ongoing discussions on the best way to approach increased direct and indirect exposures of critical financial market participants to crypto-assets more broadly. This knowledge will allow to establish the actual materiality of the underlying risks and consistently guide the decision on the proportionality of applicable requirements.

The objective of this article is, therefore, to analyse which variables influence the price at which bitcoin is traded on exchanges and how they behave across time. With this in mind, we were particularly mindful of two distinctive issues. First of all, the fact that unlike other financial instruments bitcoin lacks intrinsic value nor is it backed by a pool of assets like the so-called stablecoins. And secondly, there is no agreed theoretical model that explains ex-ante the determinants of the price of bitcoin. For this reason, we decided to use a flexible machine learning (ML) model, specifically, a long and short term memory (LSTM) neural network, in order to anticipate the price of bitcoin based on a series of potential explanatory

variables. This model is consistent with our goals and the underlying circumstances surrounding bitcoin in that it allows for a flexible approach which does not impose ex-ante restrictions on the relationship between the various features[1] and the price of bitcoin. Furthermore, the model can also accommodate multiple features in non-linear and non-stationary time series. To choose the features to be used in the model, we drew from the existing body of papers on the possible determinants of the price of bitcoin and picked a set of economic, technological and investor-related variables that seemed to enjoy the widest support.

Our aim is neither to build a perfect predictive model nor to develop a sound investment tool. In fact, we are not interested in analyzing how the price of bitcoin reacts to past prices, or to strongly endogenous variables (such as bitcoin's market capitalization). On the contrary, as stated before the purpose of this article is to understand if there is a set of factors that confidently explain the fluctuations in the price of bitcoin, and to analyze if these potential relationships are stable over time, all of this with a view to providing authorities with additional input in their open reflections on how to best deal with the emerging reality of crypto-assets.

With this in mind, our analysis takes place over three different periods. First, we address the time interval that falls between 2015 and mid-2017. This is the time frame where price growth seems to be the less pronounced and more stable. We then focus on the stretch comprised between 2017 and 2019, known to feature the first big price bubble in the history of bitcoin. Finally, we pay attention to the period between 2020 and 2021 with its large fluctuations and all-time price records. Our findings show that the LSTM model is able to replicate reasonably well the behaviour of the bitcoin price in all the above time segments. We then apply an ML interpretability technique called SHAP, to deepen our understanding of the features that are most important to the LSTM predictions. In this way we can identify the main drivers of the price of bitcoin at different points in time.

As a result of the above exercise, we learned that technological variables (such as hash difficulty, block size, number of transactions or unique addresses) played a decisive role in the first two periods, yet they became irrelevant in 2021. On the other hand, attention-related variables, such as searches on Google Trends, showed an increasing relevance over the years. In fact, they turned out to be the most important category in 2021. What's more, in stark contrast to part of the literature, we could not amass evidence that the SP500 and gold were ever among the main drivers of bitcoin price. We, therefore, conclude that the influence of specific variables in the price of bitcoin is largely unstable and seems to change substantially -and in ways hard to anticipate- across the different time periods observed. Not only does its influence vary, but our research suggests that new explanatory factors might also appear which oftentimes remain opaque, at least in its early stages. We believe these findings to be of relevance for that work that both local and global financial authorities are carrying out in order to help inform the design of prospective public policy actions.

We acknowledge the existence of both opportunities and challenges in choosing ML models over more traditional econometric techniques. On the one hand, as mentioned before, ML allows for a high degree of flexibility and better out-of-sample performance than

[1] In machine learning, "features" is the term used for the individual independent variables that are taken as an input to make predictions over a target variable. Throughout the article we will use the term "features" together with "factors", "drivers" and "variables" interchangeably.

traditional econometric techniques. On the other hand, the use of ML has a cost in terms of interpretability: i.e. traditional econometric alternatives are inherently interpretable, while ML models must resort to additional tools for this very purpose. Despite this trade-off, we believe that post hoc interpretability techniques, particularly SHAP, work reasonably well (Alonso and Carbo, 2022; Molnar, 2022), while also becoming more widely used in the context of ML (Albanesi and Vamossy, 2019; Chen et *al*. 2021). Therefore, on account of the particular circumstances of our study, we consider that ML advantages outweigh its potential limitations as it works better in predicting and replicating the data and interpretability shortcomings can be addressed in a practical way.

The paper is divided as follows. Section 2 explains the main features and recent evolution of the crypto-asset market, highlighting key public policy issues. Section 3 presents a literature review and underlines our contribution. Section 4 sheds light on the data that were used for this study. Section 5 is devoted to the empirical analysis, which is split into (i) predictions by the LSTM model (sections 5.1 and 5.2) and (ii) interpretation of those predictions and the determinants of the price of bitcoin (sections 5.3 and 5.4). Section 6 concludes.

2. Crypto-asset markets: notable characteristics, latest developments and key public policy issues

Crypto-assets, more frequently and imprecisely known as crypto- or virtual currencies, are a new type of private asset that depends primarily on cryptography and DLT (or an equivalent technology) as part of its perceived or inherent value (FSB, 2018). As such, crypto-assets embrace a diverse set of representations whose main common trait is, precisely, the fact that they usually combine several distinctive elements from a wide range of financial instruments. As a result, crypto-assets typically emerge as hybrid products that rise both significant conceptual and practical challenges in trying to accommodate them into any pre-existing legal framework[2] (Foz, 2021)

Despite the potentially manifold interpretations that may stem from the above constraints - both in relation to their nature and effective scope-, from a technical standpoint crypto-assets seem to exhibit a number of prominent and prevailing attributes at all times: (i) the deployment of a decentralized ledger that enables new ways of issuance, registration and exchange of underlying assets, (ii) the application of protocols which set out the rules by which the execution of transactions is to take place, and (iii) the emergence of a complex ecosystem featuring a wealth of participants that take up different roles as regards the distribution, validation, trading, transfer and storing of digital assets.

It is now commonplace to consider that the inception of the crypto-asset markets was a by-product of bitcoin's outbreak. However, its theoretical foundations can be traced back to the works of earlier authors like Wei Dai (1998) and Nicholas Szabo (2005) among many others. These researchers were keen on highlighting the critical part played by trusted

[2] Regardless of the growing number of initiatives that intend to provide a specific regulatory framework for crypto-assets, oftentimes where not yet in place other approaches apply. Thus, some jurisdictions are trying to address this regulatory gap by extending to certain aspects of crypto-assets' operations currently applicable requirements on account of the similarities with other types of regulated activities; e.g. payment services or financial instruments.

intermediaries (e.g. central banks and/or credit institutions) in securing the proper functioning of traditional payment circuits and yet, how this very circumstance was prone to triggering other types of negative consequences for the society as a whole[3].

In an attempt to address these challenges, both authors advocated for the establishment of an alternative monetary system based on distributed networks and cryptographic proofs. They further supported the deployment of tools that could help ensure a more predictable growth of the underlying settlement asset. However, it was indeed the actual launch of bitcoin that first provided a material example of a practical type of money holding the promise of avoiding an indigenous loss of its value over the long term (Nakamoto, 2008).

Thus, bitcoin's ultimate purpose was that of becoming a credible alternative to fiat money for which it leverages a novel electronic payment system that supports peer-to-peer exchanges of value in an allegedly safe manner. According to its proponents, such particular circumstance further facilitates the deployment of affordable and irreversible payments since transaction costs are lowered on account of forgoing the role traditionally assigned to intermediaries in all regular payment channels but cash (Conesa, 2019).

In order to achieve this goal, bitcoin hinges on a network of users -nodes- which rely on the Internet to communicate with one another on the basis of a common protocol by masking their respective real-world identities under public keys. Said protocol is widely accessible to any interested parties in its condition as open software.

In the interest of avoiding the so-called double spending problem, i.e. the possibility that the same units of crypto-assets are sent to more than one recipient at once, bitcoin marshals an inspired solution: each and every network participant enjoys full visibility over all the transactions that have taken place across them. Moreover, they all are entitled to act as potential validators. Hence, every prospective operation ends up being published in a decentralized repository and awaits final validation with the concourse of any number of parties that decide to join the respective process. The various nodes will also serve as a common database, keeping a complete or partial copy of the historic record of all executed transactions at any given point in time.

With a view to providing a consistent and unified picture of said repository and so, ensuring that it reflects the legitimate consent, bitcoin has chosen to rely on a computationally costly verification procedure. Nodes agreeing to voluntarily perform the role of validators (miners) will, therefore, have an option to be compensated for their efforts both with newly created units of crypto-assets[4] and by cashing in user fees. Yet, rewards will only be then granted to the node that first demonstrates to have rightfully solved the cryptographic puzzle posed by the algorithm.

Completing this task successfully demands a great amount of resources as it is basically a trial and error process that aims at brute-forcing a given result. Thus, in principle, there is

[3] According to these authors, salient examples of potential negative implications of intermediation include hyperinflation as well as the potential for authorities to tax or regulate social and economic activity through the threat of force.

[4] This amount is cut in half at specific points in time. It is the result of a scheduled event, known as halving, whose goal is to ensure that the maximum supply of bitcoins does not exceed a fixed referenced volume (21 million) in order to mimic the finite quantity of physical gold. By design, the number of bitcoins minted per block is reduced by 50% after every 210.000 blocks which is equivalent to about once every four years. This helps keep the pace of growth predictable. Halving seems to have some interesting knock-on effects on its market price and the behaviour of other markets (El Madhy, 2021; Meynkhard, 2019).

some randomness over which node will finally achieve this accomplishment[5]. Transactions pending validation are collected in a block which is then trimmed down to form a 256-bit block hash value. Once a block has been mined correctly it will reference the previous one, forming an unbroken cryptographic chain back to the first bitcoin block (Brühl, 2017). This setup helps warrant that ultimately everyone agrees on the transaction record. It furthermore makes it highly unlikely that anybody could tamper with blocks in the chain since re-mining all the following blocks would be computationally unfeasible.

Bitcoin's positive reception paved the way for an expansion of the crypto-asset's ecosystem which led to the emergence of a wide range of initiatives, each one with different goals and purposes. Hence, in addition to providing new means of exchange, some novel crypto-assets have proven to be highly useful in raising funds for innovative and risky projects. Likewise, others have been shaped in such a manner that they resemble an equity in the stock of a company, further furnishing their holders with voting, subscription or appraisal rights as well as with dividends and other entitlements. Eventually, certain token's main functionality is to grant their users access to products and/or services developed and distributed by its issuer over its own technical platform (utility tokens).

From the point of view of their respective users, crypto-assets are oftentimes leveraged to cater for payments, hedging or speculative purposes in a manner that is somewhat disconnect from its original intent. Along the same lines, some tokens are mainly conceived as a symbolic gesture or statement such as Dodgecoin or Jesus Coins (Kim et *al.*, 2018). In addition, so-called non-fungible tokens (NFT) have been gaining momentum lately in that they have proven particularly useful to attest the scarcity and provenance of rare assets – e.g. original artwork, pictures, collectibles, or trademarks, among many others-. This circumstance is creating new and significant opportunities to trade unique digital goods[6].

The compounded effect of the above developments alongside the consolidation of its underlying technology proposition as well as the expansion of decentralized finance (DeFi) have all caused a visible surge of crypto-asset markets. The additional impact of the recent lockdown due to the COVID-19 outbreak has significantly amplified this trend. As a result, the market has furthered its size and reached levels previously unheard of. In fact, market capitalization of crypto-assets exceeded USD3 trillion in 2021, a 400% growth in less than 12 months that can be largely explained by bitcoin and ether[7]. Moreover, both these crypto-assets ranked among the world's top 20 traded assets (Iyer, 2022).

In terms of price, bitcoin stood out as a particularly volatile crypto-asset in the period 2020/2021. At its peak, it experienced an almost 200% increase in comparison to its point of departure in January 2021. This was also about five times greater than its previous record. Other crypto-assets such as ether or binance coin followed suit, yet their rate of growth was far less pronounced. In parallel, stablecoins also saw a significant expansion, being the most traded asset class within the crypto-asset ecosystem. Their current market share stands at around 7%.

[5] Yet, miners with most efficient hardware may potentially benefit from better hash rates which is likely to increase their probability of successfully mining a block and so, capture the reward.

[6] NFTs are tradable rights to digital assets, where ownership is recorded in smart contracts on a blockchain. Due to its lack of fungibility they are mainly intended as pure assets (Dowling, 2022).

[7] Together, bitcoin and ether account for about 2/3 of the market's capitalization value.

The latest episode in the expansion of crypto-asset markets shows a number of distinctive characteristics that help tell it apart from any other previous phases. Despite the substantial role played by retail investors in driving up both volumes and values, institutional investors –including hedge funds- have been particularly active in this field[8]. Consequently, both the average size of trades and the maturity of associated portfolios has increased accordingly. Similarly, the demand for crypto-assets has soared among large IT corporations that now consider them as a useful tool for the treasury management strategies. Despite still being of a relatively low importance in terms of size, these movements seem to have had a noteworthy signaling effect on other potential investors, eventually giving rise to a bubble (Shu et *al.*, 2021).

This period has also seen a rise in the number of crypto-asset related custody and trading services that are being offered by traditional financial institutions while also featuring their mounting integration in well-established and widely-used private payment networks such as those commonly associated with cards. Again, this latter aspect is of great relevance for the future prospects of crypto-assets in that they may find it easier to enter and consolidate in the broader financial system at a faster pace.

The transformation of the crypto-asset markets is further augmented by the deployment of more sophisticated and complex investment vehicles such as, e.g. derivatives, futures or ETFs, which can help stimulate the appetite of a wealth of investors that still lay off boundary due to the uncertainties and limitations surrounding these markets.

From the point of view of financial authorities, all these changes signal the true potential of crypto-assets to become a critical element of both the financial and economic blood circuit of the society at large. They also highlight the sheer size of challenges that they need to face speedily in order to safeguard the orderly functioning of both the system as a whole and of its underlying parts. A larger footprint of crypto-assets is, thus, seen as a potential source of distortions for both financial stability and monetary policy (ECB, 2019) while also giving rise to other relevant concerns such as the effectiveness of consumer protection mechanisms (Australian Parliament, 2021), threats to financial integrity (FATF, 2014) as well as their foreseeable negative environmental impact (Moshin et *al.*, 2020).

In this context, the distinctively high volatility of crypto-assets alongside an increased reliance on leverage trading strategies (IMF, 2021) emerge as the two chief elements which largely fuel international discussions. Aspects such as their ultimate influence on market, credit, liquidity risks and how well are these being managed are, therefore, core issues for authorities to carefully look into[9]. As a result, they are growingly expected to design and deploy the most appropriate safeguards in defense of the public interest while still succeeding in keeping alive most of the promised benefits that the development of these markets could bring about.

For this purpose, authorities are presently engaged in a comprehensive and globally coordinated exercise to review thoroughly applicable regulatory and supervisory frameworks so as to decide when and how to adapt current rules and standards (e.g. Basel III) and when to complement those with novel ones (e.g. MiCA). This approach implies

[8] PwC & Elwood estimate that, between 2019 and 2020, the size of crypto-asset related AuM held by hedge funds doubled, reaching a scale of USD 2.000 million. Other research seems to confirm the growing relevance of institutional investors for the bitcoin market as well: e.g. according to data by Chainanalysis, in December 2021, bitcoin-related transactions below USD 10.000 fall by 22% whereas those with a nominal amount above USD 1 million increased by 32% in the same period.
[9] Notwithstanding other equally relevant aspects such as those related to increased operational risks due to the maturity of the technology as well as associated legal and reputational risks.

combining traditional measures -including capital, liquidity and leverage requirement- with other types of actions such as an enhanced monitoring of market behavior or promoting improved transparency, yet their proper calibration demands to deepen their understanding over which elements and forces most likely shape crypto-asset markets' behaviour.

One such fundamental aspect is the determination of triggers of price as well as the assessment of their stability over time. This knowledge will be crucial to gauge the actual materiality of the underlying risks and consistently decide on the proportionally of applicable requirements. In the next sections, we try to address some of these questions by applying a set of novel tools.

3. Literature review

Bitcoin has been of interest to both the general public and the academia since its inception, when Nakamoto wrote his seminal white paper (2008) and launched the cryptocurrency in January 2009. Due to this growing attention, there are several branches of academic literature dealing with topics that concern bitcoin. On the one hand, there is a debate on whether bitcoin could be considered as a safe haven asset, a hedge, a diversifier, or just a speculative asset. For more on this, see Bouoiyour and Selmi (2015), Bouri et al (2017), Urquhart and Zhang (2019) and Bariviera and Merediz (2021). On the other hand, there is a wealth of papers that focus on examining what variables could influence the price and return of bitcoin. This particular topic is seeing a substantial increase in the number of actual contributions which, oftentimes, exhibit very heterogeneous results. Our work is framed within this broad universe.

Bitcoin has been active since 2009 which means that existing articles address different time periods and employ very different methodologies. While some of them rely on traditional time series like ARIMA or Vector Error Corrections, others choose machine learning methods instead, like boosting or neural networks. In order to help better understand our contribution to the literature, we resolved to structure and cluster the former into three distinctive types of debates: (i) the role of economic/financial variables such as exchange rates or economic indicators, (ii) the relevance of technology and its many blockchain-specific underlying features, and (iii) the role of gold and the degree of attention/interest raised by crypto-assets among investors and the general public.

The first line of thought explores whether macroeconomic indicators or the most relevant exchange rates could explain the price of bitcoin. Using ARDL and Bayesian quantile regression, Bouoiyour and Selmi (2015; 2017) found that the Shanghai market and macroeconomics events -like Brexit or Venezuela currency demonetization- indeed exercised visible influence. Cermak (2017) leveraged GARCH models to conclude that macroeconomic conditions in China, USA, and the EU could affect bitcoin's price volatility. Moreover, using Vector Error Correction, Zhu et *al*. (2017) established that the US Dollar Index, the Dow Jones, and the Federal funds had a long term impact on the price of bitcoin.

Over and above, the critical role of the SP500 is further emphasized in three additional papers. Sovbetov (2018) resorted to an ARDL in order to infer that the SP500 had a weak and positive effect in the long run. Along these lines, Kjarland et *al*. (2018), discovered that the SP500 was among the most important variables determining the price of bitcoin. In addition, Kapar and Olmo (2020) used VECM to understand the dynamics of bitcoin returns

and concurred as well with the fact that the SP500 had positive correlation with bitcoin returns. The potential relevance of other economic factors, like return to investment, was explored by Alessandretti et *al*. (20219) via Boosting and LSTM who found it to be the most crucial factor determining the price of bitcoin. Finally, Chen et *al*. (2021) combined LSTM and random forest and found that both the SP500 and the NASDAQ, have a positive influence on bitcoin price.

Conversely, within this literature, there are also papers that question the importance of some of the aforementioned factors. For instance, leveraging ARDL and VECM, Ciaian et *al*. (2016) and Cian et *al*. (2018) posited that macroeconomic and financial variables do not have a long-term impact on the price of bitcoin. Similarly, Baur (2018) ascertained that the price of bitcoin is uncorrelated with traditional asset classes such as stocks, bonds and commodities both in normal times and in periods of financial turmoil. Finally, Pyo and Lee (2019), found that price changes in bitcoin after macroeconomic announcement are insignificant.

On the point of technology, the role of the hash rate and the difficulty of the algorithm (see appendix for definitions) seem to be overly present in the literature. Some papers find evidence that the competition in the network of producers, the rate of unit production, and the difficulty of the algorithm used to "mine" the "cryptocurrency" could be influencing the price of bitcoin -see, e.g. Kristoufek (2015), Ciaian et *al*. (2016), Bouoiyour and Selmi (2017) or Hayes (2018)-. Other research considers that the role of hash rate or hash difficulty is negligible for determining said price. For example, Fantazzini and Kolodin (2020) found that there is neither evidence of Granger-causality or cointegration between hash rate and the price of bitcoin. Kjarland et *al*. (2018) established that the hash rate is irrelevant for modeling bitcoin price dynamics.

Finally, there seems to be wide consensus on the positive impact of gold, investor attention and online searches in the evolution of the price/return of bitcoin. Bouoiyour and Selmi, (2017), Panagiotidis, (2018) as well as Kapar and Olmo, (2020) all highlighted the importance of gold. Kristoufek (2013), Kaminski (2014), Kristoufek (2015), Ciaian (2016), Kim et *al*. (2016), Sovbetov (2018), Panagiotidis (2018), Lyócsa et *al*. (2020), Chen et *al*. (2020) and Zhu et *al*. (2021) drew similar conclusions about the role on investor attention. For this purpose, they measured it either as the number of tweets, the number of online searches on Google, or the number of enquires in Wikipedia.

Some of our findings are in line with this literature. As we show in section 4.4, investor attention, proxied by the volume of both searches in Google Trends and mentions in Twitter, is one of the most important variables influencing the price of bitcoin. Regarding the hash rate, algorithm difficulty and other technology-related variables inherent to bitcoin, we gather that they exhibited a strong influence on the price of bitcoin both in 2017 and 2018, but not any longer in 2021. Finally, concerning gold and macroeconomic indicators like the SP500, we conclude that they never rank among the top factors that explain the evolution of the price of bitcoin.

Our paper is mostly related to Panagiotidis et *al*. (2018), and Chen et *al*. (2021). The former is entirely focused on the drivers of bitcoin rather than in predicting, while the later also uses LSTM and interpretability techniques to uncover the determinants of the price of bitcoin. Our contribution is twofold. We first identify which are the potential explanatory elements in the evolution of the price of bitcoin at different points of its lifecycle, including the most recent spike of 2021. In this way, we are able to provide evidence of whether these different

variables remain the same or change and are replaced by others. We further distinguish whether other, still unknown factors at the time of writing, emerge which are not captured by our model.

In addition, contrary to the approach followed by Chen et *al*. (2021), we address the issue of interpretability of the results in a different way. Instead of leveraging feature importance to uncover the relevance of each of the underlying variables to the LSTM method, we rather rely on SHAP, a global interpretability technique, that might be better suited when using correlated features and when trying to interpret neural networks (Alonso and Carbó, 2022). Moreover, unlike Chen et *al*. (2021), we try to avoid highly endogenous variables among the potential determinants, like lagged values of bitcoin price or market capitalization.

4. Data

In accordance with most the existing literature, we consider three different types of potential explanatory factors linked to the price of bitcoin: (i) the specific technology features of bitcoin, (ii) the evolution of the economy, and (iii) the level of attention/interest it arises among the public at large. Starting with the technology dimension we took into account the following elements: difficulty in finding the hash, unique addresses, commissions to miners (fees), hash rate, sum of blocks, average block size, sum of transfers, and average transfer size[10].

Regarding the economic variables, we chose to include the following ones: the price of gold and oil (separately), the SP500, the FTSE, the DOW30, the NASDAQ, and the exchange rate of several international currencies (i.e. the Euro, the British Pound, the Yuan, the Yen and the Swiss Franc) and the US dollar. Finally, as a proxy for the level of public attention we placed our focus on (i) how the search term “bitcoin” was captured in Google Trends, and (ii) the number of Tweets per day that were published with “bitcoin” as the distinctive hashtag[11] [12]. We further considered using specific term searches in Wikipedia but we discarded this approach soon as it did not enhance the information already provided by Google Trends and Twitter. While agreeing that a larger selection of Google trends’ keywords / different Twitter hashtags may help better capture the public's degree of attention or interest in bitcoin, we feel that our approach already partly addresses this concern. In fact, when searching for “bitcoin” on Google trends, we performed this search "as a topic". This implies that Google’s search engine seeks result worldwide, regardless of underlying language. Furthermore, it takes into account other searches that are considered to be related to the "topic" of bitcoin according to Google trends (Carbó and García, 2021). In any case, to more holistically reflect the actual interest of investors on bitcoin, more specialized searches could be performed the choice of the option “related topics or related searches” of Google Trends or, alternatively, using different keywords (cryptocurrency, bitcoin-usd, etc.) as in Aslanidis et al (2022) and Urquhart (2018). All this data was collected from Coinmetrics (technological factors), Yahoo Finance (economic

[10] We include in the appendix a section in which we explain in detail each of these possible determinants

[11] Due to the way in which we have obtained the data, we do not know if the intensity of searches on Google and the tweets have been generated by a specialized audience (investors) or a more general one. That is why from now on we will refer to these variables broadly as public attention variables.

[12] The information from Twitter corresponds to any tweet that has the hashtag “bitcoin”, so it can be from anywhere in the world in any language. The information from Google Trends is also collected globally and in any language

factors), Google (Google Trends), and Bitinfocharts (Number of Tweets). In total, we relied on 25 features to determine the price of bitcoin.

We worked with daily frequencies in our set of data (business days only) from January 2015 to July 2021. We further divided the sample into three periods of time. We called the first slot *launch period* (from 1 January 2015 to 1 April 2017) It distinctively shows a steady growth in the price of bitcoin. The second slot was referred to as the *expansion period* (1 January 2016 to 1 April 2018). It features the first spike in the price of bitcoin, namely in December 2017, when it topped 20.000 US dollars. The third and last slot goes from 15 June 2019 to 15 June 2021. This was branded as the *consolidation period* and runs from early post-pandemic days until reaching the heights of the price of bitcoin in March and April 2021 (i.e. 60.000 US dollars.

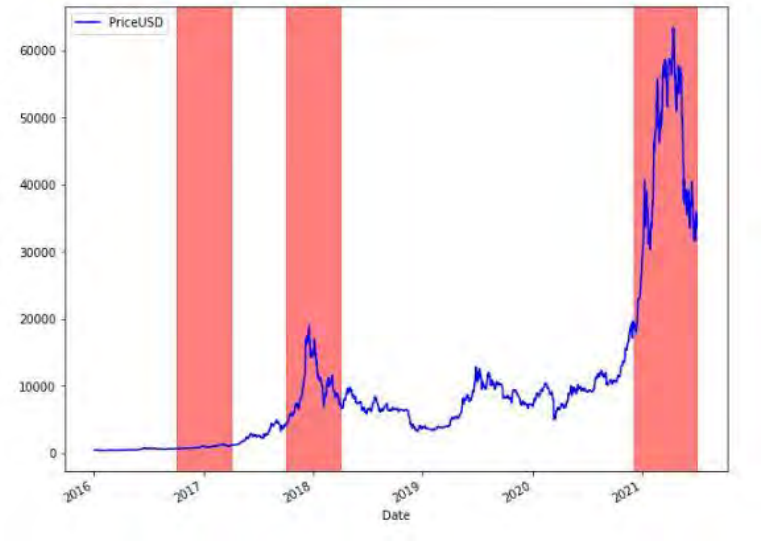
For each period, we trained, validated and tested our LSTM model, determining the target variable (i.e. bitcoin exchange price as expressed in US dollars) as a function of the 25 features considered. **Table 1** present information about their precise duration and the training, validation and test samples. The exact start and end dates were chosen at our discretion so as to ensure that the test samples had a minimum duration of six months and reflected the key events that interested us: i.e. the constant growth at the end of 2016, the first peak in the price of bitcoin at the end of 2017, and its all-time highs during the first months of 2021. In addition, when choosing the dates, we made sure to count at least 18 months of data for training and validation prior to the test samples. We carry out two predictive exercises. The main one, which we call a long-range forecasting exercise, in which we produce a single forecast for the entire test sample (six-month forecast). And a second approach, which we refer to as a rolling window exercise, in which we deliver forecasts over a sample of one month, and move the forecasts week by week. More on this in section 5.1. **Figure 1** depicts the test samples of each period. We also checked the robustness of our results by selecting different start and end dates in the interval contained four weeks before and four weeks after the original start and end dates.

Table 1. Time intervals for the three periods

Period	Training and validation sample	Test sample
Launch	01/01/2015 to 01/10/2016	01/10/2016 to 01/04/2017
Expansion	01/01/2016 to 01/10/2017	01/10/2017 to 01/04/2018
Consolidation	15/06/2019 to 23/12/2020	23/12/2020 to 15/06/2021

Source: Own authors (2022)

Figure 1: Bitcoin price evolution and test samples of each period



Source: Author's own calculations (2022).

5. Empirical exercise

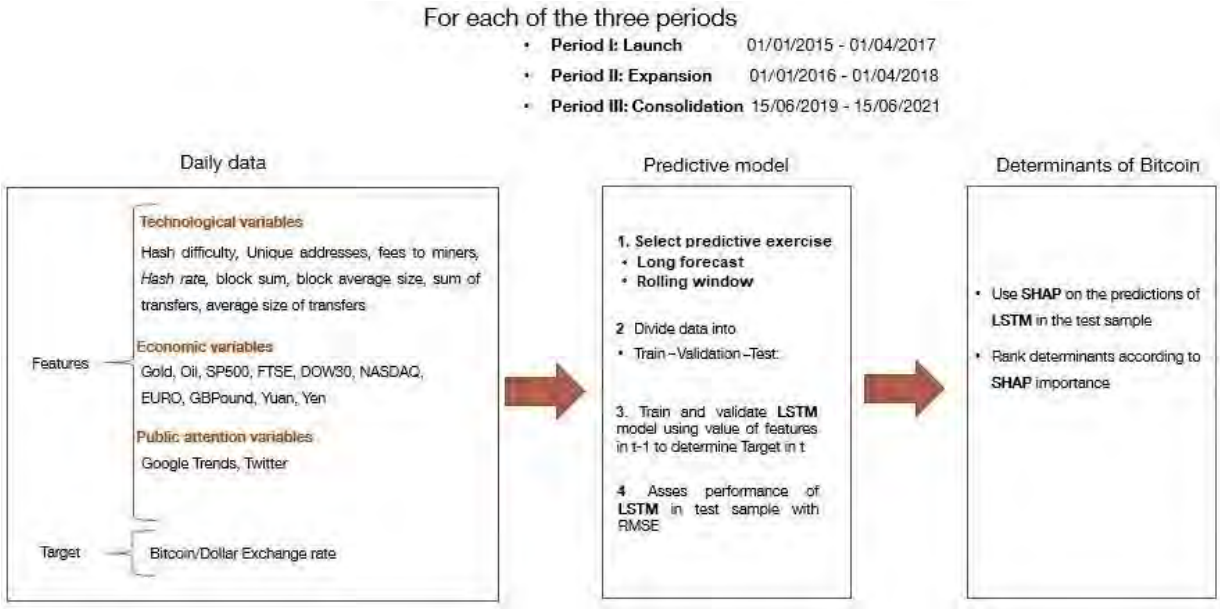
The empirical exercise proposed in this paper is summarized in **figure 2**. For each of the different time periods we chose to analyze, data on all the potentially explanatory factors -as mentioned in the previous section- was collected and further classified into three distinctive groups. The next step was to use the values of those features in *t-1* to train, validate and, finally, test our LSTM in order to see whether we could determine the price of bitcoin in *t*. We do so using two approaches, a long-range forecast approach and a rolling window approach, which will explain in detail in section 5.1. We then checked the resulting prediction error by calculating the root mean squared error metric. This allowed us to corroborate that the LSTM model was capable of satisfactorily replicating the price of bitcoin.

As explained before, our goal was not to build a model that could be used to flawlessly predict the future price of bitcoin. That is, by nature, an almost impossible task. What we pursued instead, was to develop a model that, without imposing or assuming any specific ex-ante relationship between the various features and the target variable, could roughly replicate the movements of the price of bitcoin. Finally, using SHAP, a ML interpretability technique, we checked the importance of the individual features in order to understand which ones might be the most important to determine the price of bitcoin in each period.

Below you will find all the details about our empirical study. In section 5.1 we explain the basics of LSTM model and why we used it. In section 5.2 we show the performance of LSTM based on its out-of-sample price prediction. In section 5.3 we introduce SHAP, the technique that we used to understand the outcome of the LSTM model. In section 5.4 we apply SHAP to uncover the determinants of the price of bitcoin in each of the analyzed periods. Finally, in the appendix we show how the performance of LSTM could be improved if we included the lag of the price of bitcoin as an additional variable. But this would complicate the analysis of the determinants of bitcoin, since interpretability techniques would attribute great importance to the lag in the price of bitcoin, and less to the rest of the

variables. That is why we prefer not to include the lag of the price of bitcoin as an explanatory variable, obtaining a slightly worse prediction but with an interpretability of the results that allows us to better understand the price determinants.

Figure 2: Empirical exercise proposed in this study



Source: Author's own calculations (2022).

5.1 LSTM model

The problem at hand, establishing the price of bitcoin as a function of the values adopted in the past by a set of explanatory variables, is a challenging task. Most of the data that were collected for this project are highly nonlinear and non-stationary. Therefore, the use of traditional time series will require imposing additional assumptions on the data. Since there's no generally-accepted theory that could be leveraged for this purpose, on account of the fact that our ultimate aim was to understand the extent to which certain factor movements were compatible with price movements, we chose to address this problem by relying on a Long Short Term Memory neural network (LSTM)[13]. The advantage of this approach is the flexibility that the model provides in that it does not impose any specific ex-ante relationship between the target variable (bitcoin price) and the features. Thus, it can be used in non-linear and non-stationary time series scenarios (Bala and Singh 2019, Abbasimehr et al 2020 and Wang et al 2021). The model can work well in the presence of multicollinear variables because it has a nonlinear structure (via activation functions) and also uses regularization techniques. Regularization allows to stabilize the coefficients

[13] The choice of this ML model was the result of a selection process to find the best performing one. In fact, priori to settling on the LSMT, we first tested different tree-based models, such as regression trees, random forest and XGBoost. In each of the periods, we found the prediction error of these models to be considerably higher than for the LSTM, with a root mean square error (RMSE) as high as twice the error borne by the LSTM. We tested further traditional deep learning models (without the characteristic time loop of recurrent neural networks or LSTMs), and although their performance proved to be better than with tree-based models, the prediction error was still at least 10% higher than compared with the LSTM in all the three periods considered.

assigned to the features, thus reducing the problem in the case that two features are multicollinear.

LSTM is a variation of feedforward neural networks which are capable of learning the time dimension of the data. We show a graphical representation of a simple fed forward neural network in **figure 3**, left. A feedforward neural network consists of an input layer, which corresponds to the input data, one or more hidden layers, and an output layer. The input layer has as many nodes as features in the data (four in the case of **figure 3**). The hidden layer is composed by nodes that represent linear combinations of weights of the nodes from the input layer (in this case, three nodes). These combinations are processed by activations functions (tangent, sigmoid, etc.) that result from the output of each hidden layer node. If there is more than one hidden layer (in **figure 3** there is only one hidden layer), the nodes of any hidden layer are based on the linear combinations of the output of the nodes of the previous hidden layer. Finally, there is an output layer based on a linear combination of nodes of the last hidden layer, to which we apply a nonlinear activation function.

The drawback of this architecture is that it does not take advantage of the information available from past values of the features or decisions of the net. Consequently, it might not be an ideal tool for performing time series analysis. However, this is partially solved by using Recurrent Neural Networks (RNN), which are distinguished from regular neural networks in that they can deal with sequential data and can be trained to hold the knowledge about the past. This is achieved by applying a feedback loop that is connected to their past decisions, thus ingesting their own outputs moment after moment as new input (**figure 3**, right). We explain in more detail how this feedback loop works for RNN in **figure 4** left.

Let's focus on the RNN cell at time t. The previous hidden state h_{t-1} and the input x_t are combined in a vector that will go through a tahn activation function, resulting in the output of the cell h_t, which represents the memory for the next time step. The tahn activation function ensures that the values of the vector are always between -1 and 1. This architecture allows the RNN to make use of sequential information, in a similar fashion to a Markov model. But long term information might not be used, since RNN are exposed to the vanishing gradient problem. Neural networks, including RNNs, update parameters based on an optimization algorithm called gradient descent, in which models learn via gradient values. If the gradient values are extremely small, the parameter updates become negligible, and as a result, the model stops learning or takes too long to learn.

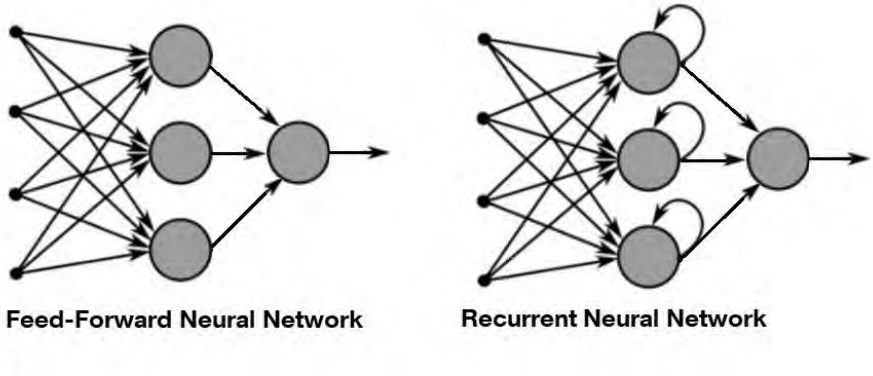
LSTM are a special version of RNN that can solve this problem. They were first proposed in 1997 by Hochreiter and Schmidhuber. The basic architecture of a LSTM cell is shown in **figure 4** right. The main difference from RNN is the way the model controls what information should be removed from memory, and what information should remain. This is done using information outside the normal flow of the network in a closed cell or memory, called cell state. This cell state can be seen as the horizontal line in the top of the cells, from h_{t-1} to h_t and to h_{t+1}. The LSTM decides whether to store or delete information from the cell state based on the importance it places on the information. There are three gates that can transform the cell state: forget gate, input gate, and output gate. These doors are neural networks that determine whether past information is eliminated (forget gate), whether or not new information is allowed (input gate), and what information is moved to the next state (output gate).

The forget gate combines the previous hidden state h_{t-1}, and the current input $x_t$through a sigmoid function, delivering a value between zero and one, where zero means non-

important and one means important. This value is then applied through pointwise multiplication to the cell state, in this way the cell state can forget unimportant information. The input gate combines as well the previous hidden state and the current input through a sigmoid function and a tahn function, and multiplies both values to regulate which information will be added through pointwise addition to the cell state. This updates the cell state according to the values that the input gate considers relevant. Finally, the output gate also combines the previous hidden state and the current input through a sigmoid function, which multiplies the tahn output of the modified cell state. This results in the new hidden state for the next time step.

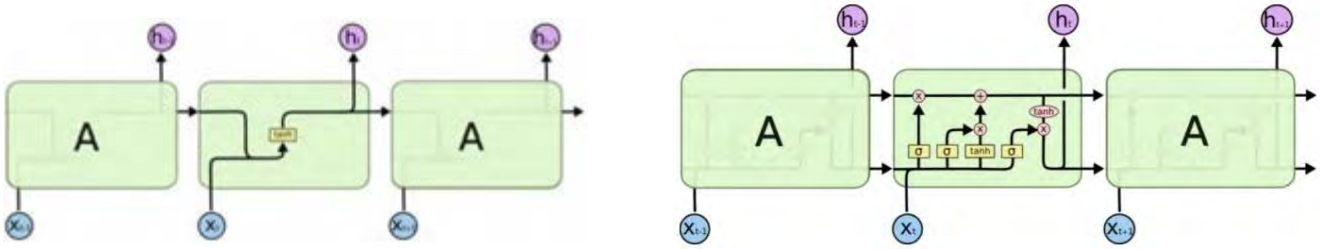
While the architecture and composition of the basic cell of a LSTM is more complicated than the one for RNN, the LSTM allows us to remember inputs for a long period of time, and it solves the vanishing gradient problem. This is the case since the gate structure of the LSTM ensures that the gradients do not converge to zero, keeping a relatively short training time.

Figure 3: Feed forward neural network and Recurrent neural networks (RNN) model



Source: Own authors (2022)

Figure 4: Recurrent neural networks (RNN) and LSTM model



Source: Olah, 2015.

As with any other ML model, we must avoid the problem of overfitting to produce a model that generalizes well to new and unknown data. If we train the LSTM model on the entire data sample, it will end up learning the optimal weights of each feature and fit the model's predictions perfectly for the in-sample data, but it will not be able to predict well out-of-sample. That is why it is common practice in ML to split data into three groups, train, validation, and test. We fit the model in the training sample. In our case, the target variable

was the price of bitcoin in *t*, while as features we used the 25 listed elements in section 3 with their values in *t-1*. In the validation set we evaluated the performance of the fitted model to select the hyper parameters and architecture of the model. In **Table 2** we summarize the range of values over which we select the hyper parameters in the validation process. Finally, we checked the accuracy of our model by making predictions in the test sample. In this way, we were able to assess the performance of the model against data sets which had not been directly involved in the training process.

We followed two approaches to perform the predictions. On the one hand, we perform what we refer to as long-range forecast approach, where we produce a single forecast for the six months of the testing sample. To further illustrate this, we present in **figure 5** what the long-term forecast approach looks like for the *launch period*, which commits from 01/01/2015 to 01/04/2017. With the long-range forecast, starting on 01/01/2015 we use roughly 18 months to train, three months to validate, and six months to predict. For the *expansion period* and *consolidation period*, we proceed accordingly. We use the three months prior to the test sample to validate, and the prior data is used for training.

On the other hand, we have also performed a rolling window approach, so that our results are not as sensitive to the specific start and end dates of the period of interest. In **figure 6**, we present how we have applied the rolling window approach for the *launch period*. We used 24 weeks to train, two weeks to validate, and four weeks to test. We then move the train, validation and test data forward one week and make a new prediction, until our predictions cover the entire test sample of the *launch period* (from 01/10/2016 to 01/04/2017, **table 1**). Therefore, in both approaches, around 70% of the data was used for training, 10% for validation, and 20% for testing.

Table 2. Parameters

Parameters	Values
Input lags	1-10 days of memory
Activation function	Linear
Learning rate	[0,1-0,001]
Number of hidden layers	3, 4, 5
Nodes in hidden layer	32, 64, 128, 256
Dropout value	[0,1-0,001]
Loss function	MSE

Source: Own authors (2022)

Figure 5: Train, validation and test, long forecast approach, launch period

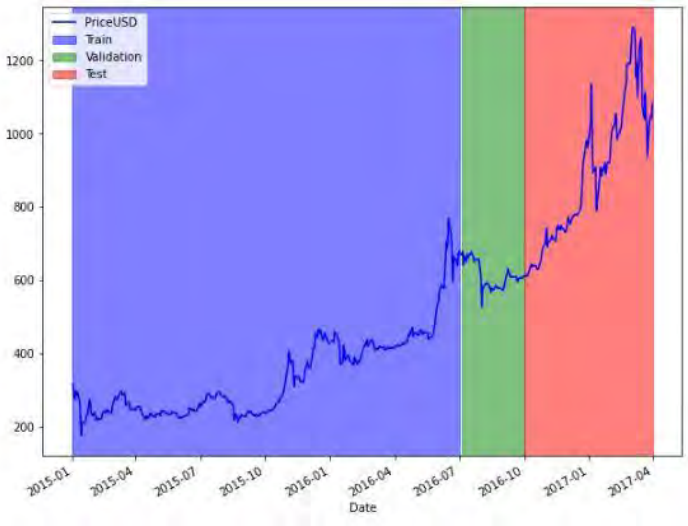
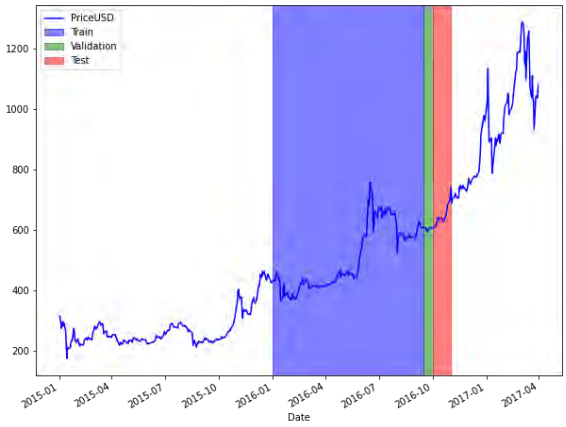


Figure 6: Train, validation and test, rolling window approach, launch period



5.2 LSTM results

We measure the effectiveness of the LSTM model through the metric RMSE (Root Mean Squared Error), which is based on the mean of the square root of the squared differences between the predicted values and the observed values. The formula for RMSE is as follows:

$$\text{RMSE} = \sqrt{\frac{\left(\sum_{t=1}^{T} \hat{y}_t - y_t\right)^2}{T}}$$

Where *T* is the number of days, $\hat{y}_t$ is the predicted price of bitcoin on day *t*, and y_t is the actual price of bitcoin on day *t.*

The results in terms of RMSE for the three periods are shown in **table 3**, for both the long-range forecast approach, and the rolling window approach. Our first finding is that the LSTM model performs reasonably well in all three periods considered. This is a particularly positive

outcome in order to cement the results of our exercise since we didn't use lagged values of the price of bitcoin as additional features. The results are very similar for both approaches, with an RMSE between 5% and 20% of the price. Another interesting observation is that the model offers its best outcome in the so–called *launch period* (5% of RMSE), followed by the *expansion period* (between 13 and 14% of RMSE). The error in the prediction is higher in the *consolidation period*, around 20%, particularly during March and April 2021.

The fact that the results of the rolling window approach are similar to those of the long-range approach suggests that the results are not strongly conditioned by the specific start and end dates of each period. In any case, as an additional robustness exercise, we tested the performance of the LSTM model by slightly altering the exact dates of each of the three periods. Specifically, we carried the long-range forecast exercise on different start and end dates, choosing dates between four weeks before and four weeks after the original start and end dates. For both the *launch* period and the *expansion* period, we observed that results do not change significantly, with the RMSE varying between 4% and 6% for the *launch* period, and between 12% and 18% for the *expansion* period. Along these lines, the RMSE of the LSTM model ranged between 20% and 26% during the *expansion* period when advancing two weeks and moving four weeks backwards both the start date and the end date. Interestingly, when pushing forward the start and end dates more than two weeks, the LSTM performed worse, with an RMSE that could reach 38%. It is worth mentioning that this, could change our conclusions around interpretability. This inferior prediction performance is explained by the fact the LSTM model cannot predict the big price increase that took place in 2021 when the training sample leaves out a critical amount of data points reflecting the price increase trend of bitcoin (late 2020). We recognize that this is a limitation of the study, and we leave for future research to delve into possible solutions to this problem.

To visualize the performance of the LSTM model, we show **in figures 7, 8 and 9** the prediction of the LSTM model with the long-range forecast approach (in orange) together with the real price of bitcoin (in blue) for each of the three periods. Regarding the *launch period* (**figure 7**), we noticed that the predicted price of LSTM followed very closely the actual price of bitcoin. The same holds true for the *expansion period,* **figure 8**. While the difference between the predicted and the actual price is higher in this latter case, given the steep price during the test sample with respect to the training and validations sample, we believe it is fair to say that the performance of LSTM remains reasonably good. **Figure 9** shows the resulting prediction values for the *consolidation period*. In this period, we did appreciate a considerable gap between predicted and actual price of bitcoin during March and April of 2021. On account of the above, we can first conclude that, based on the same features, the LSTM model's performance is generally good in terms of RMSE for the first two periods considered but it worsens considerably during the third one. These results could, however, be improved if the lag price value of bitcoin is included as an extra feature.

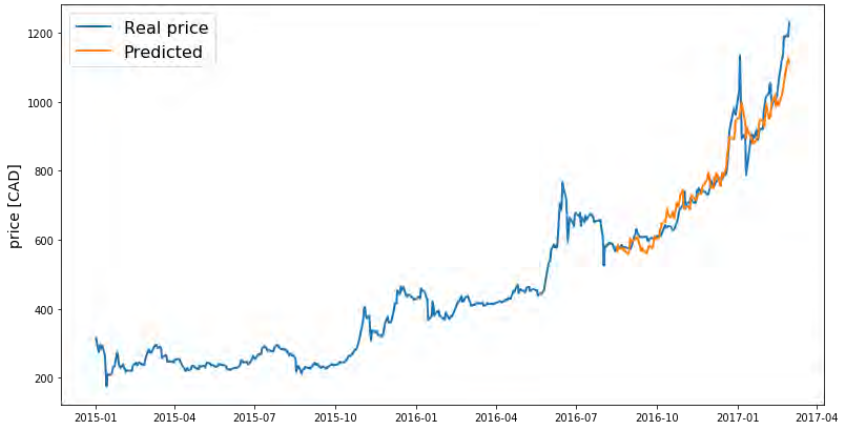
In the appendix we perform such an exercise which proves to render a considerably lower RMSE in the *consolidation period*. In other words, when included as a complementary feature, the lag of the price of bitcoin helps the LSTM model in the long-range forecast approach make a better prediction of future prices. However, this is achieved at the expense of interpretability since SHAP places a lot of weight on the lag of the price of bitcoin. Obviously, for the purposes of our exercise, these results are neither be very informative nor useful. Therefore, we refrained from taking this path for the rest of the exercise.

Table 3. Performance of the model in different periods

Period	RMSE Long forecast	RMSE Rolling window
Launch	5.7%	5.3%
Consolidation	13.2%	15.1%
Expansion	21.2%	19.1%

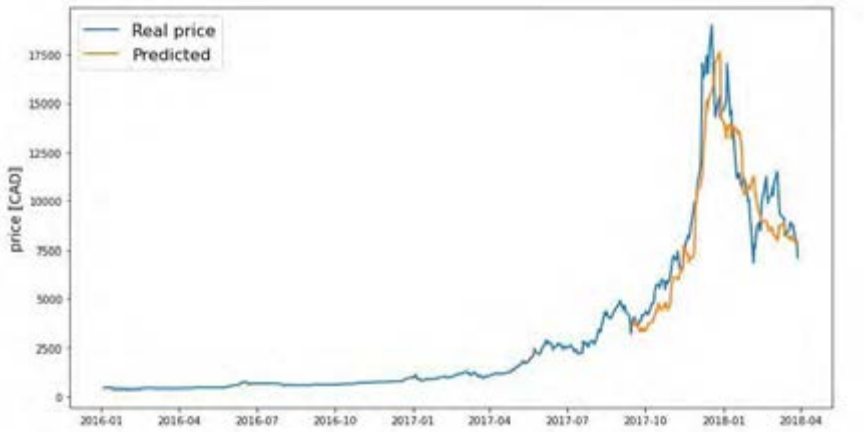
Source: Own authors (2022)

Figure 7: Predicted and actual bitcoin price. Launch period



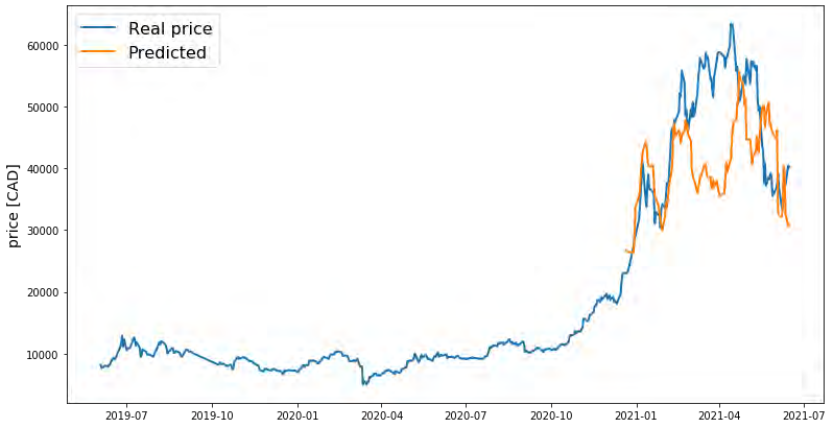
Source: Author's own calculations (2022).

Figure 8: Predicted and actual bitcoin price. Consolidation period



Source: Author's own calculations (2022).

Figure 9: Predicted and actual bitcoin price. Expansion period



Source: Author's own calculations (2022).

5.3 SHAP

The increase in the use of ML models has awaken an interest in how to explain their outcome. There are different techniques that can accomplish this (see Molnar 2020 for a detailed review and a comprehensive list of methods). Some of the most popular techniques are the so-called model agnostic or post hoc interpretability techniques, that can be applied to any model. They can either be used to explain which features matter for a single prediction (local interpretability), or to explain which features matter more in the whole dataset (global interpretability). In this paper we focus on latter aspect, since our goal is to establish which factors determine the price of bitcoin over long periods, rather than on a single day.

The two main global interpretability techniques are SHAP (Lundberg and Lee, 2017, Lundberg et al 2020)[14] and Permutation Feature Importance (introduced by Breiman 2001 for Random forest, and model agonist version provided by Fisher, et *al*. 2018). Both SHAP and Permutation Feature importance are global interpretability techniques that can yield a comprehensive ranking of the importance of features to the ML model in question. We have decided to use SHAP instead of Permutation Feature Importance as the main analysis for the following reasons.

First, SHAP is growingly gaining traction in the context of Deep Learning models (Albanesi and Vamossy 2019). In addition, it seems to have an advantage when features are correlated (Molnar 2020, Alonso and Carbó 2022). While it's true that highly correlated features will imply unrealistic permutations in both the above methods, these will be particularly significant in the case of Permutation Feature Importance. And third, while Permutation Feature Importance informs about which features matter more for a prediction's error, SHAP delivers directly the importance of a given feature in terms of its impact on the prediction itself. This allows us to compute the percentage impact of each feature on the price of bitcoin. In any case, none of these techniques is absent criticism (see, e.g., Rudin, 2019).

[14] The Shapley values can be used as local interpretability technique.

How does SHAP determine the importance of each feature? SHAP is a technique that measures the contribution of a variable to the predicted outcome, on a particular day, compared to the average prediction. These contributions are called the Shapley values. Once we have the Shapley values, for each variable and for each day, these can be added to obtain the final importance of the variable. Therefore, SHAP can be used as a local and global interpretability technique.

The approach to compute Shapley values can be explained from a game theory perspective. The game would be to reproduce the result of the model (in our case, the price of bitcoin). The players would be all possible coalitions of variables. Finally, the reward would be the contribution of each coalition towards the final outcome of the model. For illustrative purposes, here is an example. Suppose we decide to use the following variables: "Gold," "SP500," and "Hash rate". Next, consider that we want to know the extent to which the "SP500" is important in order to establish the price of bitcoin on a given day t. Based on the above, these are the four possible coalitions of variables without "SP500":

- Without variables
- Gold
- Hash rate
- Gold and Hash rate

For the four coalitions, we calculate the price in *t* with and without the "SP500". The Shapley value of the "SP500" in relation to the price of bitcoin on day *t* is the weighted average of those marginal contributions. To obtain the global importance of the "SP500" for the price of bitcoin in the test sample, we repeat the process for all days within the sample and compute the average of the Shapley values. In the appendix we provide further details as to how to find the solution analytically.

5.4 Determinants of the price of bitcoin according to SHAP

Since we considered three different periods, we ended up with three different sets of predictions. In this section, we take stock of SHAP (see section 4.3) to analyze the extent to which each of the 25 features were important for those predictions. We are going to focus on the results for the long-rage forecast approach, since the results for the rolling window approach are similar. The results can be found in **figures 10, 11 and 12**.

Period #1 – Launch.

Figure 10 depicts the top 15 features which, according to SHAP, influenced the price of bitcoin in the *launch period*. A hallmark of this initial phase was how technology-related variables inherent to bitcoin showed clear signs of playing a critical role. In fact, Hash difficulty emerged as the most important variable, with an effect on the price of bitcoin of 13% over the mean prediction in the test sample. Other proprietary technological factors like, e.g., unique addresses, number of transactions, and fees to miners ranked among the top seven features as well. All of them helped explain price behavior with an effect that is calculated to fall above2%. Regarding those variables which reflect public attention, Google Trends appeared to be the third most relevant driver, with an effect of around 4% over the mean prediction. As we will see later on, both Google Trends and Twitter, gain greater importance in subsequent periods. Finally, among the variables which reflect the status of

the economy results were heterogeneous. In particular, the Yuan appeared as the second most relevant feature, but gold and the SP500 were nowhere near the top 10 spots.

Period #2 - Expansion.

In accordance with SHAP, the top 15 features that explain the price of bitcoin during this phase are shown in **figure 11**. In 2018, technology-related variables inherent to bitcoin continued to be among the most important determinants, although they were no longer as important as was the case in the *launch period*. Yet, Hash difficulty still was the most critical price driver, with an even more pronounced effect over the mean prediction: around 20%. In contrast, other variables capturing the technology dimension like, e.g., unique addresses, number of transactions, mean block size, etc. were much less relevant during the *expansion period* in comparison to the role they played during the *launch period*.

On the other hand, in the *expansion period* the variables related to the level of public attention/interest became much more critical. In particular, Google Trends ranked as the second most important explanatory factor, with an effect of 13% over the mean prediction. Likewise, Tweets appeared among the top 5 ones. Regarding economic variables, again heterogeneity was the rule. Opposite to what was the case in the *launch period*, the variable Yuan lost all of its importance. Moreover, gold remained unimportant, and the Dow, Nasdaq and the SP500 despite showing up in the top10 spots, exhibit a rather low impact (3.5%, 2.5% and 2.4% respectively) in comparison with the top variables.

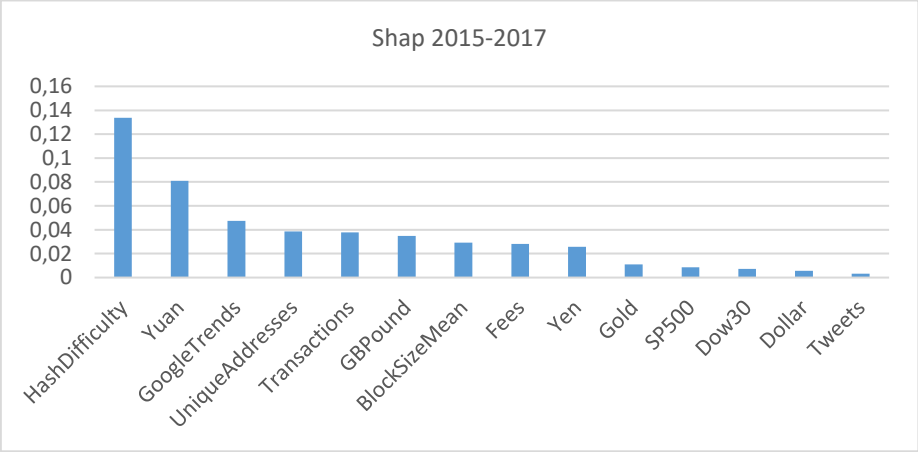
Period #3 - Consolidation.

As above, **figure 12** displays the top 15 features that help explain the price of bitcoin as derived from SHAP in the *consolidation period*. It is worth highlighting how technology-related variables associated with bitcoin lost visibly importance: while fees to miners were now the second most important aspect, with an effect of 15% over the mean prediction, other variables falling into this category did not appear among the top 15. It is especially remarkable as well how Hash difficulty, the most critical variable in both the *launch* and the *expansion period* now suddenly became completely irrelevant. On the other hand, the variables related to the level of public attention/interest definitely took over: Google Trends replaced previous candidates as the main explanatory factor with great impact: i.e. over 23% over the mean prediction. Again, the results were very different across the variables which reflect the state the economy is in. Overall they seem to have an intermediate importance, never at the top, but also not irrelevant at all. Among those, the most important ones for this period seemed to be GBPound exchange rate, the Nasdaq and the Yuan, all of them with an impact of 5% over the mean production.

Finally, in order to obtain a general overview of how each category evolved to become more or less relevant, we aggregated all the features within each distinctive group: i.e. technological variables, economic variables, and public attention variables. For each period, we combined the SHAP values of all features, and we computed which percentage belonged to each category. These results for the long-range forecast approach are summarized in **table 4.** Thus, in the first two periods, technological variables were clearly the most important ones. Interestingly in the last period, they lost relevance quite visibly: i.e. from almost 50% of all the impact in the first period, to 21% in 2021. Variables related to sentiment gained importance as the years went by. Their overall effect started at around 11% in the *launch period*, and climbed to 34% in the *consolidation period*. Economic variables did not present a clear trend. The results for the rolling window approach are

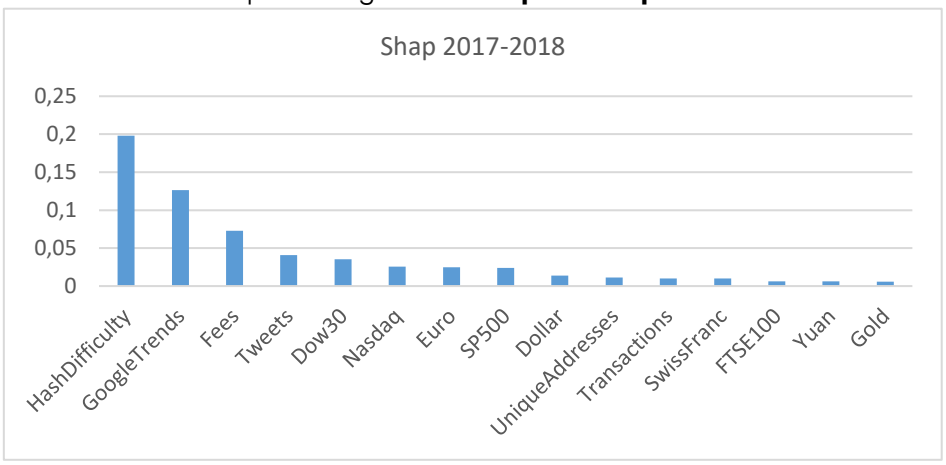
summarized in **table 5**, and although the percentages of importance are not exactly the same, we can find the same patterns: The technological variables lost relevance in the last period, while interest variables gained importance across periods.

Figure 10: Difference with respect to the mean prediction of including or not the variable, in percentage terms. **Launch period**



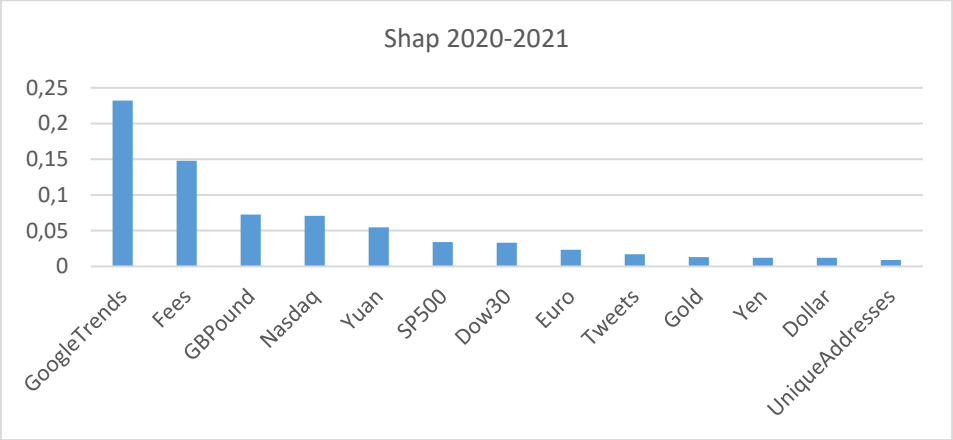
Source: Own authors (2022)

Figure 11: Difference with respect to the mean prediction of including or not the variable, in percentage terms. **Expansion period**



Source: Own authors (2022)

Figure 12: Difference with respect to the mean prediction of including or not the variable, in percentage terms. **Consolidation period**



Source: Own authors (2022)

Table 4.

Period	Technology Long forecast	Macro Long forecast	Interest Long forecast
Launch	46 %	44 %	10%
Consolidation	38 %	34%	28%
Expansion	21 %	45 %	34%

Source: Own authors (2022)

Table 5.

Period	Technology Rolling window	Macro Rolling window	Interest Rolling window
Launch	36 %	53 %	11%
Consolidation	36%	43 %	21%
Expansion	27 %	48 %	25%

Source: Own authors (2022)

6. Conclusions

In this article we have explored which elements are most likely to explain the behavior of bitcoin prices at a given point in time: i.e. in between (i) 2015 and 2017, (ii) 2017 and 2018, and (iii) 2020 and 2021. We have also analyzed the extent to which these features may be longstanding or change over the years. The aim of this exercise is to help shed some light on the degree of maturity achieved by such digital asset in comparison to other financial instruments and alternative crypto-related value propositions such as those which are backed by underlying assets (i.e. stablecoins). This way we expect to potentially contribute to the current broader public policy debate on future actions to be adopted by relevant macro and microprudential authorities for the safeguarding of financial stability.

As a point of departure, we have first taken stock of the existing literature on both the underlying nature of bitcoin (e.g. safe haven, hedge, diversifier or speculative asset) and its price/return determinants. As a result, we have identified three main set of potential drivers -(i) macroeconomic factors, technology-related ones as well as those that reflect the degree of attention that said crypto-asset may be drawing- that we have leveraged on for our purposes.

Against this light, we found that technological variables emerged as the more relevant ones for the determination of bitcoin prices during the first two periods of our sample. However, they lost all its significance as we entered the last observation period. More precisely, variables such as hash difficulty, block size, number of transactions or unique addresses rendered virtually irrelevant to elucidate the evolution of bitcoin prices as we neared 2021. Conversely, variables pointing at the degree of public attention enjoyed by bitcoin -like Google Trends- grew progressively in importance as we came closer to the present day. In fact, in stages defined by high price volatility (2018 and 2021), the interest of the public takes on a very notable role.

It is also worth noting that variables highlighting the role of macroeconomic and financial development never took up a leading place in the establishment of bitcoin prices. Throughout the entire observation period, they seem to exercise a limited, yet constant influence while also showing a great degree of heterogeneity among each of its individual components. This appears to be in contradiction with the findings of part of the literature which oftentimes underscores the critical positive impact of the SP500 and gold on the determination of price.

On account of the above, our research leads us to conclude that the formation of the price of bitcoin is still a highly complex phenomenon whose underlying causes are difficult to anticipate with an acceptable degree of uncertainty. While most of the determinants highlighted in the literature seem to clearly play a role in the evolution of bitcoin prices over time, we prove that their influence does change substantially at short notice. Moreover, possibly due to the immaturity of crypto-assets markets, oftentimes new explanatory factors emerge unexpectedly which, furthermore, may remain undetected and opaque to both investors and authorities for long periods of time. This may be one of the reason why our predictive model performs notably worse in 2021.

For the above reasons, compared to other well-known and well-established asset classes, bitcoin - and, by extension, its namesakes - seems to continue to exhibit a difficult-to-predict behavior, thus making it a high-risk investment in the current landscape. It is, therefore, advisable for financial authorities to be fully aware of this fact upon deciding, at

least, on the prudential treatment to be assigned to the potential exposures of banks to unbacked crypto-assets, in particular as regards market and liquidity risk, as well as in relation to the adoption of other relevant conduct-related measures in defense of investors and consumers at large.

This may, for example, call for deeper reflections by authorities on the implied model risk and further vindicate the amount of public warnings on crypto-assets that both national competent authorities and regional regulators have been issuing over time. In addition, such circumstance is supportive of more recent measures aimed at supervising the way these offerings are advertised in order to better cope with the existing asymmetries in end-users' knowledge and understanding of the actual risks these digital assets entail.

More broadly, our findings could also be of interest to macroprudential authorities in their assessment of the materiality of the potential risks crypto-assets place on global financial stability and the need and timeliness of the deployment of effective regulatory and supervisory actions. Against this light, financial authorities may further want to consider maintaining conservative positions regarding their regulation so as to avoid the transmission of potentially systemic risks to the financial system as a whole.

7. References

Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. Computers & industrial engineering, 143, 106435.

Albanesi, S., & Vamossy, D. F. (2019). Predicting consumer default: A deep learning approach (No. w26165). National Bureau of Economic Research.

Alessandretti, L., ElBahrawy, A., Aiello, L. M., & Baronchelli, A. (2018). Anticipating cryptocurrency prices using machine learning. Complexity, 2018.

Alonso, A., & Carbó, J. M. (2022). Accuracy and stability of explanations of machine learning models for credit default prediction. Forthcoming.

Aslanidis, N., Bariviera, A. F., & López, Ó. G. (2022). The link between cryptocurrencies and Google Trends attention. Finance Research Letters, 102654.

Bala, R., & Singh, R. P. (2019, July). Financial and non-stationary time series forecasting using lstm recurrent neural network for short and long horizon. In 2019 10th international conference on computing, communication and networking technologies (ICCCNT) (pp. 1-7). IEEE.

Bariviera, A. F., & Merediz-Solà, I. (2021). Where do we stand in cryptocurrencies economic research? A survey based on hybrid analysis. Journal of Economic Surveys, 35(2), 377-407.

Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? Journal of International Financial Markets, Institutions and Money, 54, 177-189.

Bouri, E., Molnár, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? Finance Research Letters, 20, 192-198.

Bouoiyour, J., & Selmi, R. (2015). What does Bitcoin look like? Annals of Economics & Finance, 16(2).

Bouoiyour, J., & Selmi, R. (2017). The Bitcoin price formation: Beyond the fundamental sources. arXiv preprint arXiv:1707.01284.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Carbó, J. M., & García, E. D. (2021). El interés por la innovación financiera en España: un análisis con google trends (No. 2112). Banco de España.

Cermak, V. (2017). Can bitcoin become a viable alternative to fiat currencies? An empirical analysis of bitcoin's volatility based on a GARCH model. An Empirical Analysis of Bitcoin's Volatility Based on a GARCH Model (May 2, 2017).

Ciaian, P., Rajcaniova, M., & Kancs, D. A. (2016). The economics of BitCoin price formation. Applied Economics, 48(19), 1799-1815.

Ciaian, P., Kancs, D. A., & Rajcaniova, M. (2018). The price of BitCoin: GARCH evidence from high frequency data. arXiv preprint arXiv:1812.09452.

Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. International Journal of Forecasting, 37(1), 28-43.

Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. Journal of Machine Learning Research, 20(177), 1-81.

Hayes, A. S. (2017). Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. Telematics and Informatics, 34(7), 1308-1321.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

Kaminski, J. (2014). Nowcasting the bitcoin market with twitter signals. arXiv preprint arXiv:1406.7577.

Kapar, B., & Olmo, J. (2021). Analysis of Bitcoin prices using market and sentiment variables. The World Economy, 44(1), 45-63.

Kim, Y. B., Kim, J. G., Kim, W., Im, J. H., Kim, T. H., Kang, S. J., & Kim, C. H. (2016). Predicting fluctuations in cryptocurrency transactions based on user comments and replies. PloS one, 11(8), e0161197.

Kjærland, F., Khazal, A., Krogstad, E. A., Nordstrøm, F. B., & Oust, A. (2018). An analysis of bitcoin's price dynamics. Journal of Risk and Financial Management, 11(4), 63.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. PloS one, 10(4), e0123923.

Lundberg, S.M. and S.I. Lee (2017). "A unified approach to interpreting model predictions" Advances in neural information processing systems, pp. 4765-4774.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., and Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1), 56-67.

Lyócsa, Š., Molnár, P., Plíhal, T., & Širaňová, M. (2020). Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin. Journal of Economic Dynamics and Control, 119, 103980.

Molnar, C. (2020). Interpretable machine learning. Lulu. com.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Decentralized Business Review, 21260.

Olah, C. (2015). Understanding LSTM Networks. Retrieved from https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Panagiotidis, T., Stengos, T., & Vravosinos, O. (2018). On the determinants of bitcoin returns: A LASSO approach. Finance Research Letters, 27, 235-240.

Pyo, S., & Lee, J. (2020). Do FOMC and macroeconomic announcements affect Bitcoin prices? Finance Research Letters, 37, 101386.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.

Sovbetov, Y. (2018). Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litcoin, and monero. Journal of Economics and Financial Analysis, 2(2), 1-27.

Urquhart, A., & Zhang, H. (2019). Is Bitcoin a hedge or safe haven for currencies? An intraday analysis. International Review of Financial Analysis, 63, 49-57.

Urquhart, A. (2018). What causes the attention of Bitcoin?. Economics Letters, 166, 40-44.

Wang, T., Zhao, S., Zhu, G., & Zheng, H. (2021). A machine learning-based early warning system for systemic banking crises. Applied Economics, 53(26), 2974-2992.

Zhu, Y., Dickinson, D., & Li, J. (2017). Analysis on the influence factors of Bitcoin's price based on VEC model. Financial Innovation, 3(1), 1-13.

Zhu, P., Zhang, X., Wu, Y., Zheng, H., & Zhang, Y. (2021). Investor attention and cryptocurrency: Evidence from the Bitcoin market. Plos one, 16(2), e0246331.

8. Appendix

8.1 Definitions

As taken from Coinmetrics:

Difficulty finding the hash: The mean difficulty at a given day of finding a hash that meets the protocol-designated requirement (i.e., the difficulty of finding a new block).

Unique addresses: The sum count of unique addresses that were active in the network (either as a recipient or originator of a ledger change) at a given day.

Commissions to miners (fees): The sum USD value of all fees paid by user that makes transactions at a given day. Fees do not include new issuance.

Hash rate: The mean rate at which miners are solving hashes at a given rate. Hash rate is the speed at which computations are being completed across all miners in the network.

Sum of blocks: The sum count of blocks created that interval that were included in the main (base) chain at a given day.

Average block size: The mean size (in bytes) of all blocks created at a given day.

Sum of transfers: The sum count of transfers at a given day. Transfers represent movements of native units from one ledger entity to another distinct ledger entity. Only transfers that are the result of a transaction and that have a positive (non-zero) value are counted.

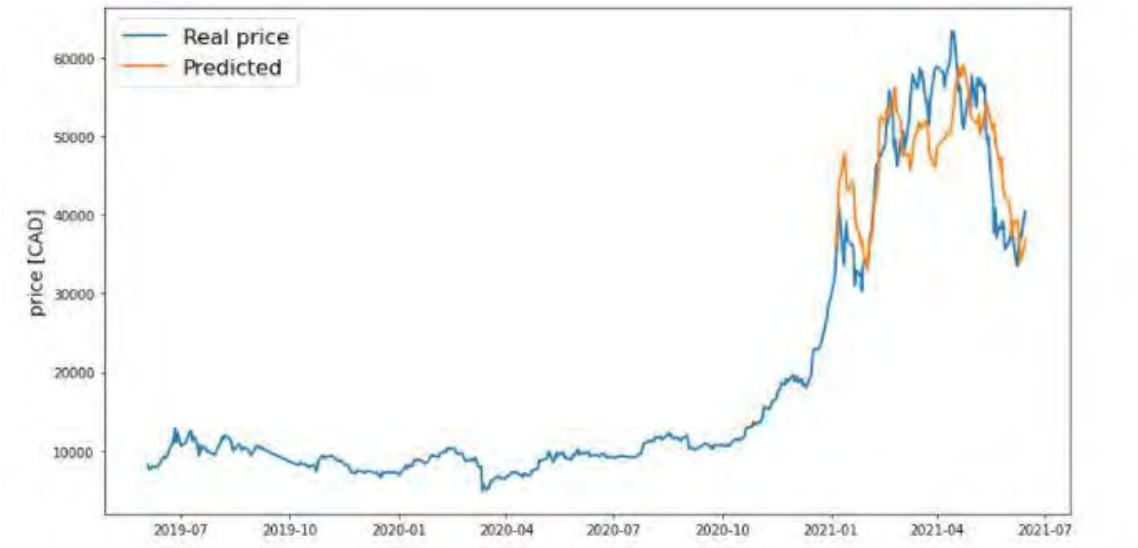
Medium transfer size: The sum value of native units transferred divided by the count of transfers (i.e., the mean size of a transfer) between distinct addresses that interval.

8.2 LSMT with lagged price

Our LSTM performed worse in the *consolidation period*, 2021, than in the other two periods. In this exercise we include as a feature the price of bitcoin in *t-1*, along with the original 25 features. With this new feature, the predictive performance of the LSTM improves considerably in every period, but particularly in the *consolidation period*. **Figure 13** shows the result from the LSTM model once we include the lag of the price of bitcoin as a feature. It can be seen that the gap between actual price and predicted price in March and April of 2021 is much smaller now than in the original exercise (see **Figure 9**).

The RMSE is of 11%, in contrast with 21% in the main exercise. Our model can predict better once we include the lag of the price. The reason for not including this variable in the main exercise is that the bitcoin price lag will bias our interpretation of the SHAP values, since it will appear as the main determinant. But this will not help us in our analysis, which is to understand what are the potential determinants of the price of bitcoin.

Figure 13: Predicted and actual bitcoin price using lag price as feature. Expansion period



Source: Own authors (2022)

BANCO DE ESPAÑA
Eurosistema

DETERMINANTS OF THE PRICE OF BITCOIN: AN ANALYSIS WITH MACHINE LEARNING AND INTERPRETABILITY TECHNIQUES

Granular data: new horizons and challenges for central banks
BoC-IFC Satellite Seminar

July 15th 2023, Bank of Canada, Ottawa

José Manuel Carbó, Sergio Gorjón

DGA FINANCIAL INNOVATION AND MARKET INFRASTRUCTURE



Why do we focus on bitcoin?

- Due to its weight and driving effect, the analysis of bitcoin could serve to illustrate possible common factors that influence or determine the general behavior of this market.

Why finding the determinants of bitcoin is relevant?

- Consolidation of bitcoin may bring risks of systemic nature.

- In particular, there is concern about its possible impact on financial stability, monetary policy, the integrity of the financial system and/or the soundness of the payment circuits.

- This concern increases for the most volatile crypto assets (bitcoin) and as the ecosystem becomes more complex and its ties with the financial system are strengthened:

  - *Investors (retail and institutional) increased exposures to crypto assets.*
  - *(Global) banks are considering offering/extending associated services.*
  - *More authorities/countries discuss the possibility of "normalizing" its use*
  - *Sophisticated product markets (ETFs)*

Why finding the determinants of bitcoin is relevant?

Improving the understanding of the factors that influence price formation, as well as its possible (in)stability over time, could help:

i. anticipate real risks to the system and,

ii. support the design of a regulatory framework that helps contain them, effectively, without compromising the opportunities.



Research question:

What variables determine the price at which bitcoin trades on exchanges?

The goal **is not** about building an investment tool or the best predictive tool

Existing literature on bitcoin price prediction

- There is an ongoing discussion on what are the factors that affect the price of Bitcoin.

EUR/USD, Yuan/USD, SP500, Dow Jones, FTSE



Determinant:

- Zhu et al 2017
- Chen et al, 2021
- Yhlas Sovbetov, 2018
- Bouoiyour and Selmi, 2017
- Cermak, 2016
- Kapar & Olmo, 2020
- Dennis van Wijk, 2013

No determinant:

- Ciain et al, 2016
- Ciaian and Rajcaniova, 2019
- Baur et al, 2018
- Pyo and lee, 2019

Hash rate



Determinant :

- Bouoiyour and Selmi, 2017
- Ciaian et al, 2016;
- Kristoufek, 2013
- Hayes, 2018

No determinant**:**

- Kjærland, 2018
- Fantazzini, 2020

Existing literature on bitcoin price prediction

- Consensus is greater on the role of gold and sentiment on the internet.

Gold



Determinant:

- Bouoiyour and Selmi, 2017
- Panagiotidis, 2018
- Kapar & Olmo, 2020

Sentiment

Determinant:

- Yhlas Sovbetov, 2018
- Panagiotidis, 2018 (Google Trends)
- Chen et al, 2021
- Ciaian, 2016
- Zhu et al, 2021
- Kim et al, 2016





- Lyocsa, Molnar, 2020
- Kaminski, 2014 (Twitter)
- Kristoufek, 2013
- Kristoufek 2015

Contribution

- Most of these articles focus on prediction:
  - Heterogeneity of results, depending on the method and the period of time used.
  - They use past prices and other strongly endogenous variables (capitalization in USD).

- We use a **flexible Machine Learning (ML) model to determine** the price of bitcoin, and employ **ML interpretability techniques** to analyze price drivers.

- **LSTM:**
  - We use a variant of Deep Learning called Long Short Term Memory (**LSTM**).
  - Flexible model, ideal because we do not know the relationship between the factors and the price of bitcoin.
  - It allows to determine its price without using highly endogenous variables or past values of the price.

- **Interpretability of** LSTM**:**
  - To interpret the determinants of the LSTM, we use **SHAP**.
  - It is one of the most robust interpretability techniques.

Possible determinants

We consider three possible groups of explanatory factors linked to: (i) the technological component of bitcoin, (ii) the evolution of the economy, and (iii) the interest it arises in the public.

Technological variables

- Difficulty finding the hash
- Unique addresses
- Commissions to miners *(fees)*
- *Hash rate*
- Block sum
- Average block size
- Sum of transfers
- Average transfer size

Economic variables

- Gold price
- Oil price
- SP500
- FTSE
- DOW30
- NASDAQ
- Euro, GBPound, Yuan, Yen

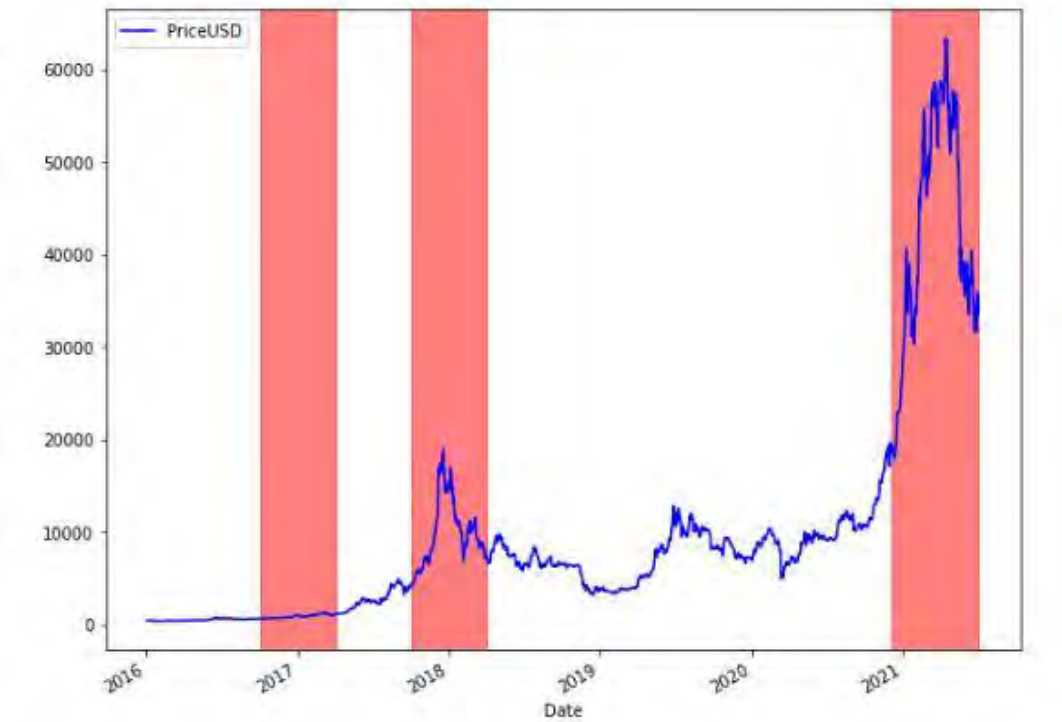
Interest

- Google Trends
- Twitter

- Sample time interval: January 2015 to July 2021
- **Daily data** from Coinmetrics, Yahoo Finance, Google and Bitinfocharts.
- **We plan to go more granular: 10 minutes by 10 minutes (block by block)**

Plots

Segmentation in three phases or differential stages

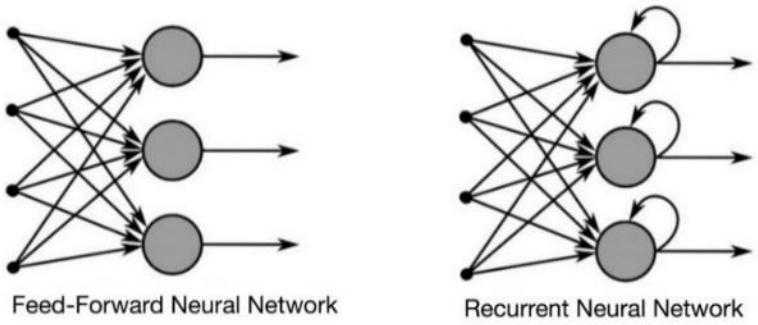


Source: Own (2021). Data from Coinmetrics

- **Stage I: Launch** 01/10/2016 to 01/04/2017
- **Stage II: Expansion** 01/10/2017 to 01/04/2018
- **Stage III: Consolidation** 01/01/2021 to 01/07/2021

LSTM model

- Neural networks are the most popular models for predicting, but they are not particularly good for time series.



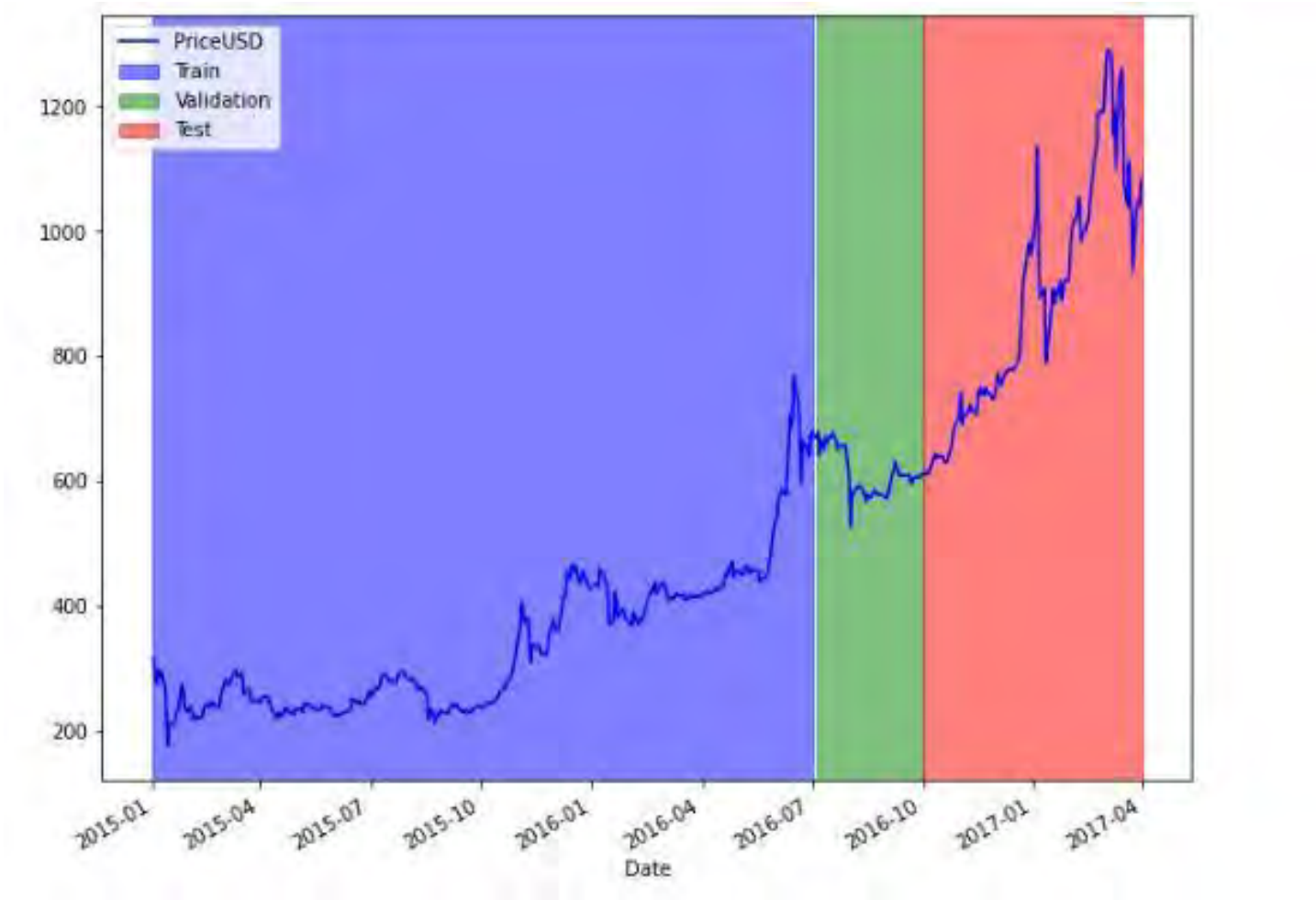
- Recurrent networks (RNN) and Long short-term memory units (LSTM) are neural networks that can have a temporal dimension.

RNN and LSTM

Recurrent networks are distinguished by a feedback loop connected to their past decisions, ingesting their own outputs moment after moment as input.

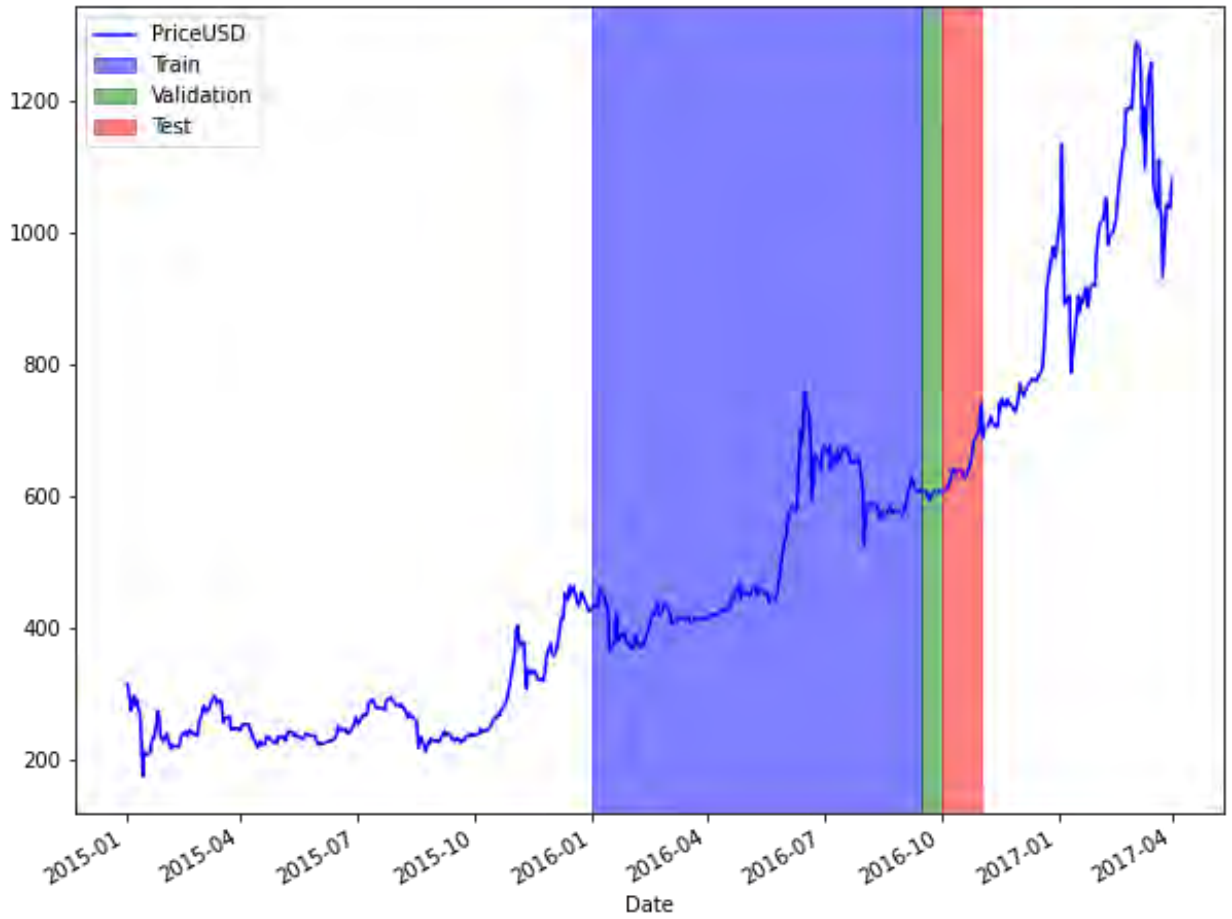
Details Cross validation

Train-Val-Test Long forecast



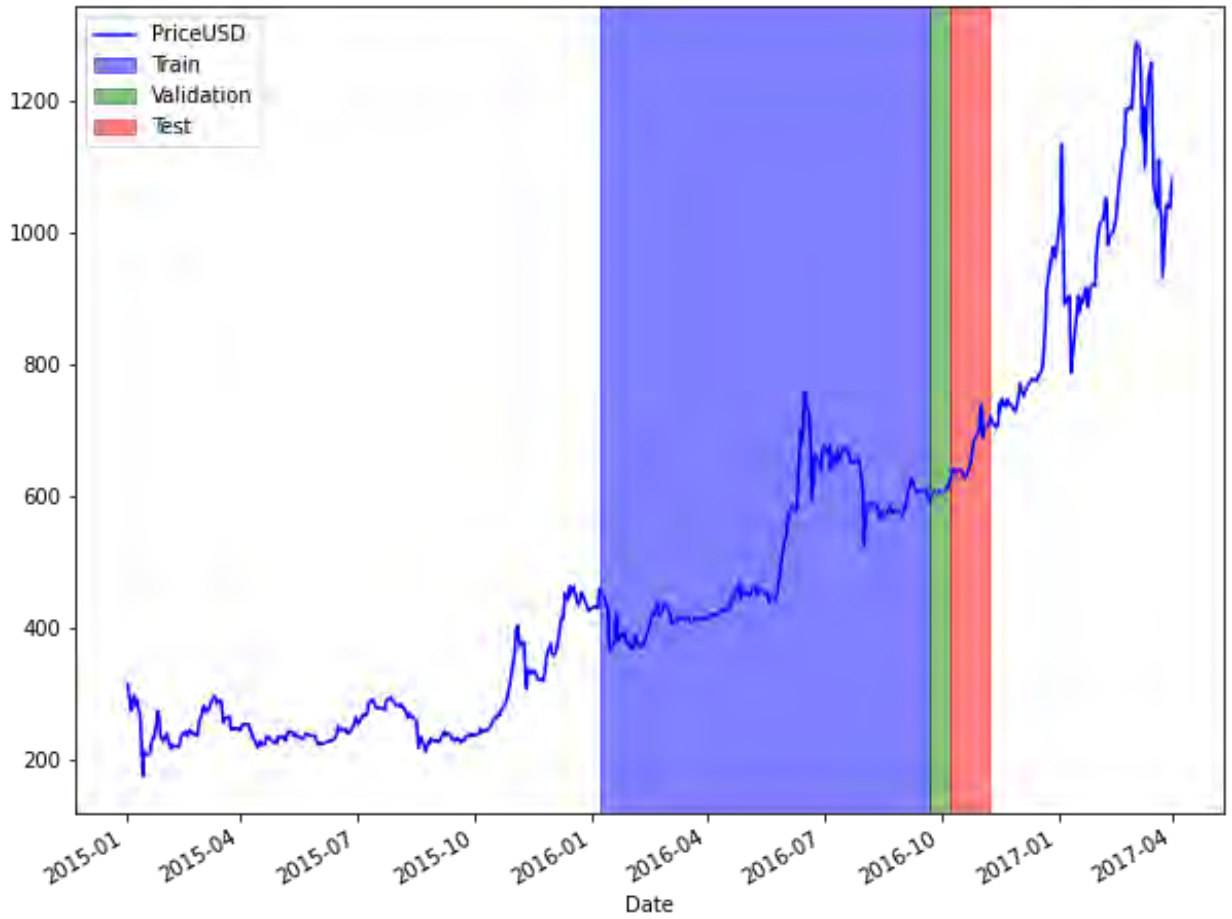
Source: Own calculations (2021)

Train-Val-Test Rolling window (i)



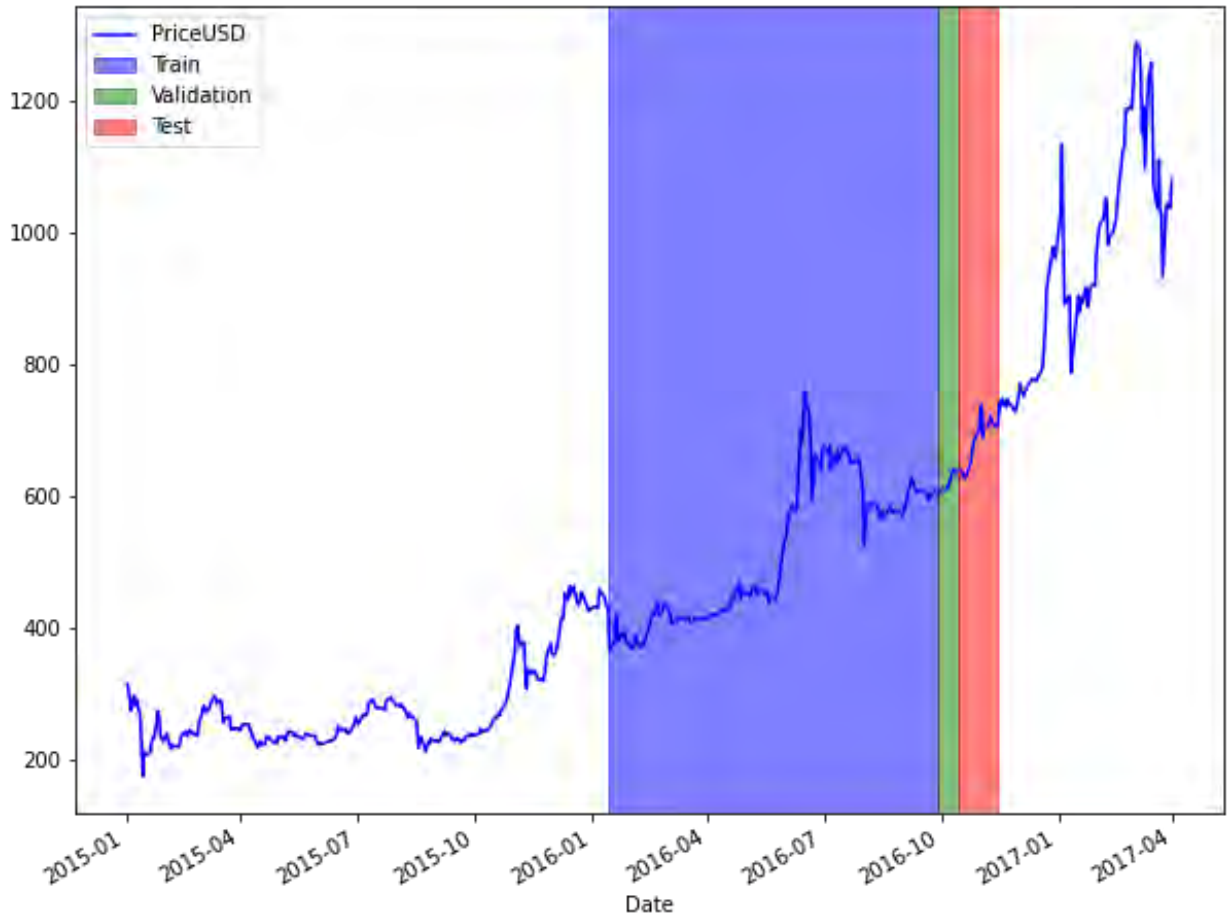
Source: Own calculations (2021)

Train-Val-Test Rolling window (i)



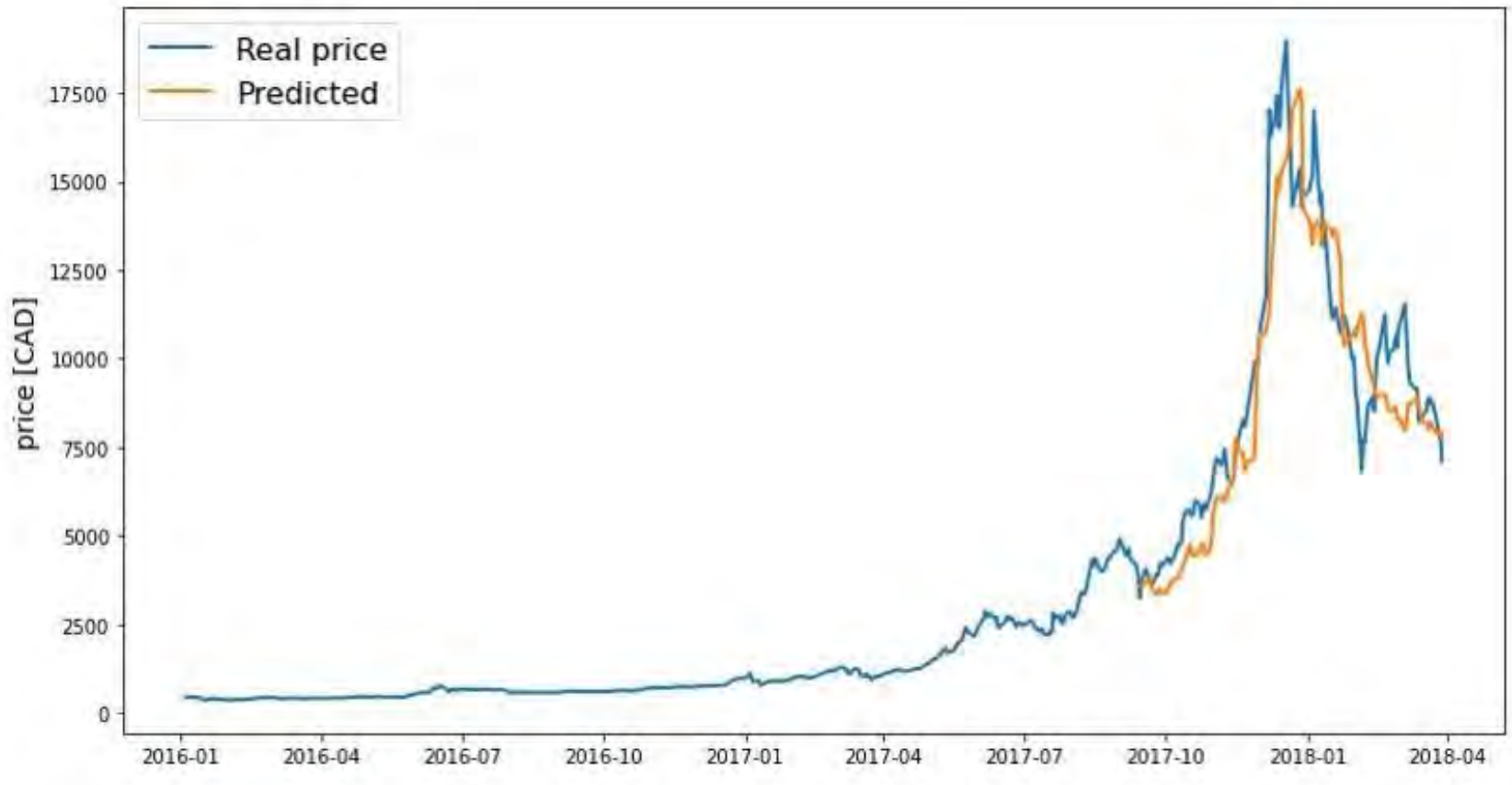
Source: Own calculations (2021)

Train-Val-Test Rolling window (i)



Source: Own calculations (2021)

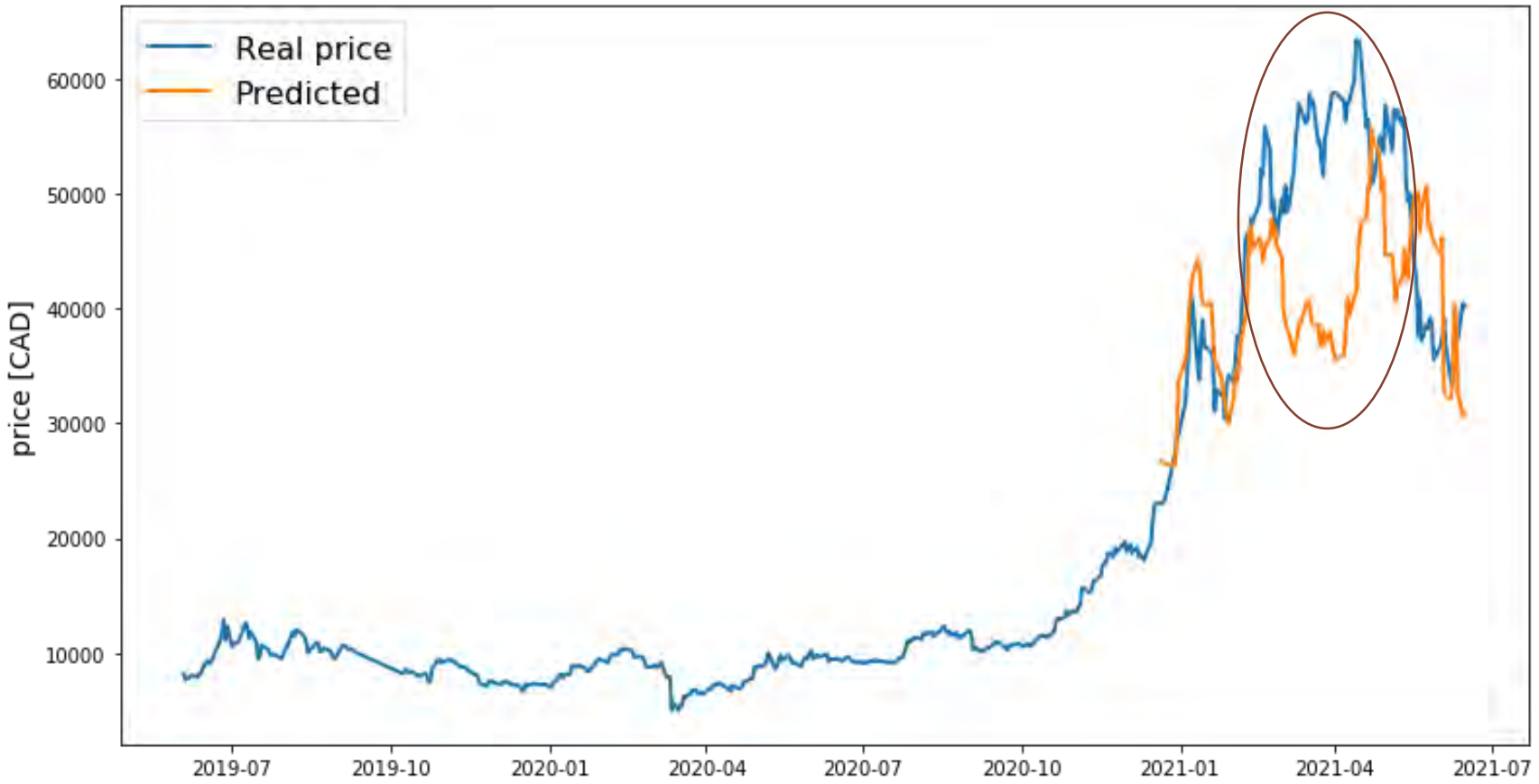
Period I: 2017-2018



Source: Own calculations (2021)

RMSE

Period III: 2020-2021



Source: Own calculations (2021)

Results 2016 RMSE

- SHAP is a technique that measures the contribution of a variable to the predicted outcome, on a given day compared to the average prediction.
  - These contributions are called the Shapley values

An example

- Suppose we have as variables"*Gold*," "*SP500*," and "*Hash rate*"**, and that we want to know the importance of the "SP500" in the price of Bitcoin on day t.**
- These are the four possible coalitions of variables without "SP500":
  - No variables
  - *Gold*
  - *Hash rate*
  - *Gold* and *Hash rate*
- For the four coalitions, we calculate the price in t with and without "SP500".
- The Shapley value of "SP500" for the price of bitcoin on day t is the weighted average of those marginal contributions.
- To obtain the global importance of "SP500" in the bitcoin price in the test sample, we repeat the process for all days and take the mean of the Shapley values.

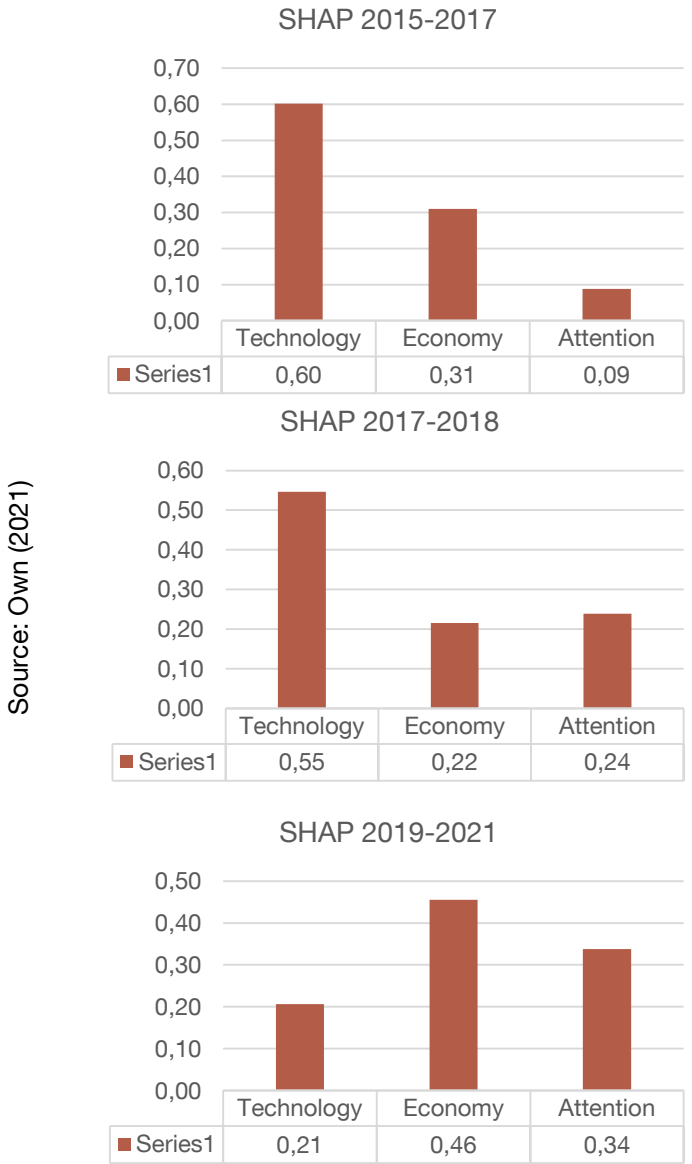
Return Formulation

BANCO DE ESPAÑA
Eurosistema

2016 detail

2018 detail

2021 detail



Source: Own (2021)

Importance of factors

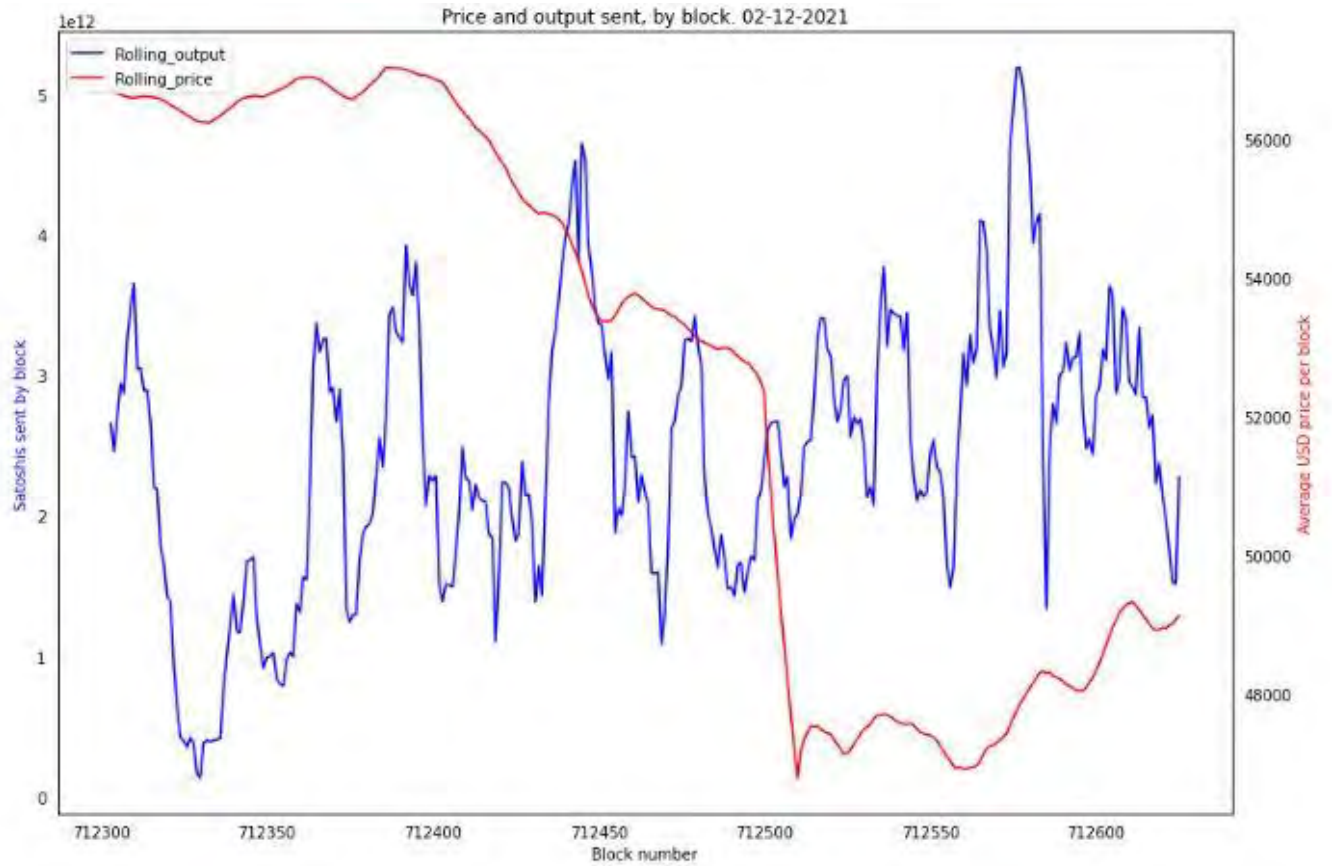
SHAP

- Importance: Difference wrt the average prediction of including or not including the variable, in %.
- In 2015-2018, the technological variables are the ones that have the greatest importance. In 2020-2021, they lose relevance.
- The variables related to sentiment gain importance as the years go by:
  - Google Trends clearly climbs positions.
  - It goes from having an effect of 5% in 2017 to 24% in 2021.
- Heterogeneity between market indicators:
  - At the aggregate level, there is no trend.

- With data obtained from University of Cardiff, we are **analyzing bitcoin data transaction by transaction**

Block	Day	Time	Sender	Receiver	Amount	Size
17839	02-12-2021	18:45	A	B	4500	0.001

- We could have averages of the technological **variables block by block**
  - Bitcoin's **blockchain produces blocks every 10 minutes approx.**, about 144 blocks per day



Next steps? Block by block

- **Opportunities:**
  - Studying intra-day price movements can reveal **details on order flows and liquidity that are not visible on a daily scale**
  - **High-Resolution Insights** into the price fluctuation dynamics
  - **Real-Time Response** Analysis: to global economic events, news, and policy changes in real-time.
  - Short-term Prediction. The block-by-block analysis could significantly enhance **short-term price prediction accuracy**

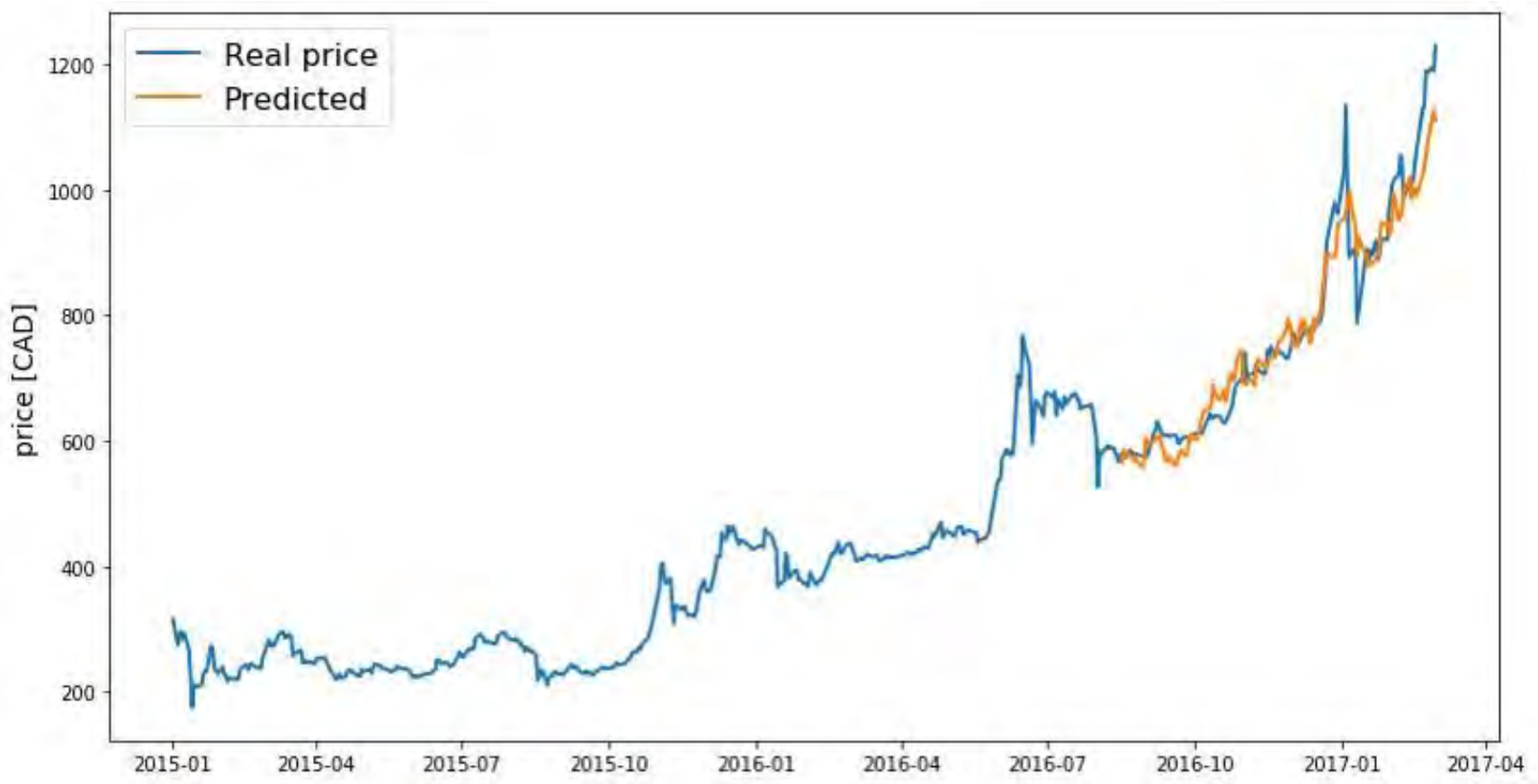
- **Challenges:**
  - **Size** of the data. 600 million transactions -> 700.000 blocks
  - **Time zone sensitivity,** the influence of different time zones and market hours can become more significant
  - **Noise** in the data, fluctuations that don't reflect underlying trends
  - **Overfitting risk**, it might pick up spurious correlations that don't hold up over time

- LSTM advantages with granular data

- The importance of the different variables in the formation of the bitcoin price changes substantially throughout the analyzed period.
  - Not only does its influence vary, but new unknown explanatory factors appear (that could explain that, in 2021, the LSTM model observes a greater error).
  - Compared to other financial assets, bitcoin - and by extension its namesakes - seem to exhibit behavior that is difficult to predict, making them a high-risk investment.
  - This lack of maturity advises that the financial authorities maintain conservative positions regarding their regulation to avoid the transmission of potentially risks.

- Other notable results are that technological factors have greater relevance only in the first observation periods.
  - Variables such as hash difficulty, block size, number of transactions or unique addresses have lost all their weight by 2021.
- In addition, in periods of high price volatility (2018 and 2021), public interest takes on a very notable role.
- Finally, the economic indicators are in an intermediate place and gold does not influence.

Period II: 2015-2017



Source: Own calculations (2021)

Return

Train-Validation-Test and cross validation

- We perform two exercises to uncover the determinants of bitcoin in each of the periods
  - **One Long forecast** : 6 months to test, 18 previous months to train, 1-2 months validate
  - **Rolling window**: 1 month to test, 6 months to train, half month to validate. Moving ahead predictions every week.

Parameters	Values
Input lags	1-5 days of memory
Activation function	Relu, Sigmoid, Tahn
Learning rate	0,01 to 0.0001
Number of hidden layers	1 to 3
Nodes in hidden layer	256, 128, 64
Dropout value	0.3 to 0.1
Longest training duration	60 seconds
Loss function	MSE

Return

Results

- LSTM presents a good performance in all three stages.
  - We measure its effectiveness through the RMSE, the square root of the squared differences between the predicted values and the observed values.
  - We get an RMSE between 5% and 18% of the price.

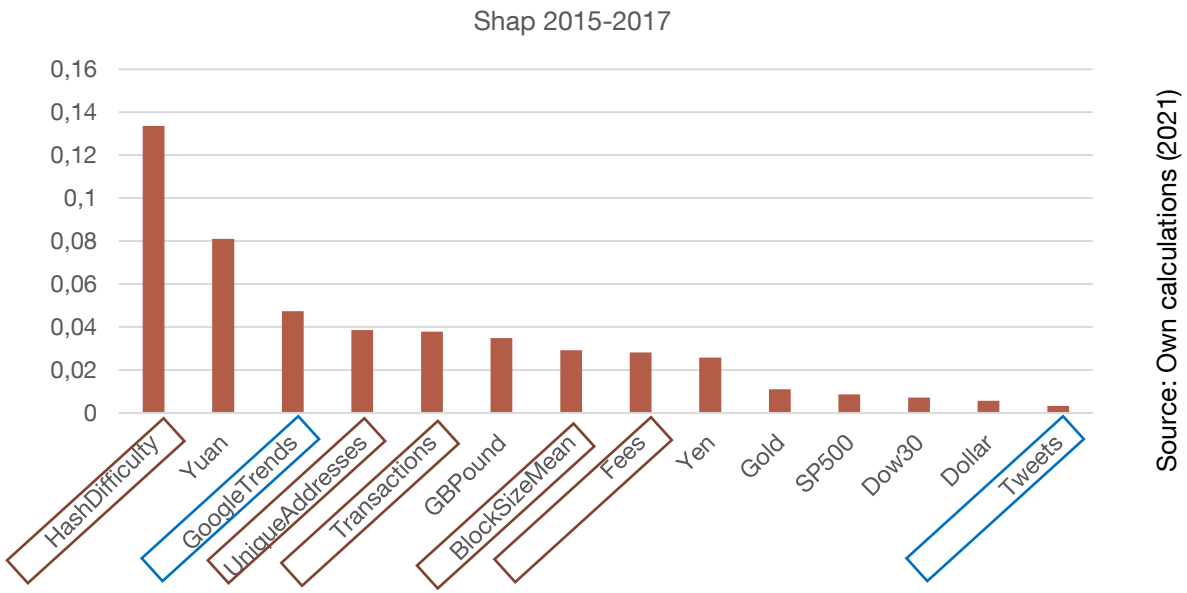
	RMSE	Test sample
Period I: Launch	5%	01/08/2016 - 01/04/2017
Period II: Expansion	12%	01/10/2017 - 01/04/2018
Period III: Consolidation	18%	01/01/2021 - 01/07/2021

- The model works reasonably well in all three stages, although there is a section of stage III in which the model perform worse (April 2021).

Return

Importance of factors in period I: 2015-2017

- Importance: Difference with respect to the mean prediction of including or not including the variable, in %.
- In 2015, the technological variables associated with bitcoin have considerable importance
  - *Hash difficulty* is the most important.
  - *Unique addresses, number of transactions, fees*, etc., appear in the top 8.
- In 2015, the variables of "interest" are further behind: eg Google Trends is the third.

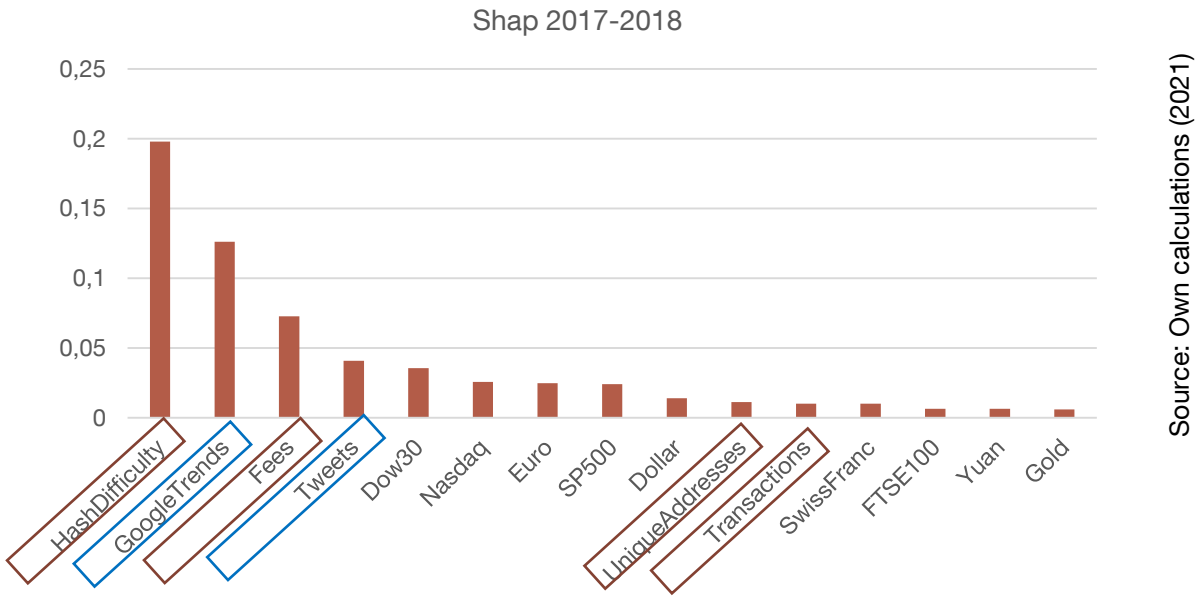


Source: Own calculations (2021)

Return

Importance of factors in period II: 2017-2018

- Importance: Difference with respect to the mean prediction of including or not including the variable, in %.
- In 2018, the technological variables associated with bitcoin are still on the top, although they lose some importance:
  - *Hash difficulty* is still the first, and *fees* advance positions*; the rest lose importance*
- In 2018, the variables of "interest" gain a lot of importance: *Google Trends* is the second and Tweets go up.

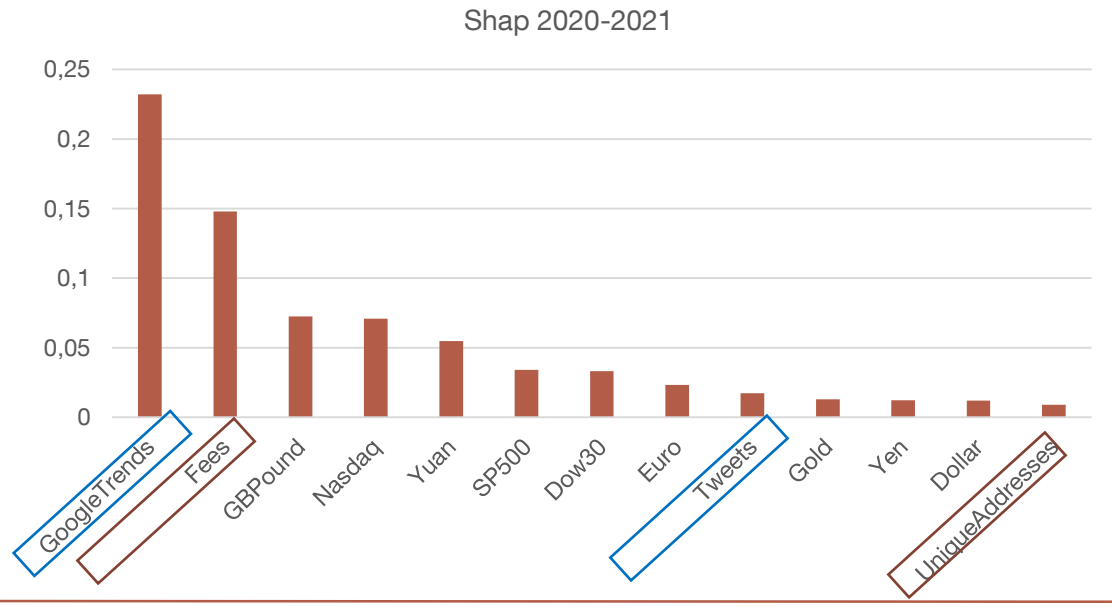


Source: Own calculations (2021)

Return

Importance of factors in period II: 2020-2021

- Importance: Difference with respect to the mean prediction of including or not including the variable, in %.
- The technological variables associated with bitcoin lose a lot of importance:
  - *Fees* remains at the top, but the rest practically disappear from the top 15.
- The variables of "interest" definitely take over:
  - *Google Trends* is the first and with great impact.



Source: Own calculations (2021)

Return

Next steps? Block by block

LSTM Advantages

- LSTMs **can learn complex patterns**, can potentially excel at short-term predictions on granular data.
- LSTMs can **be better equipped to handle this noise** and still detect underlying patterns.
- LSTMs can capture **temporal dependencies at different scales**, which can be beneficial.

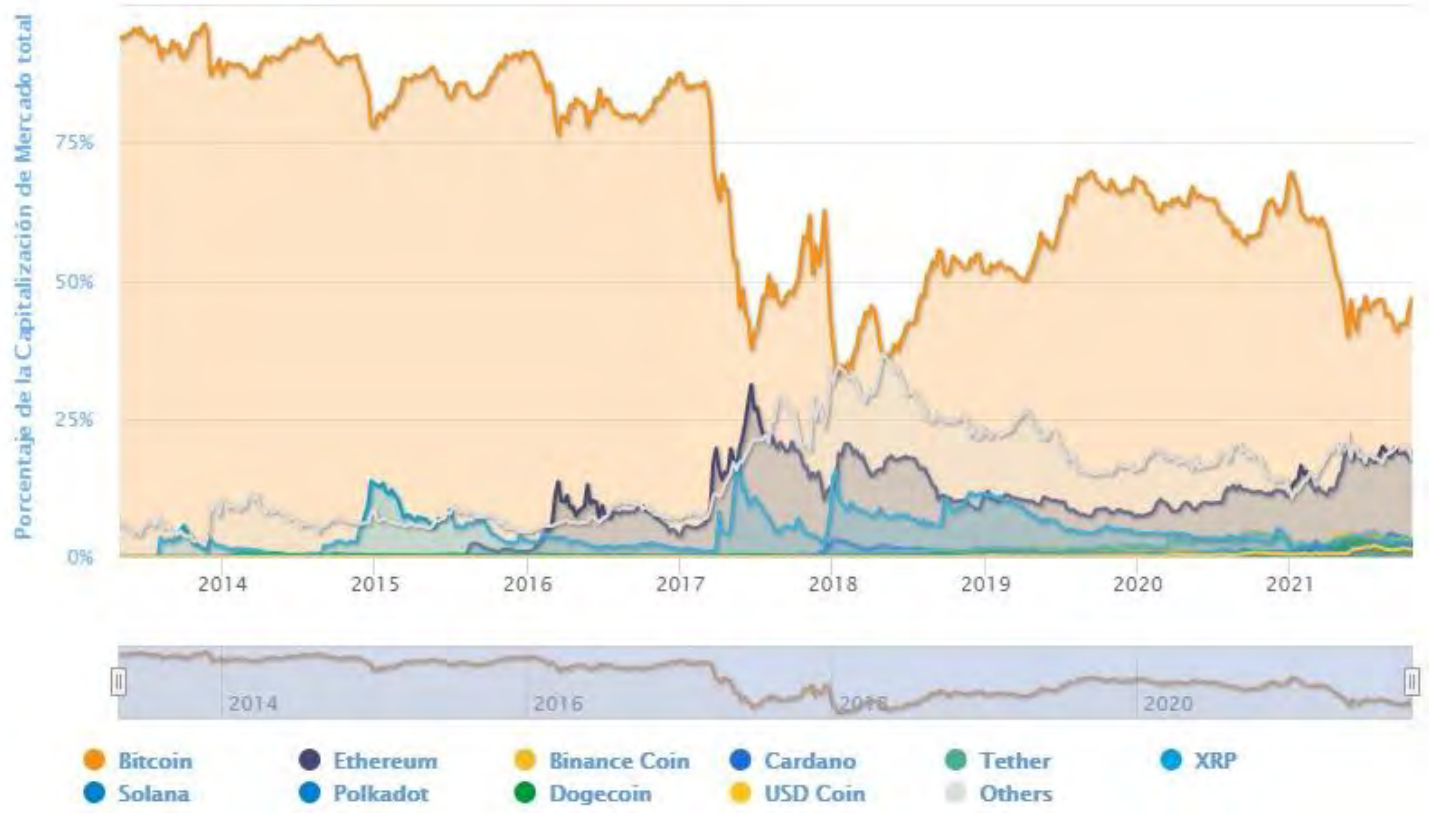
LSTM challenges

- **Computational expensive**
- **Interpretability**: But we can solve this!

Return

Some Preliminary Considerations

Relative weight of the main crypto assets in the total market capitalization: 29/5/2019 – 20/10/2021

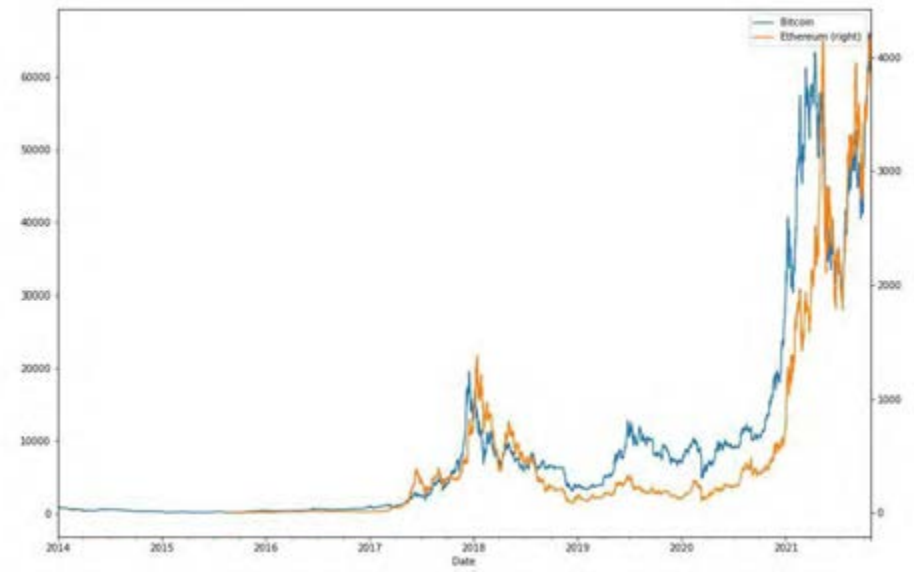


Source: Coinmarketcap.com (2021)

Return

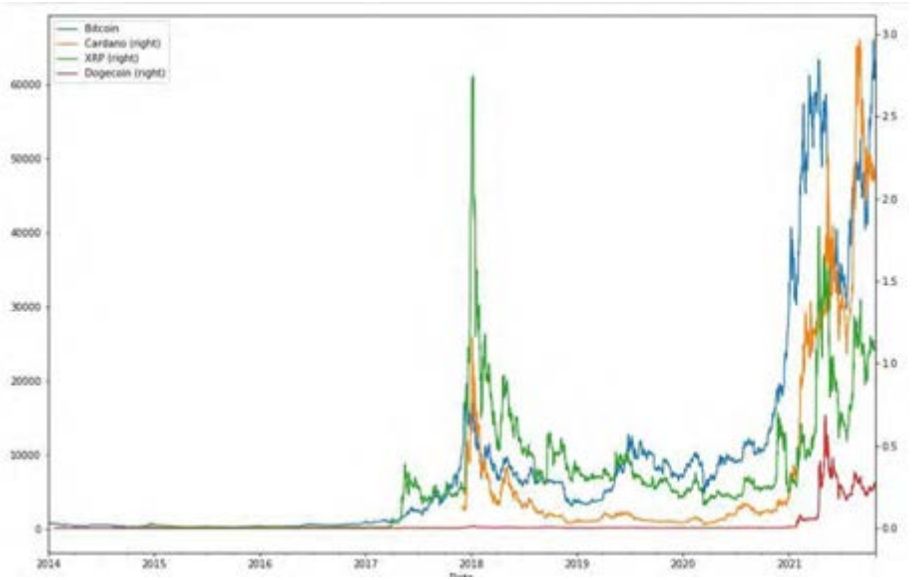
Some Preliminary Considerations

Historical evolution of the price of bitcoin and Ethereum (USD)



Source: Own (2021). Data from Coinmetrics

Historical evolution of the price of bitcoin, Cardano, XRP and DogeCoin (USD)



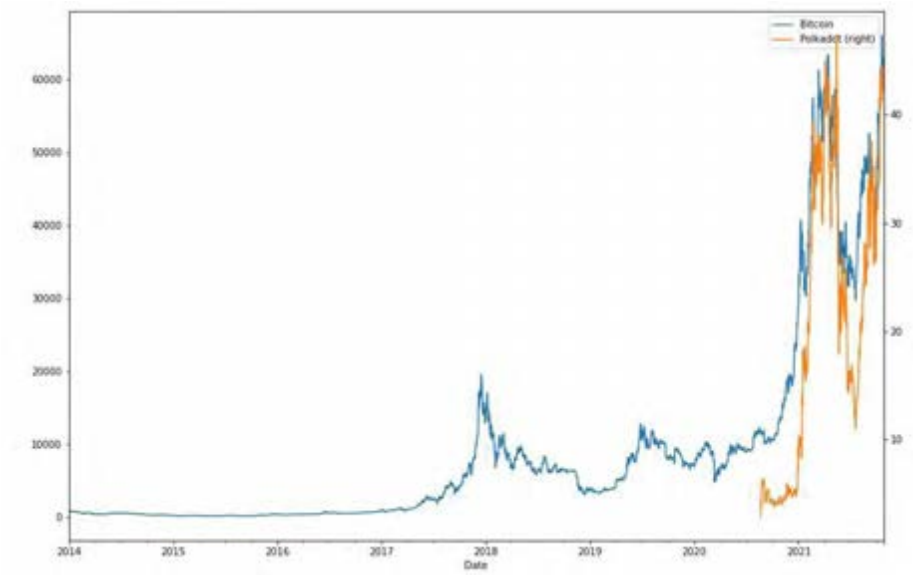
Source: Own (2021). Data from Coinmetrics

Due to its weight and driving effect, the analysis of bitcoin could serve to illustrate possible common factors that influence or determine the general behavior of this market.

Return

Some Preliminary Considerations

Historical evolution of the price of bitcoin and PolokaDot (USD)

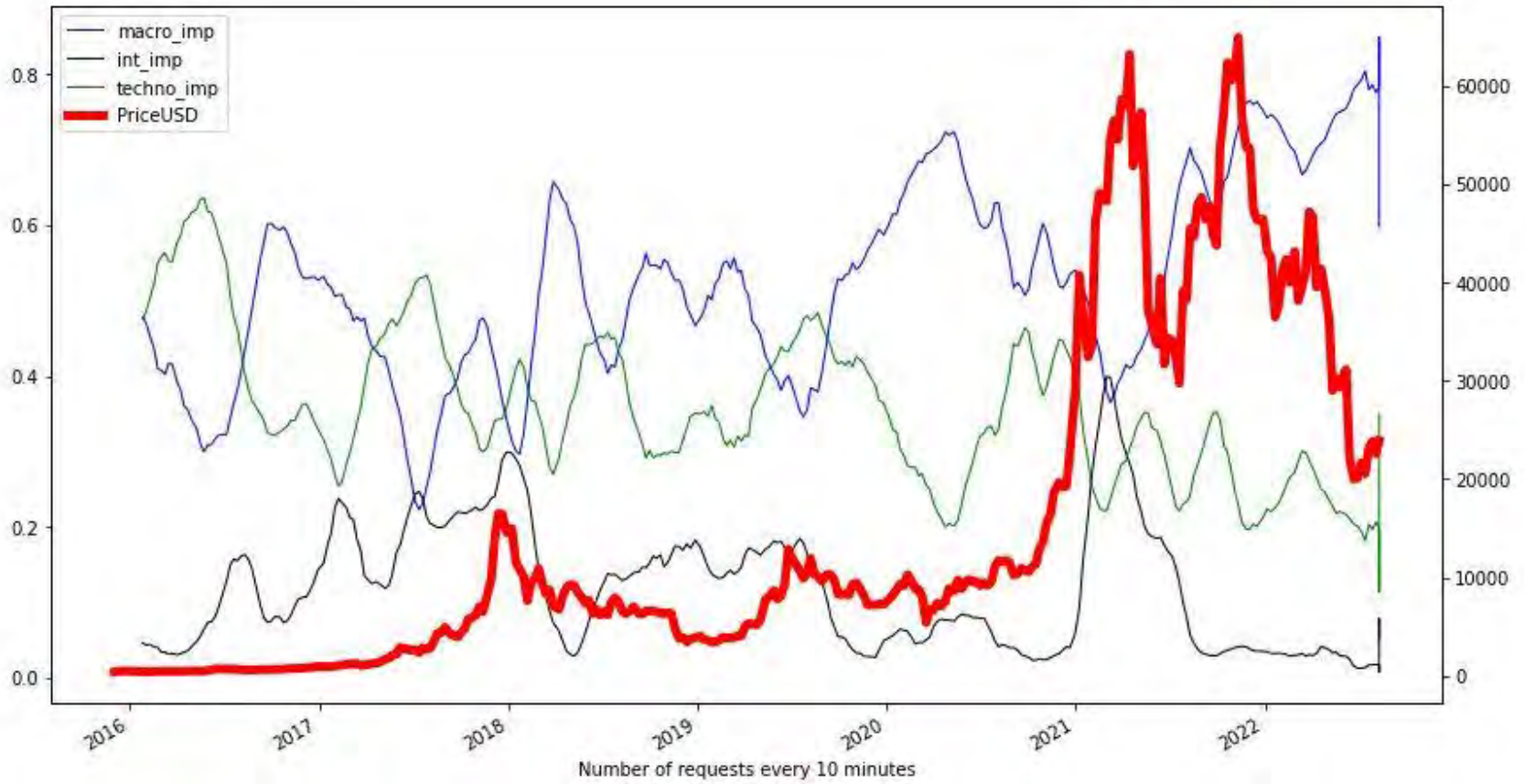


Source: Own (2021). Data from Coinmetrics

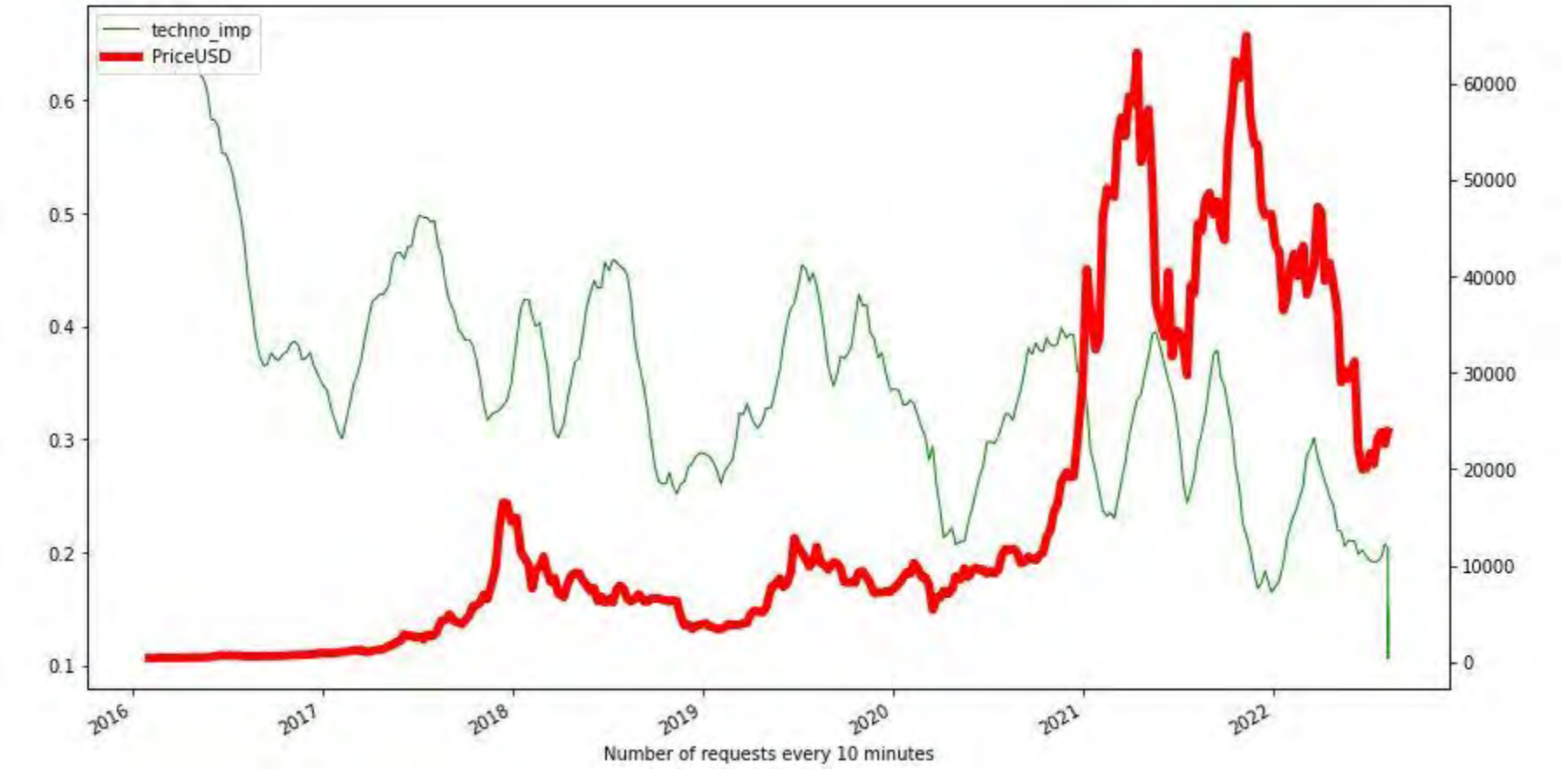
Due to its weight and driving effect, the analysis of bitcoin could serve to illustrate possible common factors that influence or determine the general behavior of this market.

Return

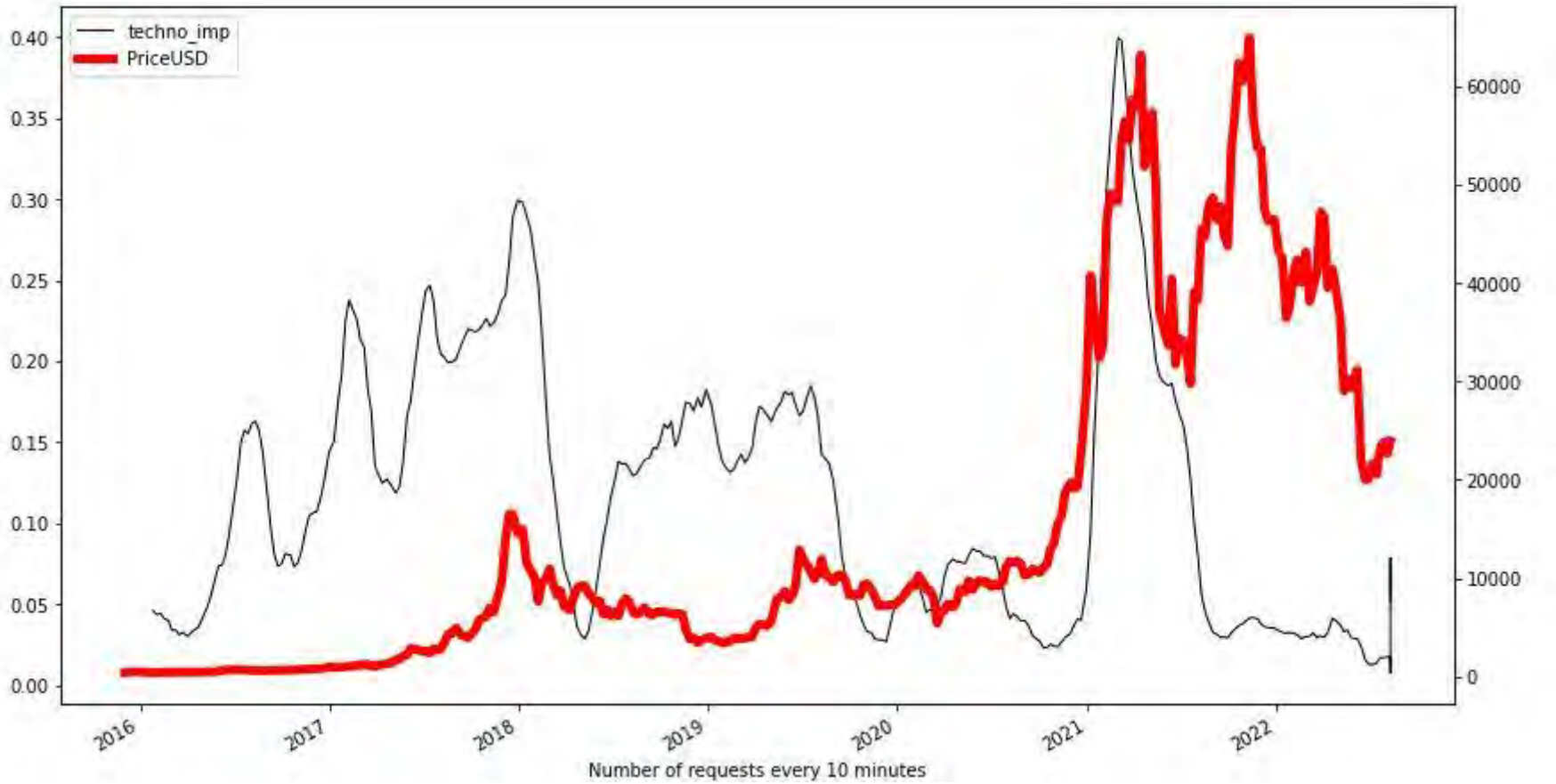
Rolling window results



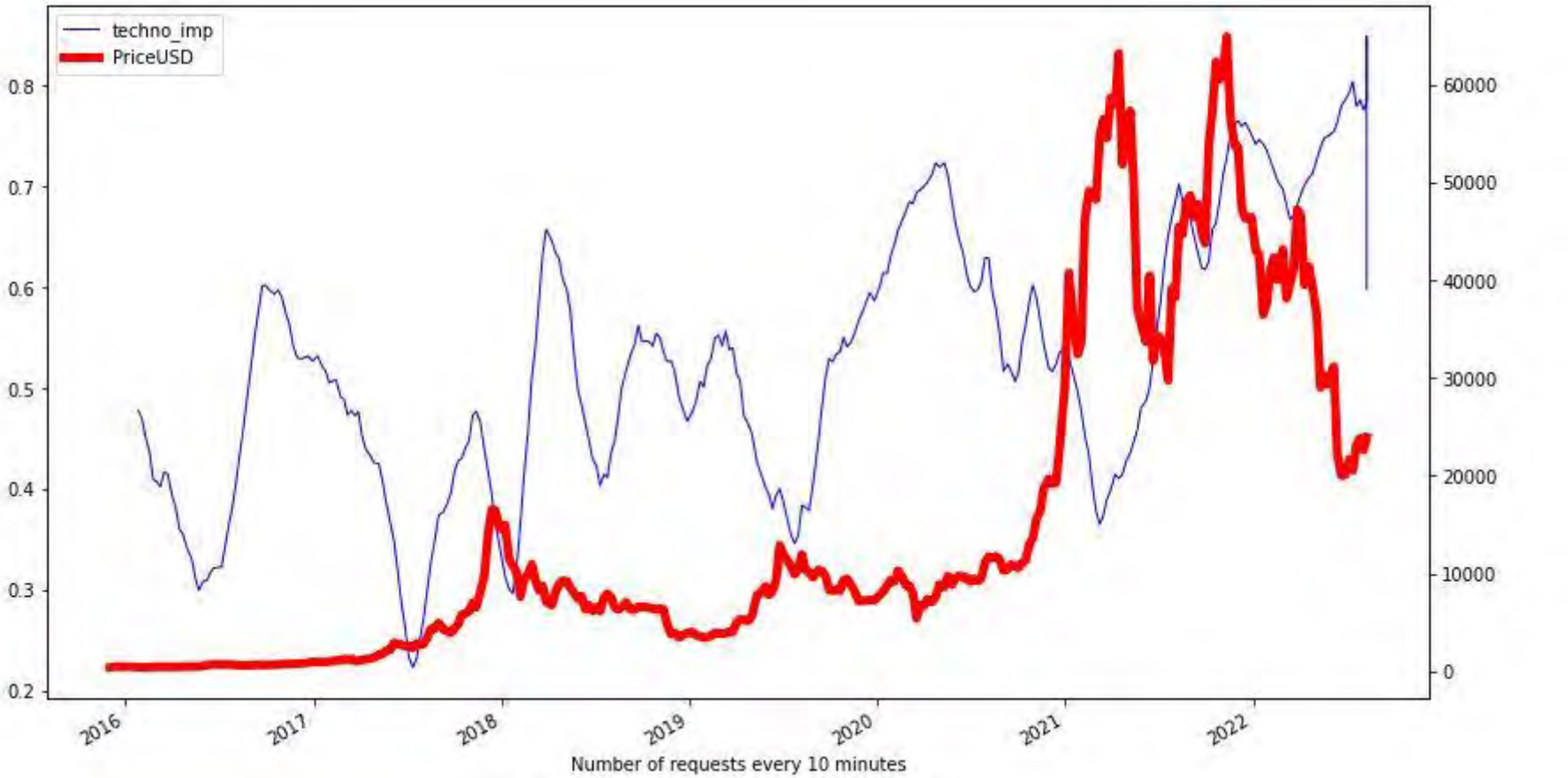
Rolling window results

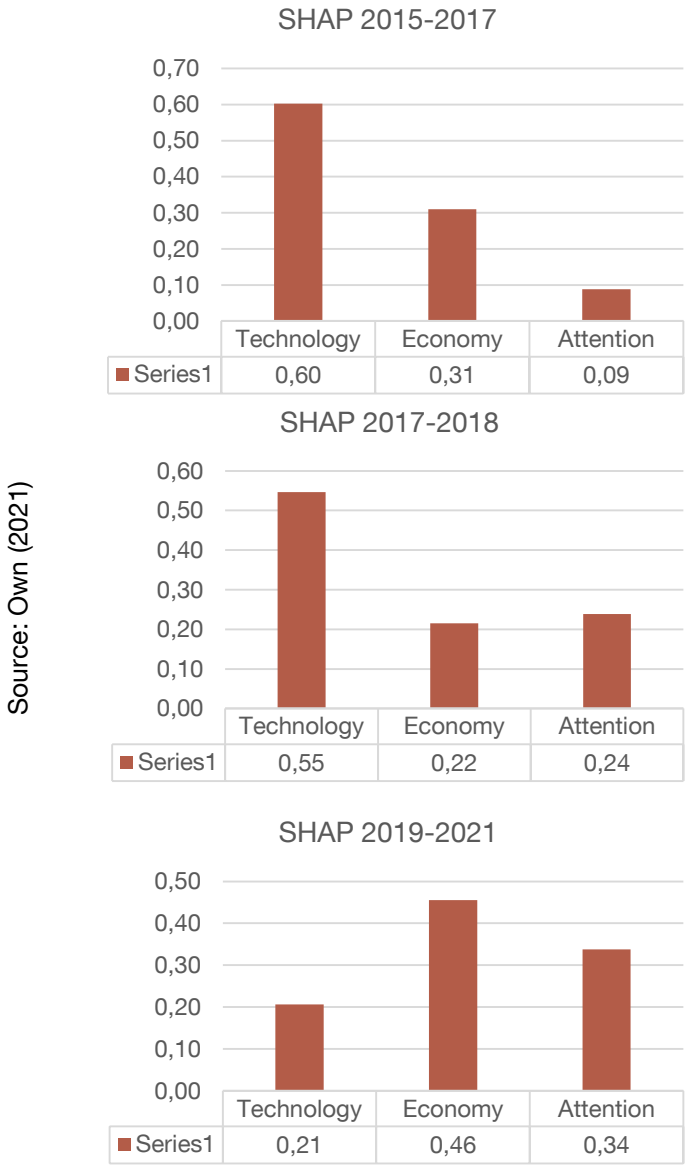


Rolling window results



Rolling window results





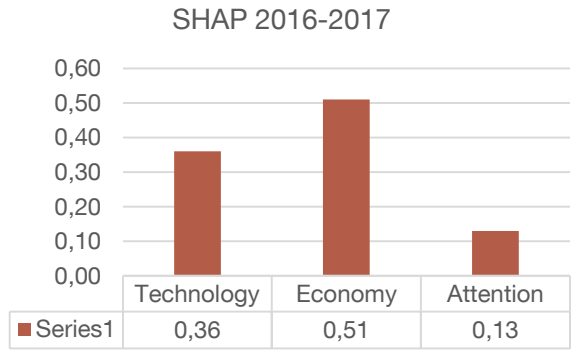
Source: Own (2021)

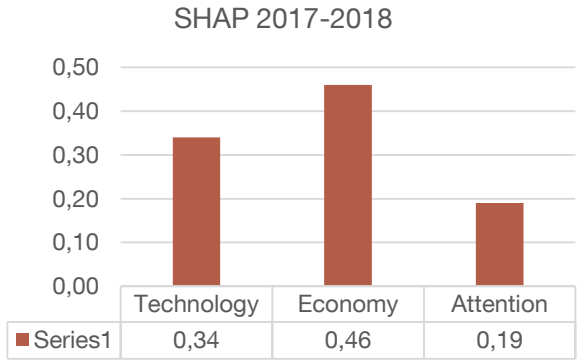
Importance of factors

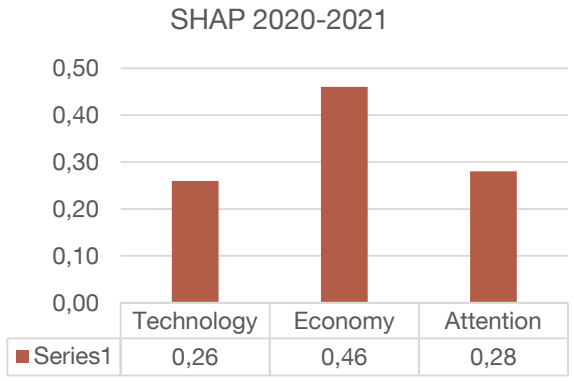
SHAP

- Importance: Difference wrt the average prediction of including or not including the variable, in %.
- In 2015-2018, the technological variables are the ones that have the greatest importance. In 2020-2021, they lose relevance.
- The variables related to sentiment gain importance as the years go by:
  - Google Trends clearly climbs positions.
  - It goes from having an effect of 5% in 2017 to 24% in 2021.
- Heterogeneity between market indicators:
  - At the aggregate level, there is no trend.

Return







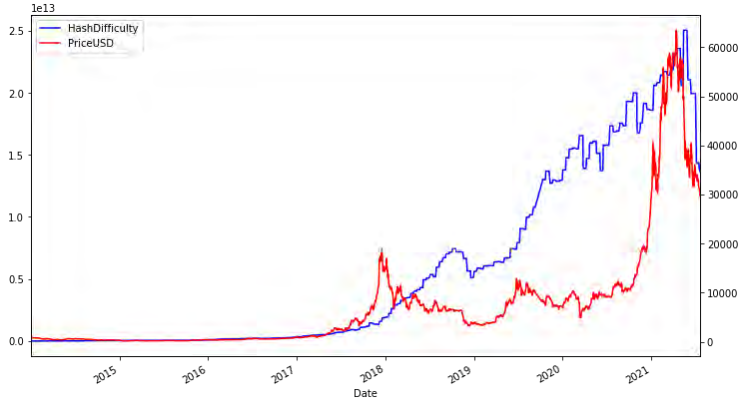
Importance of factors

SHAP "Rolling Windows"

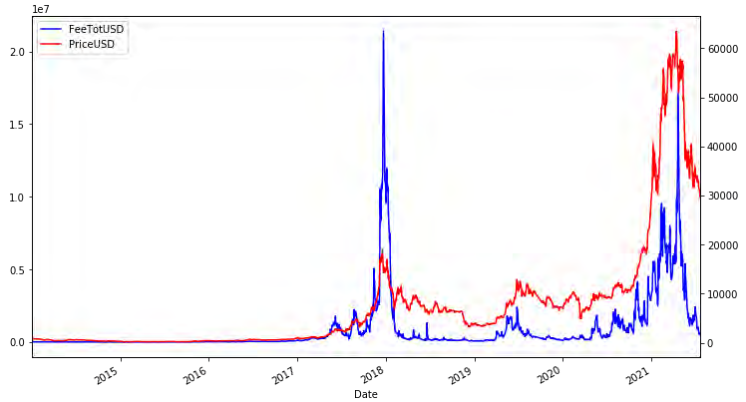
- The result shows again that the importance of the technology factors decreases and that the interest factors increase.
- No tendency in macro factors

Return

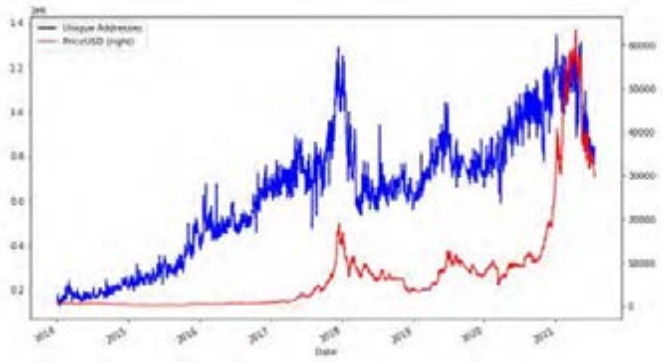
Graphs with other variables of interest



Source: Own (2021)



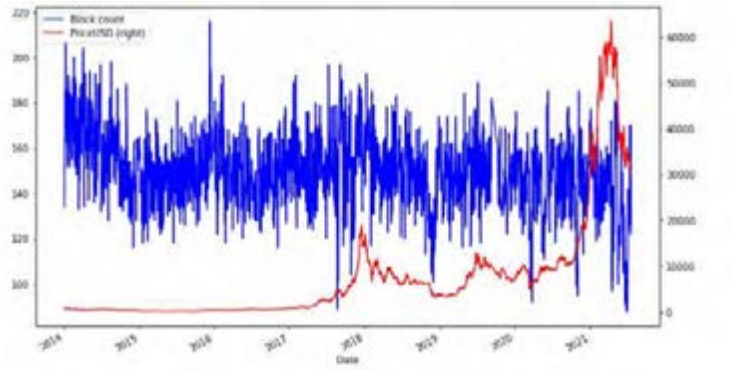
Source: Own (2021)



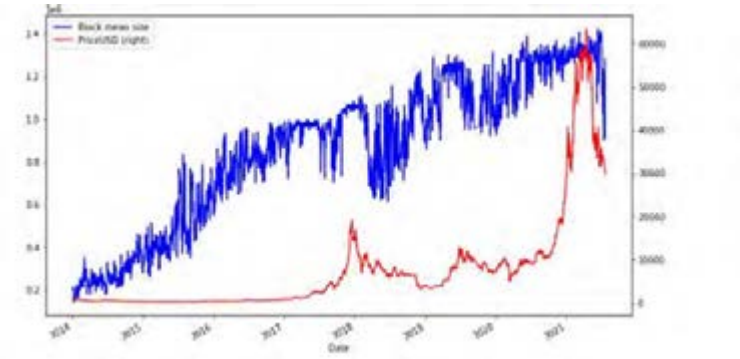
Source: Own (2021)

Return

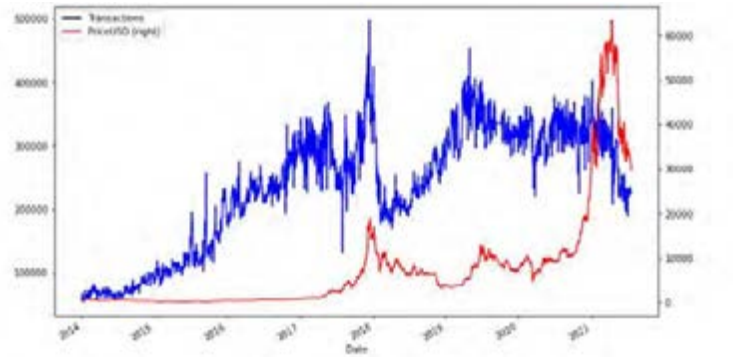
Graphs with other variables of interest



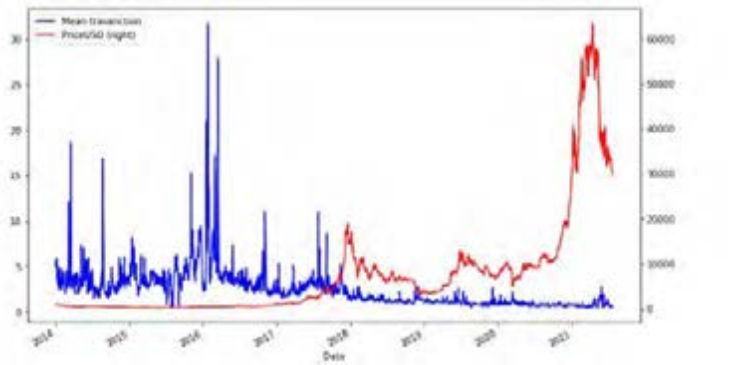
Source: Own (2021)



Source: Own (2021)



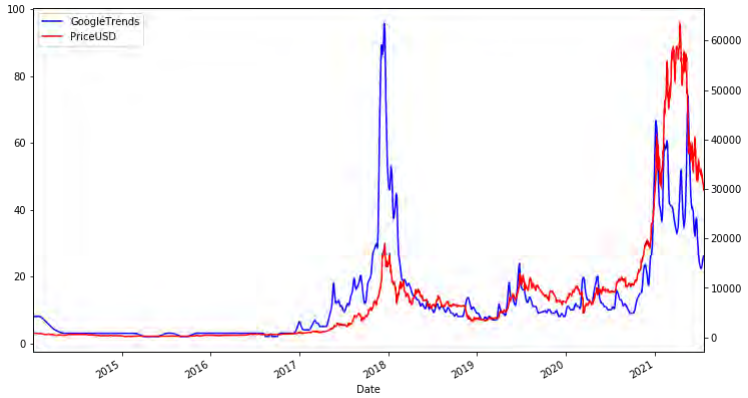
Source: Own (2021)



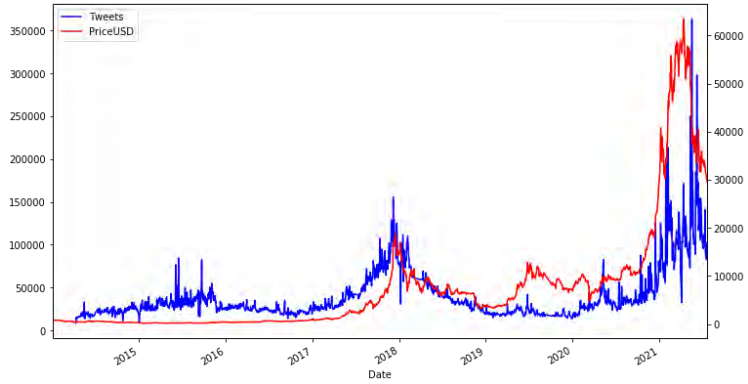
Source: Own (2021)

Return

Graphs with other variables of interest



Source: Own (2021)



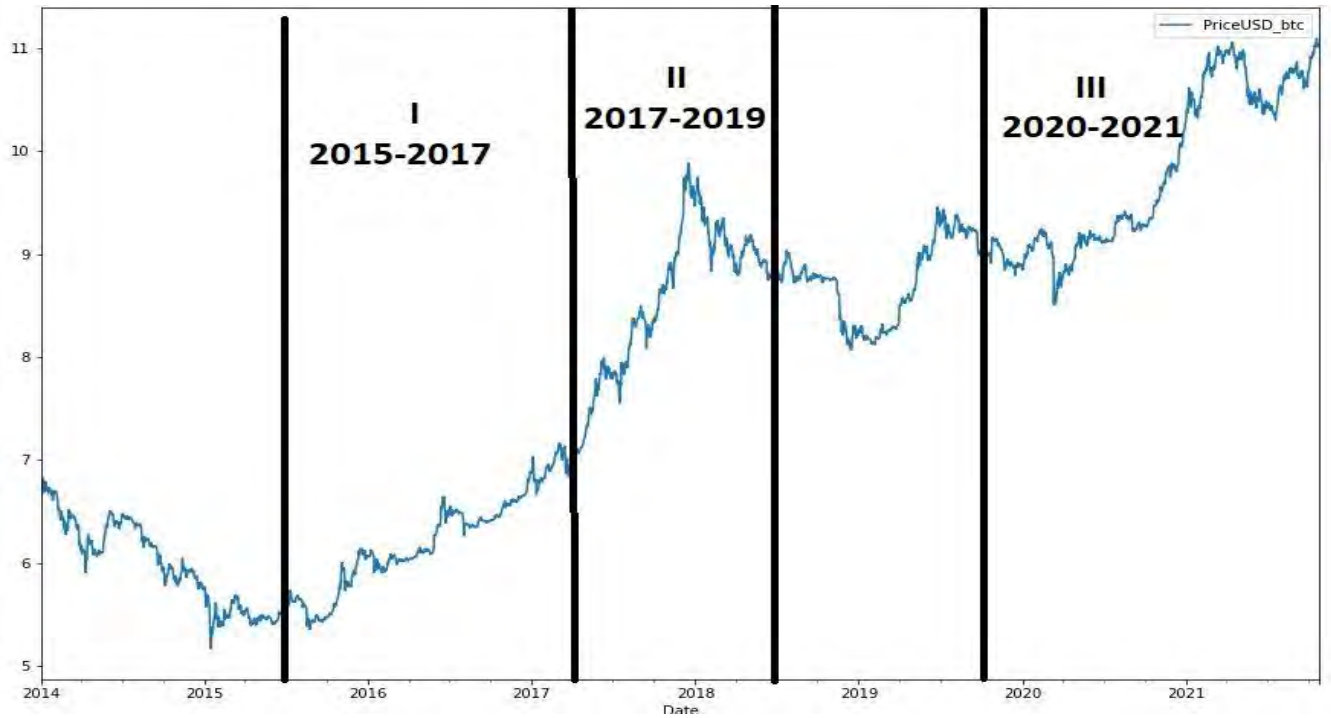
Source: Own (2021)

- **Mean difficulty:** The mean difficulty of finding a hash that meets the protocol-designated requirement (i.e., the difficulty of finding a new block) that day. The requirement is unique to each applicable cryptocurrency protocol. Difficulty is adjusted periodically by the protocol as a function of how much hashing power is being deployed by miners.

- **Mean Hash rate:** The mean rate at which miners are solving hashes that day. Hash rate is the speed at which computations are being completed across all miners in the network. The unit of measurement varies depending on the protocol.

Mean Hash rate:(BlkCnt / 144) * DiffMean * ((2^32 / 10^12) / 600))

Three stages



Source: Own (2021)

- **Stage I:** 01/04/2015 a 01/04/2017
- **Stage II:** 01/01/2016 a 01/05/2018
- **Stage III:** 01/11/2019 a 01/07/2021

Return

SHAP (Formulas)

- The Shapley value or contribution $\emptyset$ of a given feature *i* in a prediction *p* is :

$$\emptyset_i = \sum_{S \in N/i} \frac{|S|!\,(n - |S| - 1)}{n!} (p(S \cup i) - p(S))$$

  - S represents a coalition of features
  - N is the total number of features,
  - $S \in N/i$ represents all possible coalitions of features excluding feature *i*, considering all possible orders,
  - $p(S \cup i) - p(S)$ represents the difference in the predicted outcome *p* when we consider a particular coalition of features and feature *i* minus the predicted outcome when we consider the coalition of features without feature *i*.

- The term $\frac{|S|!(n-|S|-1)}{n!}$ assigns different weights to the differences, depending on the features that are in the set $|S|!$, the features that have to be added $(n - |S| - 1)$, and all normalized by the features that we have in total.

Return

SHAP

- SHAP is a technique that measures the contribution of a variable to the predicted outcome, on a given day compared to the average prediction.
  - These contributions are called the Shapley values.

- Once we have the Shapley values for each variable and for each day, they can be added to obtain the final importance of the variable.
  - Therefore, SHAP can be used as both a local and global interpretability technique.

- We can understand Shapley values from a game theory perspective:
  - The game is to reproduce the result of the model (in our case, the price of bitcoin).
  - The players are all possible coalitions of variables.
  - The reward is the contribution that each coalition has to the result of the model.

Return

Feature Importance

- It measures the impact of each feature in the dataset on a given performance metric.
  - Since the target variable in our data is a binary variable, we use as performance metric the ROC-AUC.

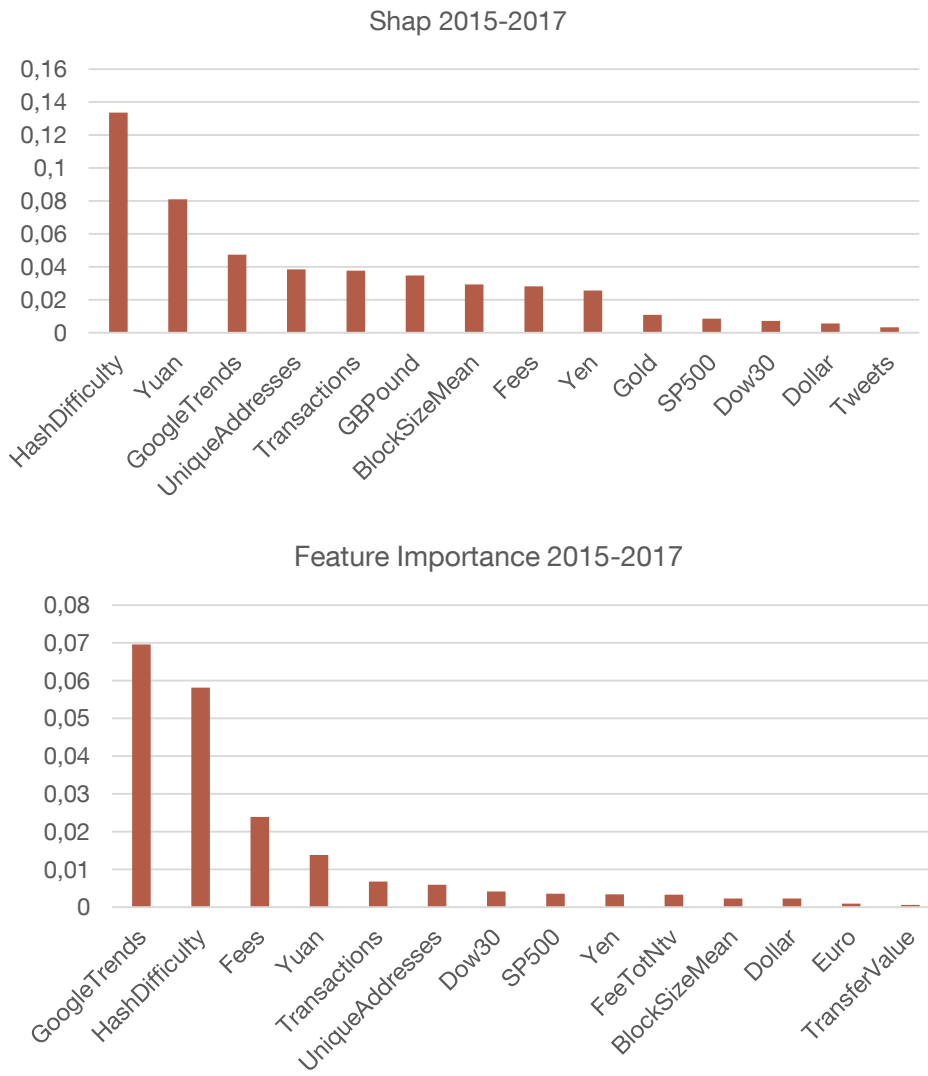
- First, we need a trained model. Then, for each of the features of the dataset, we repeat the following:
  - For the chosen feature, we shuffle its values and we compare the predictions with the original values and with the new shuffled values.
  - We repeat the process 10 times to make sure that a specific shuffling of the feature is not leading our results.
  - After 10 iterations, we compare the performance metric of the prediction with the original values and with the shuffled values.

- **If the feature is important,** then we should observe a considerable impact in ROC-AUC**.**
- **If the feature is not important,** the ROC-AUC should be similar when using shuffled values instead of original values.

- The method is considered **of global interpretability**, since it tells us the importance of a certain feature in the entire dataset, and cannot tell us the importance of that feature for a given individual.
  - One of the drawbacks: It cannot indicate us the direction of the effect of a given feature.

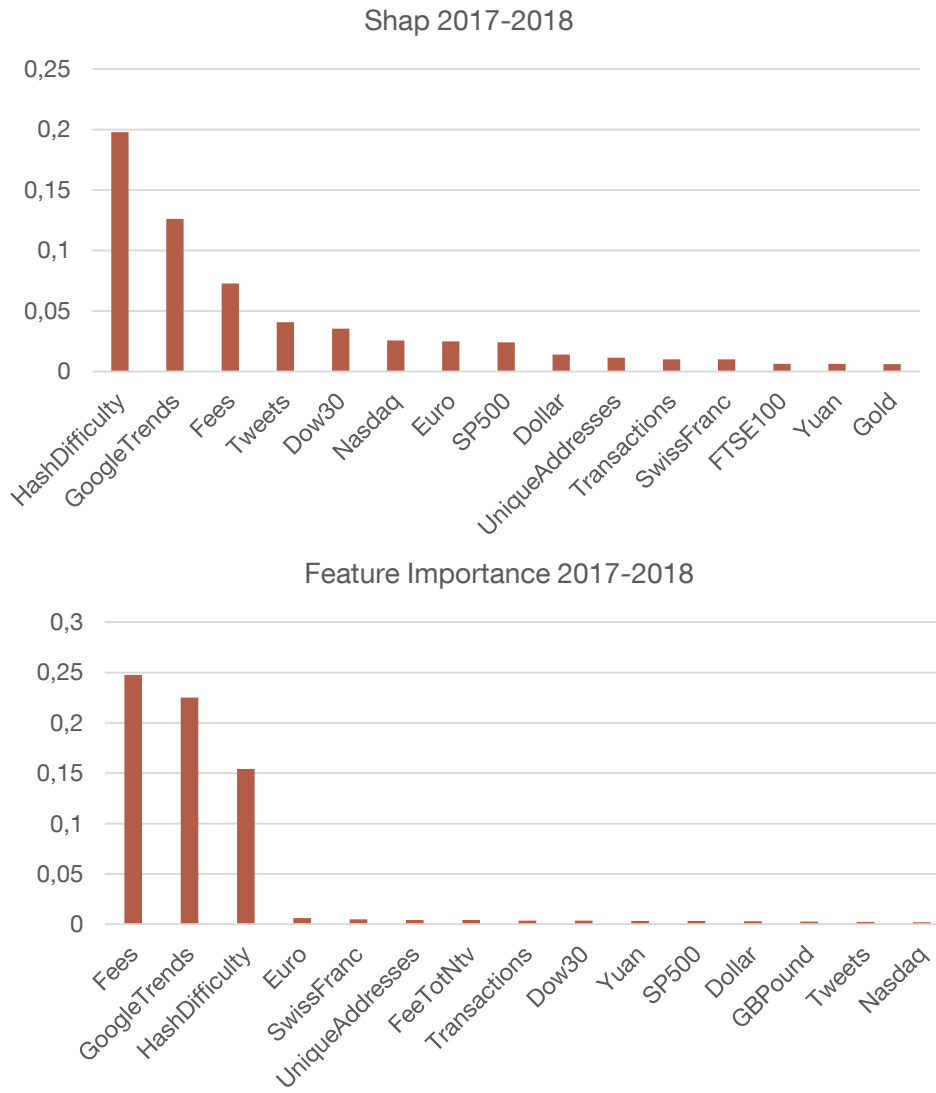
Return intro Return

Shap vs Feature Importance 2015-2017



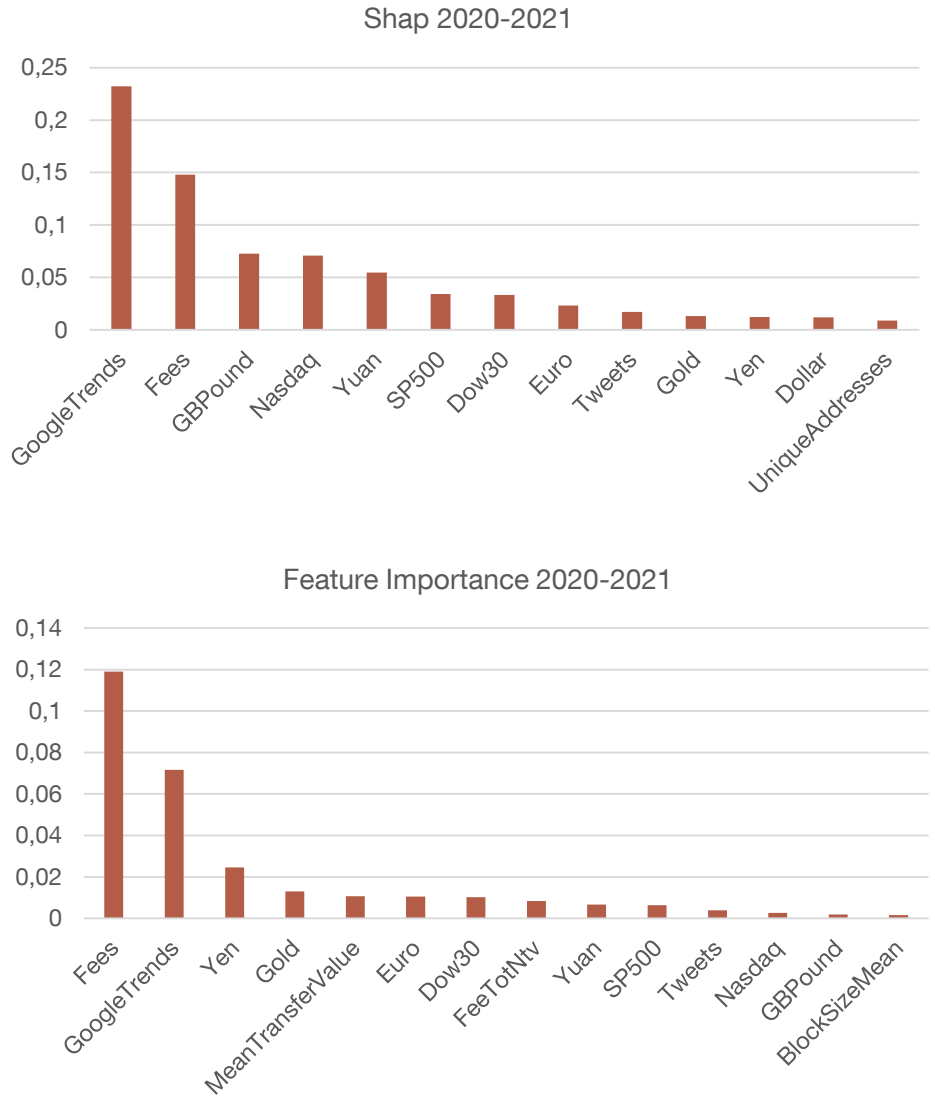
Source: Own (2021)

Shap vs Feature Importance 2017-2018

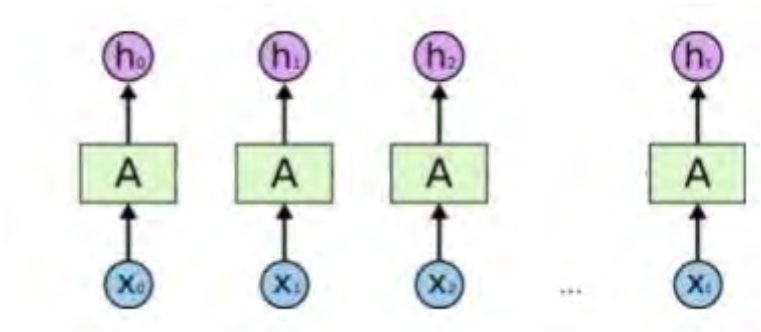


Source: Own (2021)

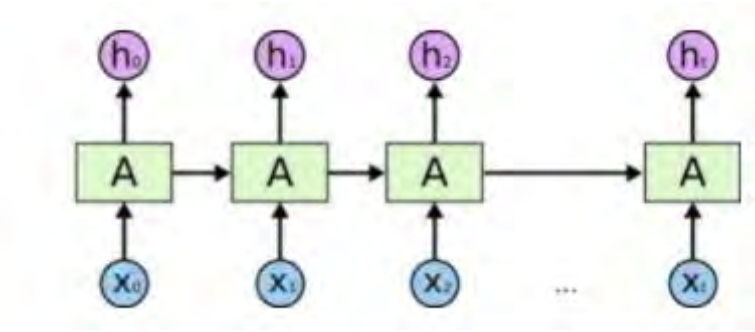
Shap vs Feature Importance 2020-2021



Source: Own (2021)



Neural Network



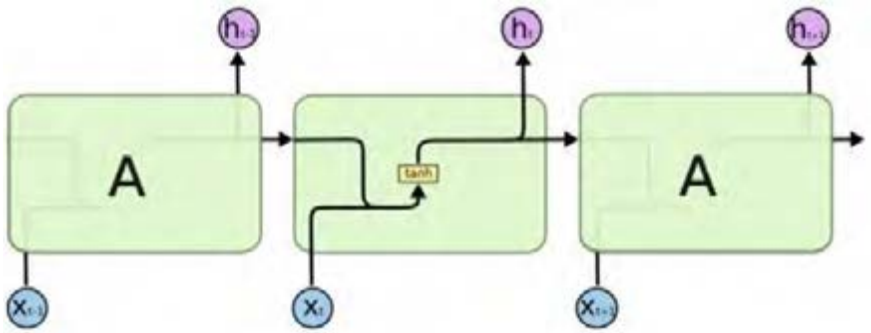
Recurrent neural network

The gradient problem: Occurs when the values of a gradient are too small and the model stops learning or, as a result, takes too long to learn. We can use LSTM.

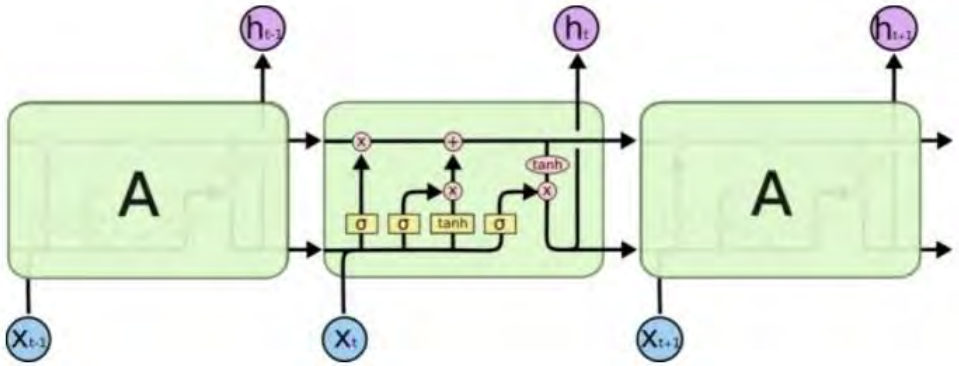
Return

LSTM model

- LSTM contains information outside of the normal flow of the network in a closed cell or memory.
- The cell decides whether to store or delete information (that is, whether it opens the doors or not), depending on the importance it assigns to the information.
  - There are three doors: oblivion, entrance and exit. These gates determine whether past information is removed (forgetting gate), whether or not to allow new information (input gate), and what information is carried over to the next state (output gate).



Recurrent neural network



LSTM

Return

References

- Panagiotidis, T., Stengos, T., & Vravosinos, O. (2018). On the determinants of bitcoin returns: A LASSO approach. Finance Research Letters, 27, 235-240.
- Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. International Journal of Forecasting, 37(1), 28-43.
- Hayes, A. S. (2019). Bitcoin price and its marginal cost of production: support for a fundamental value. Applied Economics Letters, 26(7), 554-560.
- Ji, S., Kim, J., & Im, H. (2019). A comparative study of bitcoin price prediction using deep learning. Mathematics, 7(10), 898.
- Sovbetov, Y. (2018). Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litcoin, and monero. Journal of Economics and Financial Analysis, 2(2), 1-27.
- Panagiotidis, T., Stengos, T., & Vravosinos, O. (2018). On the determinants of bitcoin returns: A LASSO approach. Finance Research Letters, 27, 235-240.
- Zhu, P., Zhang, X., Wu, Y., Zheng, H., & Zhang, Y. (2021). Investor attention and cryptocurrency: Evidence from the Bitcoin market. Plos one, 16(2), e0246331.

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Property classification with administrative data: The case of the Chilean household price index[1]

Juan José Balsa and Javiera Vasquez,
Central Bank of Chile

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Working paper

Property classification with administrative data: The case of the Chilean House Price Index[1]

Juan Jose Balsa[2]
Javiera Vásquez[3]

Department of Microdata
Division of Statistic
April 2023

Summary

Working with administrative data can be particularly challenging when it is incomplete or lacks critical information. In this paper, we present a specific case study involving the Chilean House Price Index, that is computed with a mixed stratification method where data classification into apartments and houses is necessary for the robustness of the method. In compiling this index, the Central Bank of Chile uses administrative information from the SII[4] on all effective transactions of residential properties.

We aim to provide a step-by-step approach to data classification, including testing various algorithms and optimizing the best ones for the data. Our approach helps us to classify over 3 million observations in a comprehensive and fast way. The paper discusses the House Price Index methodologies, the different method for the classification of the property, the classification algorithms used, and provides codes and insides of the challenges we encountered during the process.

[1] We thank Maria Fernanda Castillo for their contributions to this work.
[2] Juan José Balsa (jbalsa@bcentral.cl), analyst of the financial microdata group of the BCCh, Msc. in Economics from the University of Groningen and Msc. in economics analysis from the University of Chile.
[3] Javiera Vasquez (jvasquez@bcentral.cl), Head of the financial microdata group of the BCCh, Msc in Economics from the University of Chile.
[4] From the spanish "Servicio de Impuestos Internos" (Tax Administration Office)

Contents

Introduction 3
House Price Indices 4
I. Main indices methodology 5
I.1 Mean or Median 5
I.2 Stratification 5
I.3 Repeat Sales 6
I.4 Hedonic Regression 7
II. Prices Indices around the world 8
IPV Methodology 13
I. Database 13
II. IPV Calculation 14
Classification Algorithm 15
I. What is a Classification algorithm and how it works 15
II. Property Classification with administrative data 16
IPV Classification Methodology 17
I. Flow 1: Expert Criteria 18
II. Flow 2: Data-Driven classification 20
III. Bonus Track: ML Algorithm for any classification problem 24
IV. Results and uses of the data 25
Conclusion 29

Introduction

Housing represents a significant source of wealth and debt for households globally. For the case of Chile in 2021 the 63.7% of household assets correspond to real estate assets, and 54.3% of debt is constituted by mortgage debts[5]. Thus, variations in housing prices can have significant implications for household finances, with repercussions in the financial system, consumption, and the national economy.

Therefore, it is reasonable to continuously monitor the evolution of housing prices due to their effects on the country's financial stability. In fact, after the severe consequences of the subprime mortgage crisis, there has been increased interest in the systematic analysis of price and activity indicators in the real estate sector.

In addition, the increases in monetary policy rates during 2022, caused by the global high inflation scenario that affects debt and access to credit for households and companies in this sector, make it highly relevant to review and continuously improve this indicator.

Different methodologies for constructing housing price indices have been documented in the literature, and the choice of the most appropriate one will depend mostly on the data available and the use of the index . For example, the Eurostat handbook indicates that if the purpose is to estimate the dynamics of prices to track housing market inflation, the optimal way is to gather information on current transaction prices. On the other hand, if the objective is to estimate the balance of a country's wealth, information on housing sales should be supplemented with information on the stock of properties that are not for sale. However, in practice, the availability of information determines the methodology to be used.

Within the mission of the Central Bank of Chile (BCCh) is to ensure financial stability, and analyzing the real estate sector is crucial, where a price index is a central tool. Therefore, starting in 2014, the BCCh developed the Housing Price Index (IPV), which is currently calculated on a quarterly basis and published in the Statistical Database. The main objective of this index is to study the behavior of prices for urban residential properties in the Chilean market.

The IPV[6] is calculated using the mixed stratification method[7], where the 27 strata are defined based on seven geographic zones, housing condition (new-used), and housing type (house or apartment)[8]. The general IPV, which represents transactions throughout the national territory, is published along with specific indicators by geographic zone, property condition, and housing type. In compiling this index, the BCCh uses administrative information from the SII[9] on all effective transactions of residential properties.

A central part of the process to ensure the correct generation of the IPV is the stage of cleansing and imputing administrative records sent by the SII in both the F2890 form and

[5] Chilean Household Financial Survey available on Encuesta Financiera de Hogares (EFH) - Banco Central de Chile (bcentral.cl)
[6] From the spanish "Índice de Precios de Vivienda" (Household Price Index)
[7] For more detailed information, please refer to Section two of this document.
[8] The strata of new houses in the city center of Santiago has very few transactions so it is eliminated.
[9] From the spanish "Servicio de Impuestos Internos" (Tax Administration office)

the Real Estate Registry (CBR). These cleanings are complemented with the creation of new variables that are not in the records but are requiered for the generation of the strata, such as classification by housing type, and the condition of the property (new or used). Given the complexity of the previously described process and in line with delivering more and better statistical information, the BCCh during 2021 conducted a review of the cleansing and imputation processes of the information provided by the SII, along with the incorporation of new sources of information to enrich the IPV calculation. The review led to a series of improvements in the calculation of the indicator, reducing its volatility. The objective of this document is to present one of the key stages of the process of creation of the IPV, the property type classification among houses and apartments.

Classifying administrative data can be challenging, and in the case of the IPV, the challenge was to differentiate between houses and apartments. Various techniques were used for this classification, including manual classification, regular expressions, close neighbors, text mining, and machine learning algorithms such as logistic regression, random forest, k-means, k-nearest neighbors, and support vector machines.

The first section of the document describes house price indexes and their characteristics, main methodologies, and some international experiences. The second section presents a summary of the IPV methodology. The third section discusses classification algorithms and their evolution in the literature. Finally, in the fourth section, we make a zoom in the house type classification, the different approaches, we present a ready to use toolkit for others to use and some results.

House Price Indexes

A price index is an indicator that reflects the evolution over time of the price level of a segment of the economy. According to the Eurostat manual (2013), for a price index to be useful in the decision-making of various economic agents, it must meet the following characteristics: i) transparent methodology based on objective and documented criteria, ii) observed prices from actual transactions, and iii) easy access for user monitoring.

Housing is a capital good that has certain characteristics that make it difficult to construct a price index for this market. As the Eurostat Housing Price Index Manual (2013) states, houses are fundamentally heterogeneous since there are no two identical houses, either due to their physical characteristics or their location. This implies that the price of one property is not necessarily a good predictor of the price of another.

In addition to this, there is the problem of temporality, as it is not possible to exactly equate two properties due to the depreciation they undergo and the repairs or changes that owners may make over their useful life.

The house price index (HPI) consists of an indicator that measures the change in the price of residential real estate over time. It is often used as an indicator of the health of the housing market and can provide information on the state of the economy.

According to Himmelberg, Mayer, and Sinai (2005), the HPI is typically calculated from price data on actual home transaction, considering factors such as location, size, property

type, and overall condition. These data are then compiled into a single figure, representing the average change in housing prices in each area. This provides a useful tool for tracking trends and changes in prices over time.

The relevance of the HPI in the real estate sector can be diverse. For example, for homeowners, the HPI is used as an indication of the current value of their property, which can be useful to make decisions about selling or borrowing. Englund, Hwang, and Quigley (2002) note that the HPI can help homeowners assess the equity they have in their homes and make informed decisions about their financial future. For real estate agents, on the other hand, the HPI can help them set sale prices and predict housing demands. Meanwhile, for investors the HPI is a key indicator for assessing the performance of real estate assets and identifying potential growth or diversification opportunities. Case, Quigley, and Shiller (2013) note that the HPI provides a useful benchmark against which investors can compare their portfolio performance and make informed investment decisions.

In addition, the HPI is also important for policymakers, as it can help them identify and respond to changes in the housing market. For example, if the HPI shows that house prices are rising at a fast rate, policymakers may consider implementing measures to control the credit market, like raising interest rates. On the other hand, if the HPI indicates that house prices are falling, policymakers may consider introducing stimulus measures to support the housing sector and thus the economy.

I. Main indexes methodology

I.1 Mean or Median

The simplest analysis of housing price trends is based on measuring the central tendency of the distribution of prices of properties transacted in a particular period, using the mean or median. A significant disadvantage of simple indices (mean or median) is that they produce noisy estimates of price variations[10], since the set of homes sold in a period is often small and not necessarily representative of the stock of homes. Thus, changes in the composition of properties sold affect the sample of properties on which the mean or median is being calculated. Additionally, simple indexes are subject to biases when there are changes in the quality of properties over time or when certain types of homes are sold more than others.[11]

I.2 Stratification

The stratification method, also known as "composition adjustment," is based on the creation of strata or groups of transactions that aim to neutralize variations in the composition of properties sold. [12]

Stratification involves breaking down the total sample of transactions into a series of sub-samples or strata. Then, for each stratum, the price index based on the mean or median is obtained to calculate the aggregated housing price index as a weighted average of the indexes of each stratum. That is:

---

[10] Eurostat manual (2013). Pg 12.

[11] Prasad, N. L., y Richards A. (2008). Improving Median Housing Price Indexes Through Stratification.

[12] Moreover, stratification allows for the calculation of sub-indices for different segments of the housing market.

$$P^{0t} = \sum_{m=1}^{M} \omega_m^0 P_m^{0t}$$

Where P_m^{0t} is the index of the stratum m that compares the average price of the period t is the period with base 0, and ω_m^0 denotes the weight of the stratum m.

The formation of the strata is based on certain variables, such as geographic location, type of housing, housing condition, property size, etc. In this way, changes in the composition of housing between different periods are controlled, but not within each stratum. The definition of the stratification variables is key and should allow for the construction of the most homogeneous strata possible. It is important to note that if variables are not being considered in the stratification, changes in the composition according to those variables will not be controlled for.

Regarding the value of the weights, it depends on the type of index being estimated. If the main objective is to monitor the price variation of the stock of housing, then weights based on the existing stock of housing should be used, i.e., the value of the stock corresponding to each stratum. On the other hand, if the goal is to measure the variations in prices of the housing transacted in the market, weights based on sales should be applied.

The main advantages of the stratification method are that it is simple to implement and reproducible by any user who knows the stratification variables. It also allows for the construction of different sub-indexes within the housing market. However, this method assumes that the stratification variables are sufficient to control for all changes in composition.

If the stratification is too granular, the indexes calculated in each stratum can be very unstable due to the small sample size available in each cell. On the other hand, if the stratification is not granular enough, the composition variables will affect the indexes. Therefore, it is essential to achieve a balance in the number of variables used for stratification.

I.3 Repeat Sales

The name of this method indicates that it uses information about properties sold more than once. This way, it controls for the characteristics of the property, assuming that they do not change between periods and that price changes do not reflect changes in the quality or composition of the sample. However, implicitly, new properties for sale are excluded from an index constructed with this methodology.

To construct the house price index based on repeated sales, a simple regression model must be estimated that requires very few variables: the price and the sale date of the property. Thus, if there is a sample of properties sold in a period of time t=0,...,T, with their respective transaction prices, the linear regression model to be estimated is:

$$ln\left(\frac{p_n^t}{p_n^s}\right) = \sum_{t=0}^{T} \gamma^t D_n^t + \mu_n^t$$

Where p_n^t is the sale price of the property n in the period t; D_n^t is a dummy variable that take value 1 in the period of re-selling n, -1 in the period of the previous sale and 0 in any other case and μ_n^t is the error term.

Thus, the repeat sales index that goes from period 0 to period t is obtained as follows:

$$P_{VR}^{0t} = \exp(\hat{\gamma}^t)$$

While this method is simple and easy to implement, a problem with the repeat sales method is that the property sold at two different points in time is not necessarily identical due to factors such as depreciation and renovations. Eurostat (2013) recommends excluding properties that are sold quickly or that are not resold for prolonged periods from the calculations, that is, limiting the observation period for resales.

I.4 Hedonic Regression

The hedonic regression method assumes that the heterogeneity of goods can be represented through a series of observable attributes or characteristics. In the case of real estate properties, these attributes are related to the structure and location of the property. In this context, demand and supply for properties implicitly determine the marginal contributions of characteristics to property prices, and regression techniques can be used to estimate these marginal contributions.

The key assumption for the use of the hedonic regression method is that the price of property n at time t, P_n^t, is a function of a number K of characteristics that are measured through the variables z_{nk}^t:

$$P_n^t = f(z_{n1}^t, z_{n2}^t, \dots, z_{nK}^t, \varepsilon_n^t) \quad con \quad t = 1, \dots, T.$$

Where ε_n^t is a stochastic error term

To estimate the marginal contribution of each characteristic using the regression method, it is necessary to parameterize the function that relates the price of the property to its characteristics. The two most used hedonic specifications are:

- linear model:

$$P_n^t = \beta_0^t + \sum_{k=1}^{K} \beta_k^t z_{nk}^t + \varepsilon_n^t$$

- Log-linear model:

$$\ln P_n^t = \beta_0^t + \sum_{k=1}^{K} \beta_k^t z_{nk}^t + \varepsilon_n^t$$

In both specifications, it is allowed for the coefficients associated with the characteristics to vary over time, which is consistent with the idea that market conditions change in time;

however, these variations are gradual, allowing for the assumption of constant coefficients over time. Thus, the linear logarithmic model can be reduced to:

$$\ln P_n^t = \beta_0^t + \sum_{k=1}^{K} \overline{\beta_k} z_{nk}^t + \varepsilon_n^t$$

Therefore, if we have samples of transactions for each period, we could estimate the above model for each period separately, or combine all the periods and include dummy variables for those periods:

$$\ln P_n^t = \beta_0 + \sum_{t=1}^{T} \delta^t D_n^t + \sum_{k=1}^{K} \overline{\beta_k} z_{nk}^t + \varepsilon_n^t$$

An important aspect of hedonic models is that, like any regression model, all relevant characteristics of the property that determine the price should be included as explanatory variables. Otherwise, a bias in the housing price index is introduced due to omitted relevant variables.

Then, the coefficients of time-related dummy variables are used to obtain an indicator of the variation in housing prices, controlling for the characteristics[13]:

$$P^{0t} = \exp(\hat{\delta}^t)$$

II. Prices Indexes around the world

At an international level, both public and private organizations have constructed housing price indexes with the objective of providing useful information for research and decision-making. These indexes use different sources of information and methodologies, depending on the quality and availability of data. Below are some examples of Housing Price Indexes grouped according to the three methodologies mentioned in the previous section.[14]

Hedonic Regression

In Sweden, the SCB (Sweden Central Bank) uses an estimated value of properties based on hedonic regressions. They use a method that combines the use of transaction prices and appraisal prices. The properties are divided into 12 strata (according to price) based on estimated values (appraisals). The weights are calculated based on the stock of housing in each stratum, which allows for the calculation of the average sales prices in different strata and then the estimation of an index.

In addition, another index exists in Sweden: The Nasdaq OMX Valuegard-KTH, which, unlike the former, is calculated by a private company and is commercialized. It is based

[13] For further insight into other methods for obtaining the housing price index using hedonic models, see the Handbook on Residential Property Prices Indices (RPPIs), Eurostat (2013).
[14] Further information on Owusu-Ansah, A. (2018)

solely on actual transactions of houses and condominiums. This index is calculated for three metropolitan regions and uses a hedonic regression method with time variables.

In Germany, the Hypo Real Estate Index (HyperIndex) has been computed since 2008. Its main purpose is to provide useful information on the real estate market, especially because the German Banking Act stipulates that financed housing transactions must be constantly monitored. This index uses data from the German banking association on real transactions. There is an index for houses and apartments, a national index is published, and there is a specific index for some cities.

Stratification

The Australian Bureau of Statistics (ABS) publishes a bi-annual index of housing prices for the 8 capital cities in Australia, as well as a weighted national average. The index is constructed using the mean and is compiled using a mixed stratification based on two characteristics: the long-term price level for the suburb where the property is located, and the socioeconomic characteristics of the neighborhood.

Repeat Sales

In the United States, the most recognized index is the S&P/Case-Shiller, based on the repeat sales method. The index monitors the growth rate of residential properties at both the national level (published quarterly) and for 20 metropolitan regions (published monthly). The data for the construction of the index are managed by Fiserv Inc., a leading provider of information technologies that receives information on home sale prices from multiple sources and then cross-validates them before including them in the database.

In Canada, Teranet (an international leader in electronic property registration) along with the National Bank of Canada, produces a publicly available price index. This corresponds to an independent representation of the growth rate of individual house prices. The index covers eleven metropolitan areas of Canada. Table N°1 summarizes some international experiences in the development of housing price indices:

Table N°1. International Experience in House Pricing Index (HPI)

Country	Institution	Frequency	Data	Source	Methodology	Coverage
Mixed Adjustment						
Australia	Australian Bureau of Statistics	Quarterly	Sample	State and Territory Land Titles Office/Real Estate Agents	Stratified Median	National Index + 8 capital cities
Greece	Bank of Greece	Quarterly	All transactions	Banks and credit institutions/ Real Estate Agents	Geometric Mean Stratified	National Index + Categories[15]
Turkey	Central Bank of the Republic of Turkey (CBRT) Turkey Statistical Institute (TurkStat)	Quarterly	Mortgage transactions	Commercial Banks and Real Estate appraisal companies	Stratified Median	National Index + Capital (all dwellings)
Hedonic						
Sweden	Sweden Central Bank	Quarterly	Real transactions of purchases of new dwellings and Purchases of existing dwellings	Real Estate Tax Assessment Register	Mixed Method: Hedonic Regressions + SPAR	National Index
Germany	Hypo Real Estate Holding	Quarterly	Transaction prices both cash and mortgage for new dwellings and existing dwellings	Association of German Pfandbrief Banks (vdp): Transactions of 13 financial institutions Expert Committees for Property Valuation GEWOS	Hedonic Regressions	National Index
United Kingdom[16]	The Nationwide Building Society	Monthly	Approved mortgage transactions	Nationwide: Mortgage Data	Hedonic Regressions (strictly cross-sectional hedonic model SCS)	National Index + 13 regions in the UK

[15] Capital city, big cities, small cities, urban area and rural areas.
[16] Wood, R. (2005), A Comparison of UK Residential House Price Indices.

United Kingdom	The Halifax	Monthly	Approved mortgage transactions	Halifax: Mortgage Data	Hedonic Regressions (strictly cross-sectional hedonic model SCS)	National Index + 12 regions in the UK
Ireland	Central Statistics Office (CSO)	Monthly	All transactions	Irish Tax Office	Hedonic Regressions	National Index + Capital (all dwellings)
Finland	Official Statistics of Finland (OSF)	Monthly	All transactions	Finnish Tax Administration's asset transfer tax data	Hedonic Regressions	National Index + categories[17]
France	National Insitute of Statistics	Quarterly	Sample	No direct sources	Weighted average indices adjusted with hedonic methods	National Index + Capital City + Suburbs (all dwellings)
Norway	Statistics Norway	Quarterly	All transactions	FINN.no (website) Real Estate Norway, covering transactions of existing dwellings	Hedonic Regressions	National Index + Capital City + Suburbs (all dwellings)
Poland	National bank of Poland	Quarterly	Sample	380 district governor's offices (local administration units)	Hedonic Regressions	National Index + categories[18]
Spain	National Institute of Statistics	Quarterly	All transactions	Notarial Certification Agency (ANCERT).	Mixed Method: Combine hedonic regressions with stratification	National Index
Repeated Sales						
United States	S&P/ Case-Schiller National Home Price Index	Monthly	Used dwellings transactions	Fiserv Inc[19]	Repeated Sales	National Index + 20 metropolitan regions
Colombia	Central Bank of Colombia	Quarterly	Used dwellings transactions with mortgage	Information on mortgage loans reported by various financial institutions	Repeated Sales	National Index
Canadá	National Bank of Canada / Teranet	Monthly	Used dwellings transactions	Teranet (electronic property registration system)	Repeated Sales	National Index + 11 metropolitan regions

[17] Capital city, big cities, urban area and rural areas.
[18] Capital city, big cities, small cities, urban area and rural areas.
[19] Provider that collects data from multiple sources and then cross-checks them.

United Kingdom	Office for National Statistics	Monthly	Used dwellings transactions	HM Land Registry (Real Transactions)	Repeated Sales	National Index (just for England and Wales)

IPV Methodology

In 2014, the Central Bank of Chile began calculating the IPV based on administrative data from SII, corresponding to actual housing transactions nationwide registered through form F2890, complemented with information from the Real Estate Registry (CBR). In 2014, the method of stratification was defined with 14 strata, constructed through the variables of geographic zone (seven zones) and type of housing (houses and apartments). The calculation requires a series of depurations and imputations that allow for comparable information between strata over time. In issue 107 of the series of Statistical Economic Studies of the Central Bank of Chile, published in June 2014, the methodology and main results of the IPV are presented.

Later, in 2019, the data imputation process and methodology were improved[20] to expand it to 27 strata defined by the variables of geographic zone (seven zones), type of housing (houses and apartments), and condition of the housing (new and used). [21] Finally in 2021, we generate a new set of improvements in the depuration methodology which are explain in the document 139 of the series of Statistical Economic Studies of the Central Bank of Chile, published in July 2023.

I. Database

The IPV uses two sources of information for its calculation, both corresponding to quarterly data provided by the SII:

1- Declaration Form for Real Estate Transfer and Registration (F2890)

Quarterly, the SII provides the BCCh with a database of transactions recorded in the F2890. Additionally, since 2015, the SII has made available to the Central Bank the information from the online F2890 form, which is processed online in the main notaries of the country. Having the online F2890 form reduces the lag time for publishing the IPV, as this information is provided monthly and the lag time for registering sales is less than two months. [22]

2- National Registry of Habitational and Agricultural Real Estate (CBR)

This information is used to complement the information available through the F2890, mainly to impute the variables target code (variable that identifies properties for residential use), built square meters, land square meters and address information (that allows for better classification of houses and apartments) when this information is not available.

[20] Press Release Central Bank of Chile (2019). "Actualización metodológica del Índice de Precios de Vivienda que elabora el Banco Central de Chile".
[21] In the City Center of Santiago, there are no transactions for new houses, so this stratum is removed, resulting in 27 strata instead of 28.
[22] However, the online F2890 is preliminary information, subject to corrections, and does not include all the variables of the regular F2890, and more variables may have missing values.

II. IPV Calculation

The first stage of calculating the IPV involves a series of processes to consolidate the different sources of information and clean the data from these administrative records[23].

These stages can be summarized by the following set of processes:

1- Compilation and formatting of databases: joining the databases, modifying the variables to the required formats, and creating missing variables in the process
2- Removal of duplicates: first, general observations that are identical in all variables, and then a precise specification (ID, date, amount) that refer to the same property in the same period
3- Correction of data entry errors: review and resolution of problems related to data entry
4- Detection of social housing: Elimination of social properties
5- Classification of property type: identification of transactions between houses and apartments
6- Imputation with historical information: destination code variables, built square meters, appraisal, land, and date of registration.
7- Identification of warehouses
8- Imputation of construction

After the data cleaning process, the IPV is calculated. This process can be summarized in the following steps:

1- For each of the 27 strata and for each quarter, the simple average of transaction prices in Unidades de Fomento (currency indexed to inflation) per square meter (UF/m^2) is calculated.
2- A weighted average price is calculated for each quarter. The weights are annual and obtained by dividing the sum of square meters transacted in the stratum during the previous year by the total square meters transacted in the previous year.
3- A chained quarterly price is calculated using prices from the previous year.
4- The house price index is calculated using as a base the chained average price from 2008.

As mentioned in earlier sections, besides the 3-month lag in receiving the information, the IPV stabilizes after two quarters. This is mainly because the registration of new housing units takes longer due to the longer period between sale and delivery of the property. Therefore, the IPV in its last two quarters is considered provisional. For more detail, review the methodological document[24].

[23] The F2890 form is manually entered, which means that this information is not free from errors.

[24] Balsa, J y Vásquez, J (2023). Índice de Precios de Viviendas Banco Central de Chile 2022. Estudios Económicos Estadísticos N°139.

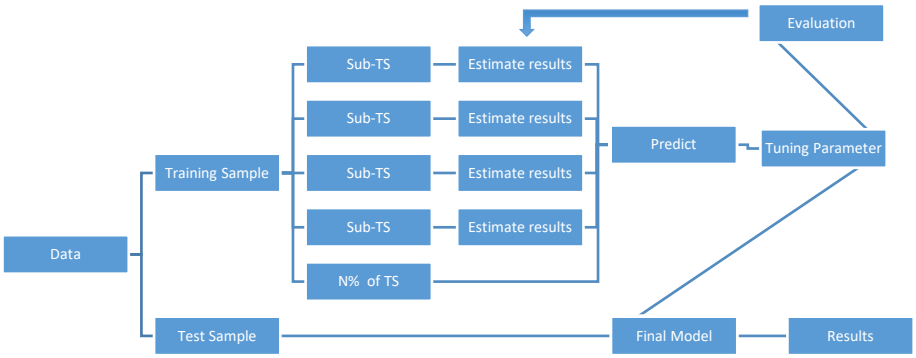
Classification Algorithm

I. What is a Classification algorithm and how it works

Classification algorithms are used to classify data into different classes based on specific given criteria. According to Alpaydin (2010) and Kotsiantis (2007), these algorithms analyze incoming data, learn the patterns and relationships between the various features of the data, and make predictions about the class to which the data belongs. The main goal is to develop a model that manages to predict the class of new data based on its features. The classification process encompasses data preprocessing, feature selection, model training (if need it), evaluation, and finally deployment.

For this paper, we will focus on supervised classification algorithms. As explain in Balsa (2018), Figure I resume the process of a supervised algorithm: first it splits the sample into a training sample and a test sample. Then, for the training sample, the algorithm will create subsamples and leave a n percent apart. After that, the algorithm will start to estimate results from the subsamples and use them to predict the outcomes of the n percent that it was left aside. Later, the tuning parameter is chosen, based on the one that minimizes the loss/cost function, that is normally defined as the sum of the squared residuals, the entropy or other measures in the cross-validation samples. The final model performance is assessed by calculating the lost/cost function of model predictions on the held-out test sample, which was not used at all for model estimation or tuning.

Figure I: Supervised ML Algorithm



There are several types of supervised classification algorithms, including decision trees, kneighbors, random forests, logistic regression, neural networks, among others. Hastie, Tibshirani, and Friedman (2009) and Bishop (2006) note that these different algorithms work in different ways, such as logistic regression that uses a logistic function to model binary dependent variables, decision trees that make decisions by partitioning the data based on feature values, and neural networks that use interconnected nodes to process information and make predictions.

As expected, each algorithm has its own strengths and weaknesses. For example, a logistic regression is easy to interpret, and can handle multi-class classification problems, according

to Kotsiantis (2007). However, it may not work appropriately with non-linear boundaries and can also have high bias. Decision trees, as described by Alpaydin (2010), are easy to interpret and visualize capturing non-linear relations between features. Nevertheless, they can be prone to overfitting, and it is very sensitive to small changes in the data. Lastly, neural networks, as highlighted by Hastie, Tibshirani, and Friedman (2009), are highly flexible and can learn complex decision boundaries, making them very effective for all kind of classification problems, but they can be difficult to interpret, and they consume significant computational resources, making them slower than other algorithms.

One popular classification algorithm is the k-nearest neighbors (k-NN) algorithm, which is a type of instance-based learning. According to Alpaydin (2010) and Bishop (2006), the k-NN algorithm assigns new data points to the class of most of its k-nearest neighbors in the training data set. The value of k is typically chosen based on a validation set, and smaller values of k can lead to higher variance and overfitting, while larger values can lead to higher bias.

The k-NN algorithm has several strengths and weaknesses. As highlighted by Alpaydin (2010), the algorithm is simple and easy to understand, and can be effective for low-dimensional data sets. It can also handle multi-class classification problems and non-linear decision boundaries. However, the algorithm can be computationally intensive and may require significant memory storage, especially for large data sets. In addition, it can be sensitive to outliers and noise in the data.

Overall, the k-NN algorithm is a useful classification algorithm for low-dimensional data sets and can be effective for multi-class classification problems. However, it may not be the best choice for large data sets or when computational resources are limited.

These sources provide a comprehensive overview of classification algorithms and their role in machine learning. They highlight the different types of classification algorithms, how they work, and their strengths and weaknesses. The information provided is useful for anyone interested in understanding the classification process in machine learning.

II. Examples of Property Classification with administrative data

As far as we are concerned, administrative data can be a valuable source for property classification. In this section, we present some examples of countries that use administrative data to classify properties, as well as the methodologies they use.

1. United Kingdom:

The Valuation Office Agency (VOA) uses administrative data from local authorities and the Land Registry for property classification. This data can include information about property location, size, type, age, and architectural characteristics. Although specifics about the use of machine learning are not publicly available, the agency has mentioned in general terms the use of machine learning algorithms for property valuation.

2. Canada:

The Municipal Property Assessment Corporation (MPAC) uses data from property sales, physical inspections, building permits, and other local authority data, using variables that include property size, location, age, type and data from physical inspections. They often use a blend of statistical modeling and machine learning for their methodologies. For example, they use multiple regression analysis to assess property values, considering different variables such as property size, location, among others. As for property classification tasks, such as distinguishing residential properties from commercial ones, they often use decision tree algorithms.

3. Netherlands:

The Dutch Land Registry (Kadaster) and Statistics Netherlands (CBS) use administrative data for property classification. The use a range of this data including property transaction data, physical property attributes. To do this, these organizations use different automated valuation models (AVMs) which could include decision tree algorithms, random forests, or neural networks, trained on historical transaction data.

Overall, administrative data can be useful to get a comprehensive view of the property market, as it includes information on various characteristics of the properties, such as location, size, type, and condition. While there are some differences in the specific variables considered and the algorithms employed, the main goal for organizations is to provide an accurate and fair property valuation. By using administrative data, governments and organizations can create models that assign a value to each property based on its unique characteristics. In addition, these models can help to identify trends and patterns in the housing market over time.

IPV Classification Methodology

As explained earlier, the IPV is obtained using the mixed stratification method, for which three variables are used: property type (house or apartment), condition (new or used), and geographic zone.

The property type variable does not exist in the F2890 database or in the auxiliary databases of the process. However, it is constructed based on the available variables in these sources through an iterative process. Two strategies were followed: one called "Expert Criteria" that uses logical criteria from the researcher to generate the classification, and another strategy called "Data-driven" that uses machine learning algorithms to classify the properties.

Although one might think that it is easy to classify properties using only a criterion based on the square meters of land, where all properties with zero land should be classified as apartments, this is no true. For example, in the case of Type A houses in condominiums (i.e., houses located on common land), even though they are registered with zero land they are houses and have shared land with other houses in the condominium.

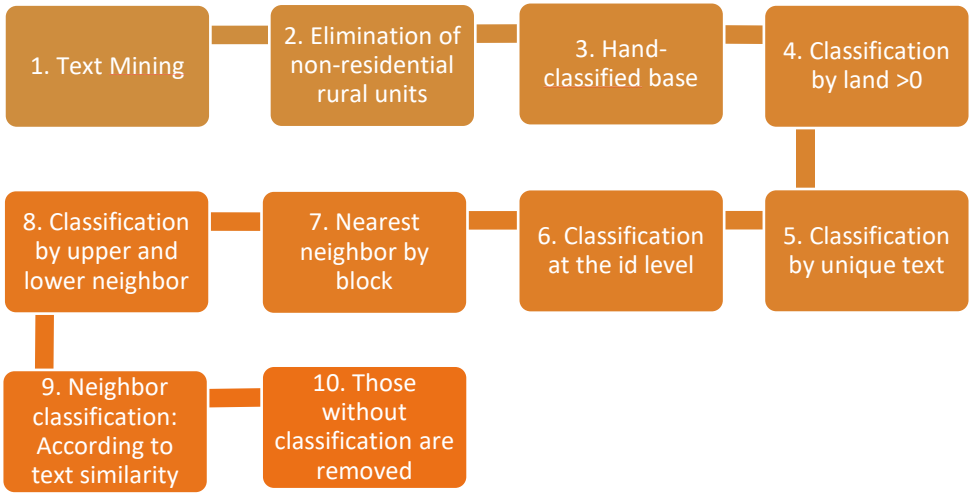
In this section we explain both criteria and the results and uses of the data generated from the process.

I. Flow 1: Expert Criteria

The "expert-criteria" classification process of the IPV works through 10 stages, which are summarized in Figure II:

1. *Text-Mining*: We review the text variables related to the address and look for characteristics that indicate whether they are houses or apartments, as well as whether the properties are rural (plots, land, etc.).
2. *Elimination of non-residential transactions (rural units)*: All properties that were classified as rural in the previous stage are deleted. (3.3% transactions eliminated).
3. *Hand-classified base:* A historical database is created with different units and it is checked with external databases, having around 40.000 units. (It is in continuous growth). (1.27% classified).
4. *Classification by land over zero*: if the Land information is over zero, we assume that it is a house. (58.2% classified)
5. *Classification by unique text:* if was only classified in one type we keep that one. For example, if the address only has information that says is an apartment, we keep it as an apartment. (36.6% classified)
6. *Check classification at the id level:* if all the transactions of a unit have the same classification, we maintain it if not we de-classified the unit. (0.2% classified)
7. *Nearest neighbor by block*: If all the units in a block had the same classification, we complete the ones without classification by the one in the unit. (0.01% classified)
8. *Classification by upper and lower neighbor*: As the name says if my neighbor unit upper and lower has the same classification, we imputed that classification to the unclassified unit. (2.6% classified)
9. *Neighbor classification according to text similarity*: We compare with soundex and levenshtein how similar are two addresses are and we classify by it. (0.1%)
10. *Elimination of non-classify observations*: We eliminated all observations that are not classify. (less than 0.001%)

Figure II: Expert Criteria Flow Resume



The "Expert Criteria" classification process, although works well in correctly classifying around 3 million properties, had errors when classifying new properties. This is because new properties have less information available regarding their addresses, and their land information may not be up to date. For example, Images 1 and 2 show two cases where the "Expert Criteria" misclassified properties as houses when they are apartments.

Image 1: Misclassification in the "Expert Criteria"



Image 2: Misclassification in the "Expert Criteria"



These errors, although they may seem to involve a few units, have a significant impact on the IPV. Since it is a stratified index with 27 zones and a quarterly frequency, the inclusion of hundreds of apartment units (which have a lower price per square meter compared to houses) can result in variations of over 2% in the annual variation of the index in a specific stratum and even up to 1% in the overall index.

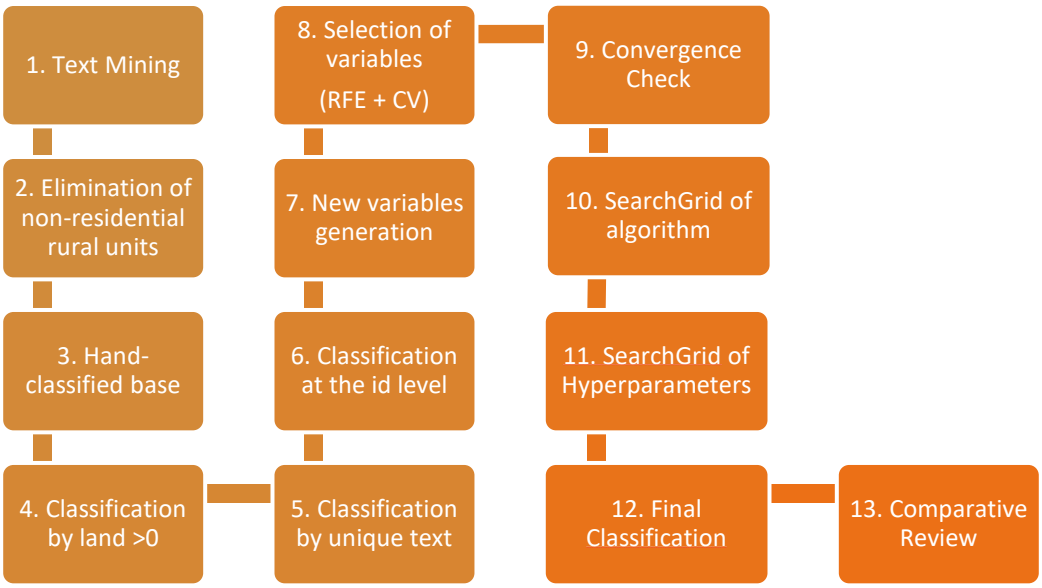
II. Flow 2: Data-Driven classification

Given the errors of the "Expert Criteria", it was decided to experiment with new automated classification methods, which led to a supervised machine learning algorithm.

Like the previous criterion, the first 6 steps were maintained, which were used as the training database for the algorithm. This database consists of approximately 3 million transactions. The selection of the expert criterion cutoff at this step is based on the absence of classification errors up to that point and the consideration of "solid" classification criteria in a complementary manner.

The "Data-Driven" method is explained below (Figure III) with some examples of the code used:

Figure III: Data-Driven Flow Resume



7. *New variables generation:* The database is linked to other sources such as the census and variables are created by block and municipality related to the percentage of each type of property, credit ratio, number of units, among others.
8. *Selection of variables*: We use a recursive feature elimination method with a Random Forest algorithm and five splits of the database for cross-validation. The method chooses ten variables: *Construction*, *Land*, *Credit Ratio*, *Price Ratio*, *Construction Ratio*, *Number of units per block*, *Two measures of quality*, *percentage of apartment by the last census*, *logarithm of the public valuation of the property*.

```
#%% 1.8♦. Feature Selection

"""
Recursive Feature Elimination with cross validation
"""
min_features_to_select = 1  # Minimum number of features to consider
rf = se.RandomForestClassifier(random_state=123)

rfecv = RFECV(
    estimator=rf,
    step=1,
    cv=cv,
    scoring="accuracy",
    min_features_to_select=min_features_to_select,
    n_jobs=25,
)
rfecv.fit(x, y)

print(f"Optimal number of features: {rfecv.n_features_}")

# get the selected features
selected_features = np.array(x.columns)[rfecv.support_]
print("Selected features: ", selected_features)
```

9. *Convergence Check*: One important step is to verify if all the chosen algorithms converge to results. Without this preliminary step, the code could run indefinitely without reaching a conclusive outcome. SVM was eliminated for non-convergence.

```
#%% 1.9. Model Selection
"""
Convergence Check
"""
rf   =   se.RandomForestClassifier(n_estimators=100,   max_depth=5,   n_jobs=6,
random_state=123)
rf.fit(x_train, y_train)
rf_pred=rf.predict(x_test)

svm = SVC(C=1, kernel="rbf", gamma="scale",  random_state=123)
svm.fit(x_train, y_train)
svm_pred=svm.predict(x_test)

lr = LogisticRegression(C=1, solver="lbfgs",  random_state=123)
lr.fit(x_train, y_train)
lr_pred=lr.predict(x_test)

knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(x_train, y_train)
knn_pred=knn.predict(x_test)
```

10. *Searchgrid of algorithms*: We use the search grid technique to systematically explore different combinations of algorithms and hyperparameters to find the best model performance. This helps automate the process of finding optimal hyperparameter settings.

```
#%% 1.9. Model Selection
"""
Model SearchGrid
"""
x= x[Ganadoras]

classifiers = [
    se.RandomForestClassifier(n_jobs=7,random_state=123),
    LogisticRegression(n_jobs=7),
    KNeighborsClassifier(n_jobs=7),
    ]
param_grids = [
    {
        "n_estimators": [100,200,300],
        "max_depth": [3,5,7],
     },
    {
     "C": [0.1,1,10],
     "penalty": ["l1", "l2"],
     "solver":["lbfgs","liblinear"],
     },
    {
     "n_neighbors": [3,5,7],
     "weights": ["uniform", "distance"],
     "p":[1,2],
     },
    ]

best_score=0
best_params= None
best_model= None
```

```
counter=0

for classifier, param_grid in zip(classifiers, param_grids):
    print(fr"starting {counter} of 255")
    grid_search= GridSearchCV(classifier, param_grid, cv=5, n_jobs=7)
    grid_search.fit(x_train, y_train)

    if grid_search.best_score_ > best_score:
        best_score= grid_search.best_score_
        best_params = grid_search.best_params_
        best_model = grid_search.best_estimator_
    counter=counter+1

test_score=best_model.score(x_test, y_test)
print("best Score:", best_score)
print("best Parameters:", best_params)
print("Best model:", best_model)
print("Test Score:", test_score)
```

11. *Searchgrid of Hyperparameters*: We use the same technique to choose what hyperparameters give the best model performance for the algorithm that won. [25]

```
#%% 1.9. Model Selection
"""
Hyperparameter searchgrid
"""
classifiers = [
    KNeighborsClassifier(n_jobs=10),
    ]
param_grids = [
    {
     "n_neighbors": list(range(1,15)),
     "weights": ["uniform", "distance"],
     "p":[1,2],
     "leaf_size": list(range(1,30)),
     },
    ]

best_score2=0
best_params2= None
best_model2= None

cv = KFold(n_splits=5)

for classifier, param_grid in zip(classifiers, param_grids):
    grid_search=   GridSearchCV(classifier,   param_grid,   cv=cv,   n_jobs=3,
pre_dispatch='n_jobs')
    grid_search= grid_search.fit(x_train, y_train)


    if grid_search.best_score_ > best_score2:
        best_score2= grid_search.best_score_
        best_params2 = grid_search.best_params_
        best_model2 = grid_search.best_estimator_

test_score=best_model2.score(x_test, y_test)
```

[25] Care must be taken with these steps since the excessive search of hyperparameters in very small databases can lead to overfitting. Therefore, all processes are run with a cross-validation process to avoid sample bias.

12. *Final Classification*: The transactions are classified with the selected algorithm and hyperparameters.
13. *Comparative Review*: We compared the results of Flow 1 with Flow 2 and manually reviewed the differences.

All these previous processes were run in Python 3.7 using the modules of Sci-kit learn, numpy, pandas, joblib and os. The database had 34 variables and a training set of 3M transactions with 21 variables pre-selected with expert criteria. The algorithm selects 10 variables: *Construction*, *Land*, *Credit ratio*, *Price ratio*, *Construction ratio*, *N° of unit per block*, *Quality 1 and 2*, *percentage of apartment by Census*, *Log of public valuation* and runs with a cross-validation of 5 splits.

We compare the results of 4 algorithms: *Random Forest*, *Support Vector Machine*, *Logit*, *Kneighbors*, and we choose the algorithm with best accuracy. The winner was Kneighbors with the following hyperparameters:

Neighbors = 13 | weights = 'distance' | p = 1 | leaf size = 15

The algorithm predicts with an accuracy of 99.2% in-sample and 98.9% out-of-sample. When comparing the classification of flow 1 with flow 2, we can see that flow 2 correctly classifies the buildings solving the previous problems, but it does have some issues with new houses. That's why we decided to use both processes and compare them.

It's also important to note that the process improves with each iteration through the hand-classified database. The first time we ran it, the difference between Flow 1 and Flow 2 was 250,000 transactions, and it has been decreasing to 7,000 transactions.

This process is also re-tunable and is recommended as market trends change or updates to external data sources are incorporated.

III. Bonus Track: ML Algorithm for any classification problem

During the creation of this classification process a change of emphasis was made from pure property classification to an easy-to-use code for any classification process of the Central Bank of Chile.

This process can be used for both dichotomous and multicategory problems, as well as focusing on specific problems such as unbalanced samples, all through the choice of algorithms to incorporate in the searchgrid or limiting the searchgrid within a subset of hyperparameters. For example, if the problem is highly unbalanced, one can choose Kernel-Support Vector Machine or Logit with an L1 regularization.

In order to be able to modify the code, it is only necessary to have a pre-classified database to train the algorithms. However, it is also advisable to have some criteria for the pre-selection of variables.

IV. Results and uses of the data

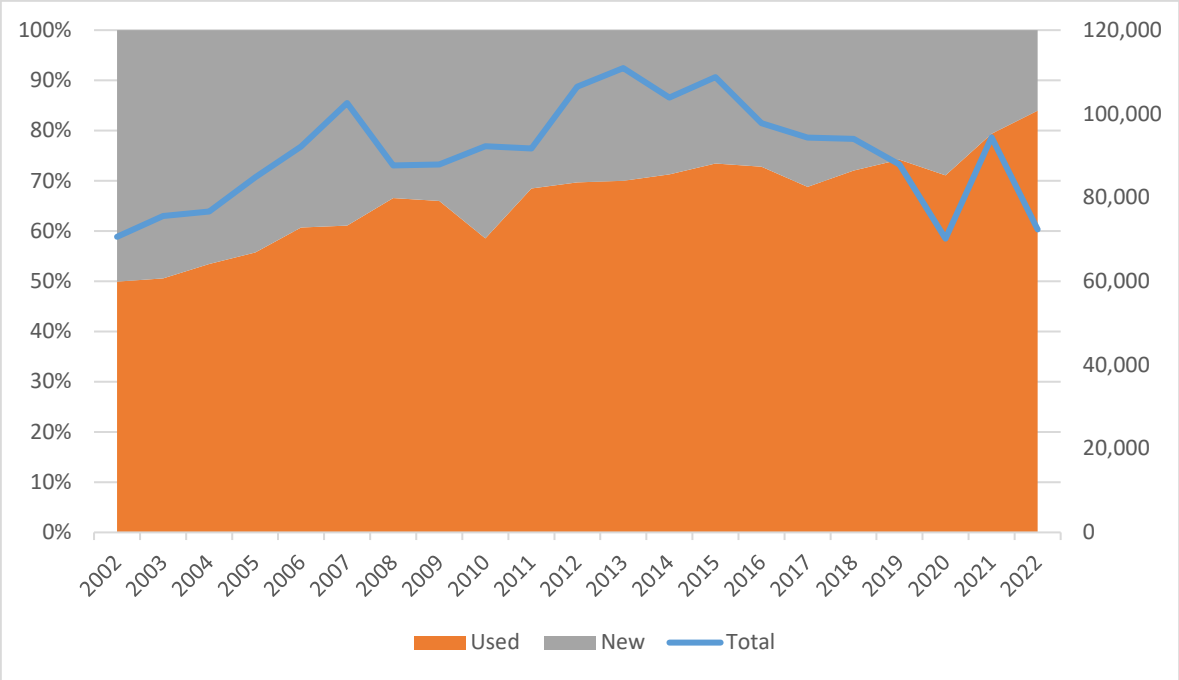
The purpose of this section is to show how the Central Bank uses the data classified by the algorithm. The final consolidated and cleaned database contains data from the first quarter of 2002 through the fourth quarter of 2022 and has approximately 3.2 million residential property transactions nationwide. Of these transactions, 47.8% correspond to properties in the Metropolitan area, 61.2% to houses and 42.4% to new properties. The consolidated and cleaned database has an average price of 27.4 UF/m2 and a standard deviation of 19.33 UF/m2, the maximum price value in UF/m2 is 200.3 UF/m2.

The number of annual transactions increased by 35% between 2002 and 2022[26] (76% in 2021), from 94 thousand to over 125 thousand. However, since the peak in 2015 (186 thousand), the number of transactions has been declining reaching even less than 130 thousand by 2022. The market for houses (70 thousand) in 2002 more than tripled the number of apartments (24 thousand). This dynamic has been adjusting over the years due to the increase in urbanization and gentrification of cities, with the housing market (72 thousand) being only 30% higher than the apartment market (56 thousand) by 2022.

On the other hand, the apartment market has grown substantially between 2002 and 2022, going from 24 thousand to 56 thousand units annually, finding its peak in 2019 with 82 thousand transactions. The most significant growth is found in new apartments, which from 11 thousand units in 2002 reach 28 thousand in 2021, almost 2.6 times more than in 2002 and reaching up to 4.5 times more in 2017 (50 thousand). Transactions of used apartments increase from 13 thousand to 27 thousand in this same period, with 2021 being the year with the highest number of transactions (38 thousand).[27]

[26] It is important to note that the 2022 base does not yet contain all transactions and could increase by up to 10% based on analyses performed for other years.
[27] Graph N°2.

Graph N°1. Number of transactions per year and composition between new and used (Houses)



Graph N°2. Number of transactions per year and composition between new and used (Apartments)

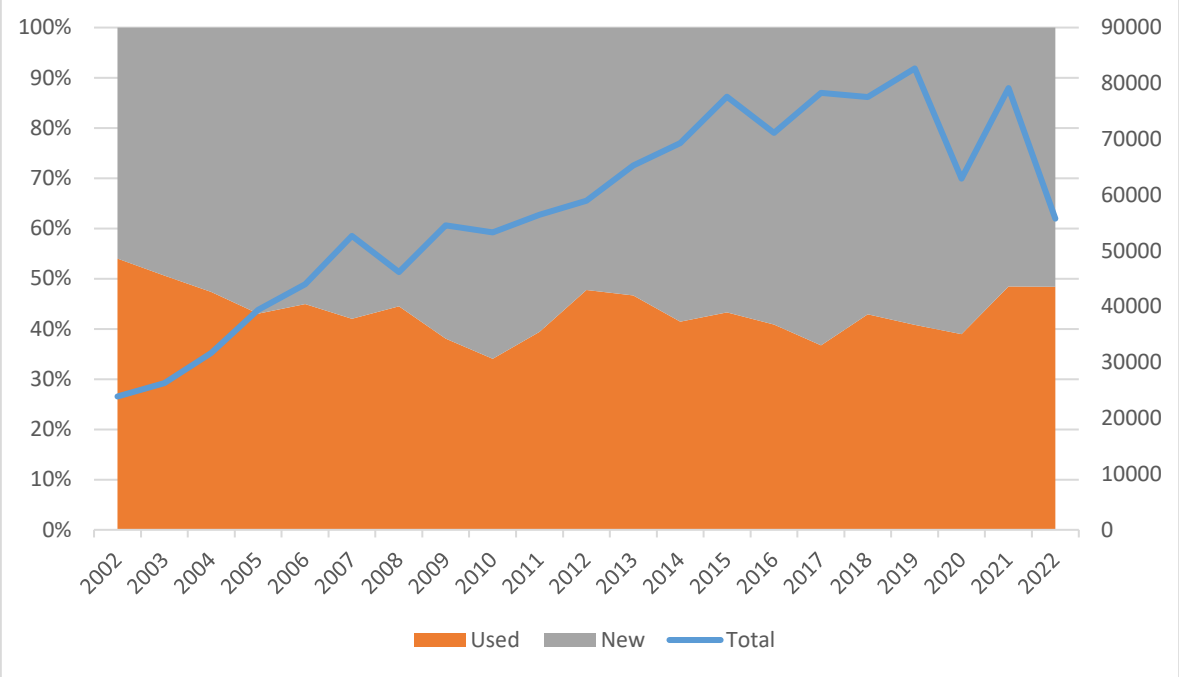


Table N°2 shows the number of transactions during 2022 for each of the 27 strata. In total there were 128 thousand transactions, which are broken down between 88 thousand used

and 40 thousand new, of which 55 thousand corresponded to apartments and 72 thousand to houses. In the Metropolitan Region, the east and west zones registered the highest number of transactions, with 15,803 and 15,348 respectively, with apartments being the most traded in the east zone and houses the most traded in the west zone of the capital. At the national level, the central zone registered some 37 thousand transactions, dominated by used houses (19 thousand).

Table N°2. Number of transactions per stratum (2022)

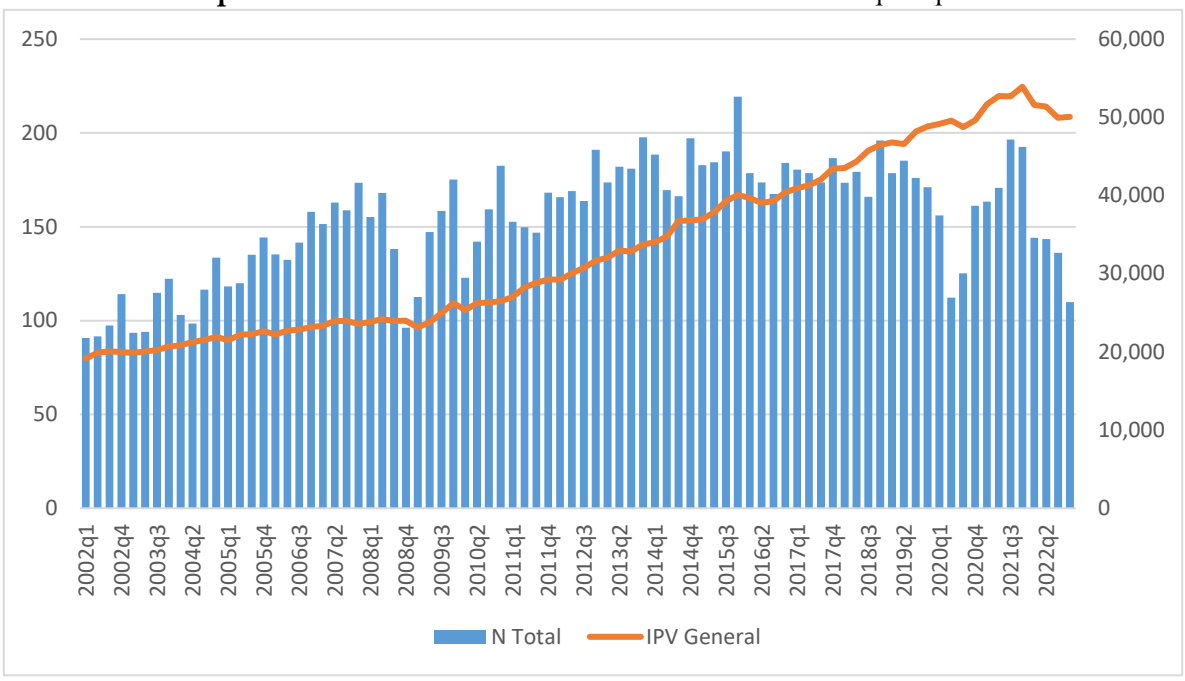
	Houses			Apartments		
Zone	Used	New	Total	Used	New	Total
MR Central	630	0	630	3729	3317	7046
MR East	3935	205	4140	6739	4924	11663
MR West	8587	1777	10364	2541	2443	4984
MR South	7002	1686	8688	2199	3219	5418
Total MR	**20,154**	**3,668**	**23,822**	**15,208**	**13,903**	**29,111**
North Zone	4774	314	5088	2027	1392	3419
Central Zone	18958	4016	22974	6826	7304	14130
South Zone	16878	3594	20472	2909	6171	9080
Total	**60,764**	**11,592**	**72,356**	**26,970**	**28,770**	**55,740**

Graph N°3 and N°4 show the general dynamics between the IPV and the number of transactions. With them, the Central Bank can study the dynamics of this indicator at different junctures:

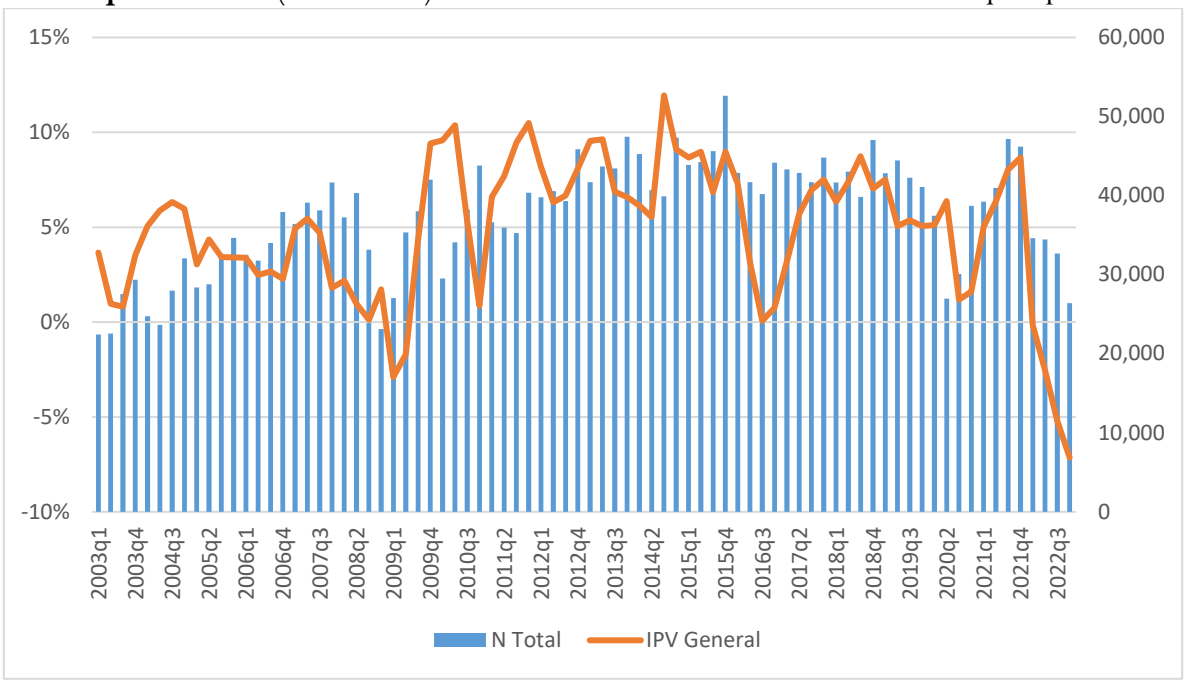
1. Housing prices have risen 161% nationwide from the first quarter of 2002 to the fourth quarter of 2022, with the western part of the metropolitan region being the area with the highest increase (245.3%). In the case of house condition, new homes (175%) are leading the increases compared to used homes (156.8%).
2. In the first quarter of 2009, housing prices had one of their largest declines nationwide (-2.9%), caused mainly by the effects of the subprime crisis. This drop was driven by used homes (-8.6%), particularly houses (-10%), while new homes slowed, but did not enter to negative territory (2.2%).
3. The fall in prices at the end of 2010 is consistent with the effects of the earthquake, which, unlike in the subprime crisis, it was new homes that led the falls (-4%).
4. The increase in prices in 2014 that was linked to a drop in sales of construction companies explained by the economic conditions of the time (11.9%).
5. The confluence of factors derived from the social outbreak and the pandemic, evidenced by an increase in credit restrictions and the unemployment rate, as well as the decrease in real remunerations, have caused a notorious repercussion in the number of transactions and in the decrease of the price by 11.9% as of the fourth quarter of 2022. This drop is the largest recorded since the existence of the index.
6. Finally, and linked to the previous point, there is a decoupling in the trend between new and used homes, where new homes have remained in positive values during 2022, unlike used homes that have entered negative territory, with the MR (-11.9%)

and especially the eastern zone (-12.6%) contributing the most to the fall in the index.

Graph N°3. General IPV and Number of transactions per quarter



Graph N°4. Var (12 months) General IPV and Number of transactions per quarter



Conclusion

The global housing indexes are of great relevance for the decision making of various agents of the economy, which has increased since the subprime crisis and in the face of the recent pandemic. Their interconnection with the financial system has led the Central Bank of Chile to pay particular attention to them in recent years.

However, the creation of these indicators, especially if they use administrative data, is not without challenges. This paper presented a solution to the handling of this data for the calculation of the Housing Price Index (IPV), managing to classify between house and apartment 3.2 million transactions automatically, with an error rate close to zero and thus reducing substantially the working hours involved (from 5 working days to just one morning).

Interestingly, the process is available not only for property classification, but also as a first approach to any classification problem, whether it is dichotomously balanced, multi-categorical or unbalanced.

Among the main results we found that the IPV constructed with this methodology has had the theoretically expected behavior in the face of the shocks experienced by the Chilean economy and especially by the construction industry.

Particularly, the index correctly reflects the price dynamics caused by the social outbreak and the COVID-19 pandemic, showing a strong fall in housing prices (-11.9%), which is congruent with the greater credit restrictions, the increase in unemployment, the fall in real wages and the deceleration in the construction sector.

It would be interesting to review the results of the methodology in the face of other classification problems, looking at its consistency and effectiveness. It would also be interesting to analyze how much training data is needed for it to work properly.

References

- Alpaydin, E. (2010). Introduction to machine learning (2nd ed.). Cambridge, MA: MIT Press.
- Balsa, J y Vásquez, J (2023). Índice de Precios de Viviendas Banco Central de Chile 2022. Estudios Económicos Estadísticos N°139.
- Balsa, J. (2018). Using causal tree algorithms with difference in difference methodology: a way to have causal inference in machine learning. Available in https://repositorio.uchile.cl/handle/2250/168527
- Central Bank of Chile (2014). Índice de Precios de Viviendas en Chile: Metodología y Resultados. Estudios Económicos Estadísticos N°107.
- Bishop, C. M. (2006). Pattern recognition and machine learning (1st ed.). New York, NY: Springer.
- Case, K. E., Quigley, J. M., & Shiller, R. J. (2013). Wealth effects revisited 1975-2012. Cowles Foundation Discussion Paper No. 1933.
- Englund, P., Hwang, M., & Quigley, J. M. (2002). Hedging housing risk. Journal of Housing Economics, 11(4), 310-342
- Eurostat (2013). Handbook on Residential Property Prices Indices (RPPIs). Eurostat Methodologies & Working Papers.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). New York, NY: Springer.
- Himmelberg, C., Mayer, C., & Sinai, T. (2005). Assessing high house prices: Bubbles, fundamentals, and misperceptions. Journal of Economic Perspectives, 19(4), 67-92.
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. Informatica, 31(3), 249-268.
- Owusu-Ansah, A. (2018). Construction and Application of Property Price Indices. Routledge Studies in International Real State.
- Prasad, N. L., y Richards A. (2008). Improving Median Housing Price Indexes Through Stratification. Journal of Real Estate Research, vol. 30, No. 1, págs. 45–71.
- Press Release Central Bank of Chile (2019). "Actualización metodológica del Índice de Precios de Vivienda que elabora el Banco Central de Chile": https://www.bcentral.cl/c/document_library/get_file?uuid=e8d65791-b51b-43a3-3e6b-94fbb52732de&groupId=33528
- S&P Dow Jones Indices. (2021). S&P/Case-Shiller U.S. National Home Price Index. Retrieved from https://www.spglobal.com/spdji/en/indices/real-estate/sp-case-shiller-us-national-home-price-index/

- Wood, R. (2005), A Comparison of UK Residential House Price Indices. Real Estate Indicators and Financial Stability, BIS Papers No 21, Banco de Pagos Internacionales.

Property classification with administrative data: The case of the Chilean House Price Index

Department of Microdata

Statistics Division , July 2023

The Challenge



The Chilean House Price Index (IPV)

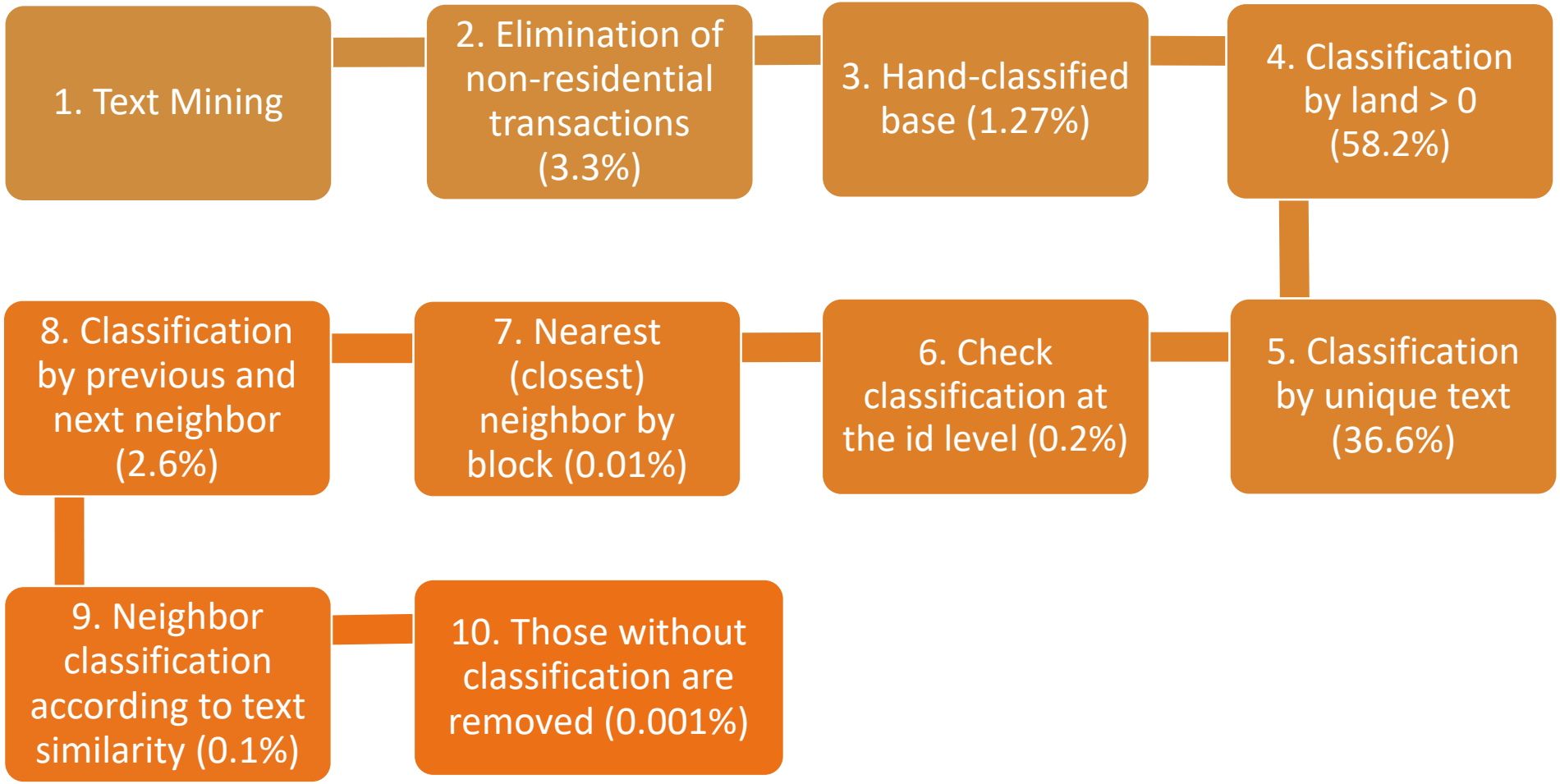
- Since 2014, the Central Bank of Chile calculates the House price Index based on administrative data from the Tax Administrative office, corresponding to actual housing transactions nationwide registered, complemented with information from the Real Estate Registry.
- The IPV is an experimental statistics and it's created based on a stratification method with 27 strata defined by:
  - Geographic zones (7)
  - Type of Housing (houses and apartments)
  - Condition of the property (new and use
- Of course, the distinction between houses and flats (apartments) is not part of the information reported by the tax authorities, so it has to be classified within the IPV calculation process.

banco central Chile

Solution

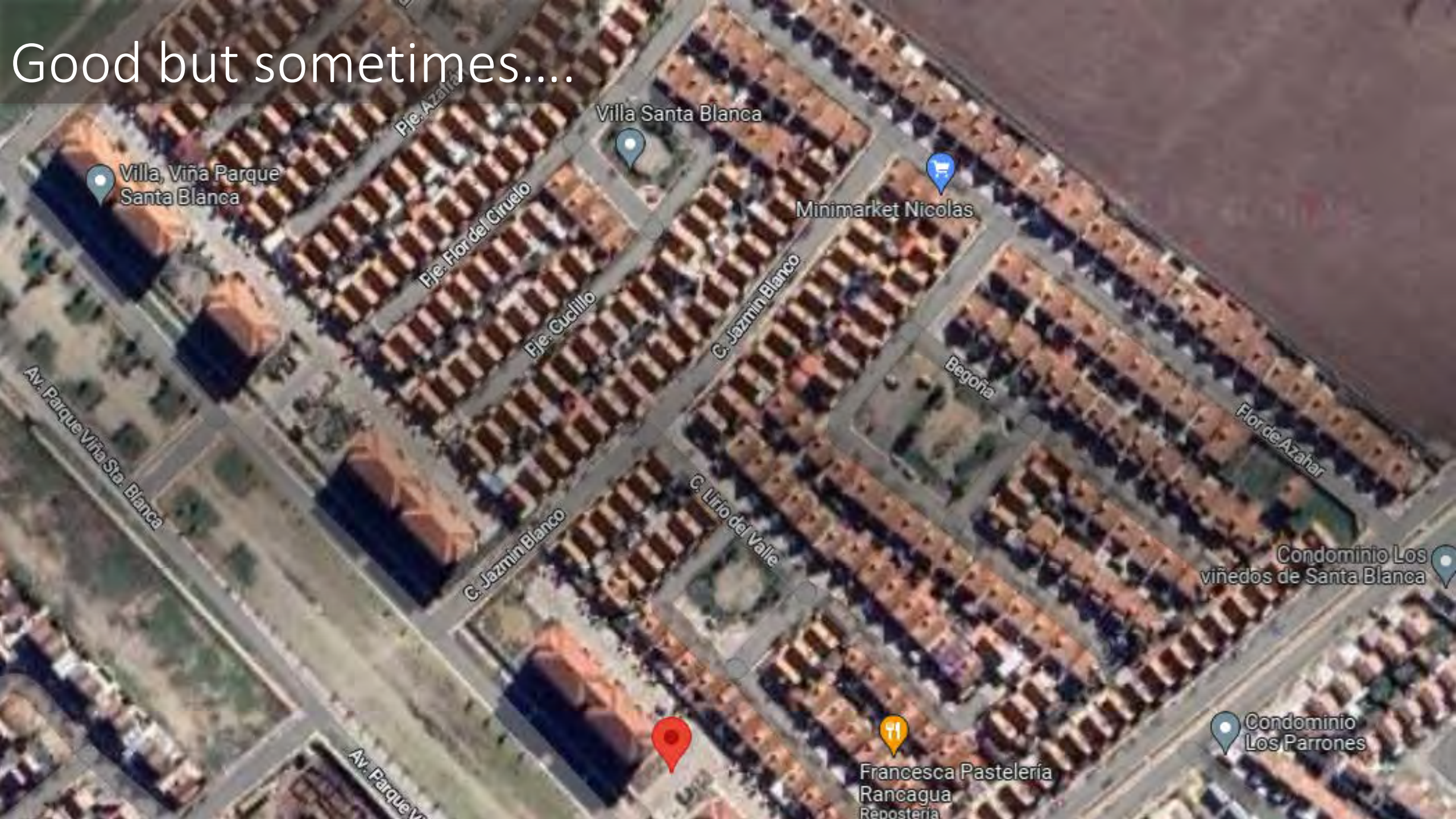


Manual Process



banco central Chile

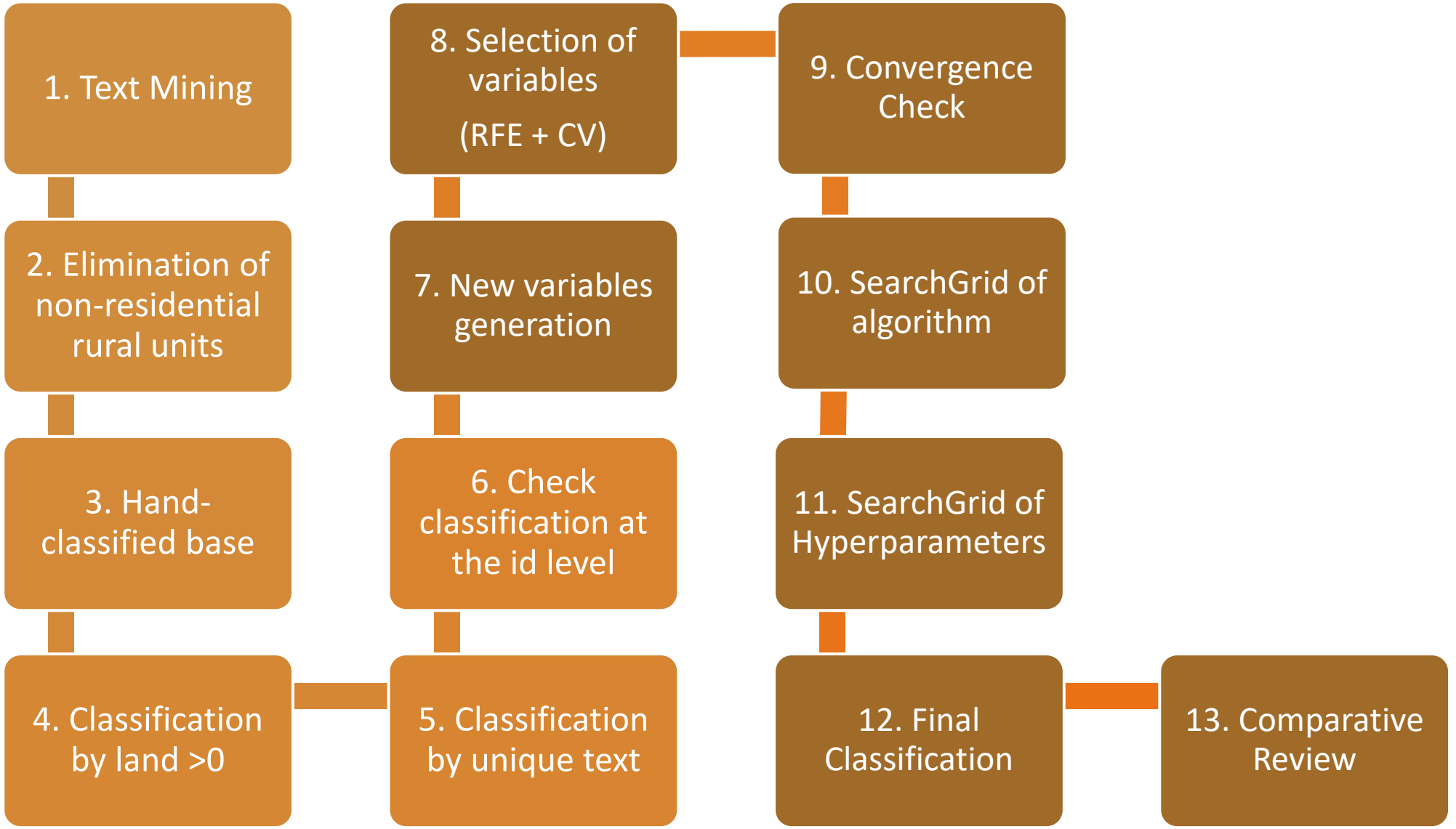
Good but sometimes....



Leave it the way it is? Change or Fix the previous algorithm? Machine Learning - AI?

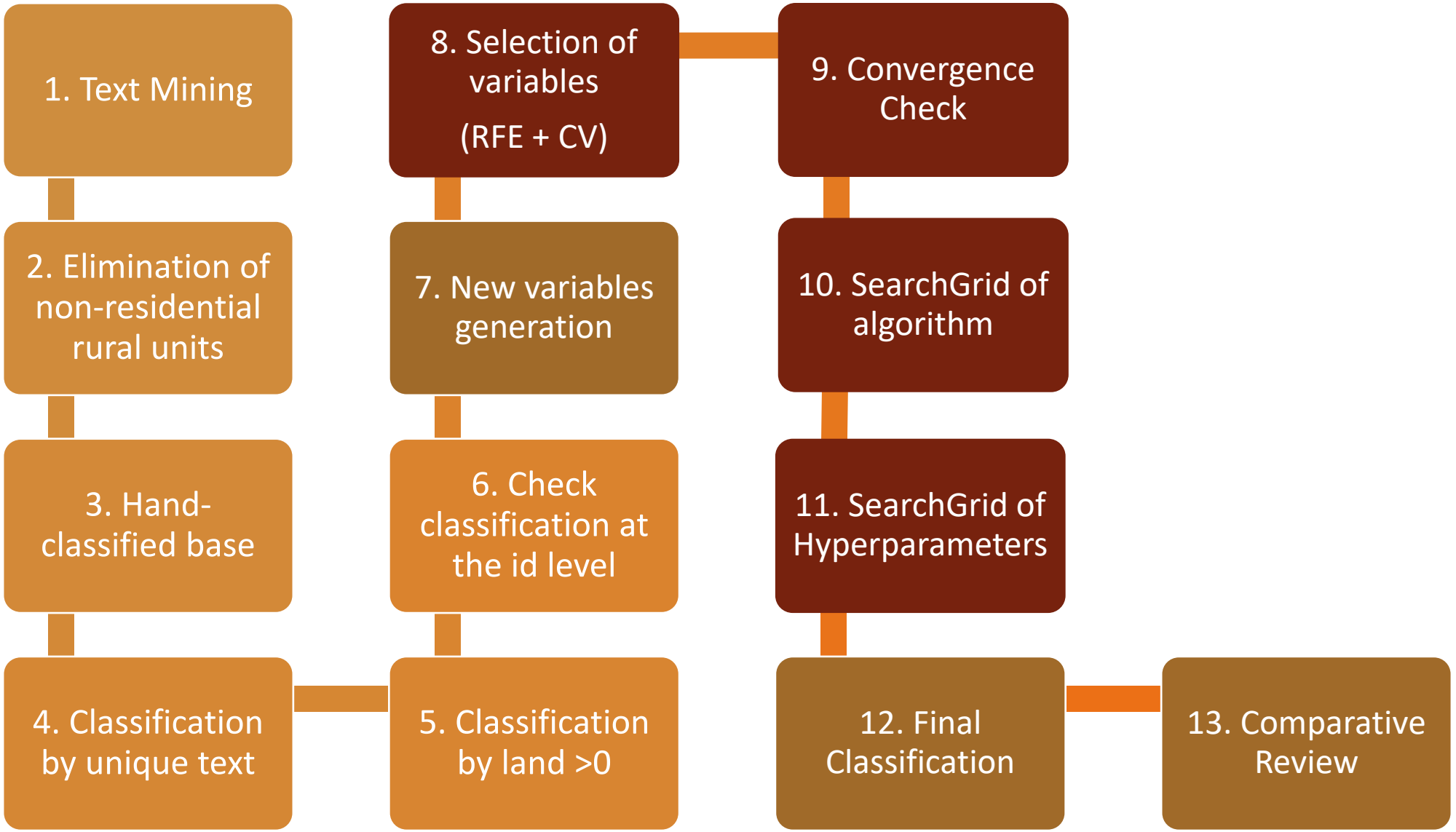
Improved process with ML

Data-Driven Process



banco central Chile

Data-Driven Process



banco central Chile

Characteristics of the Process

- The database had 34 variables and a training set of 3M transactions with 21 variables pre-selected with expert criteria.
- The algorithm selects 10 variables:

  Construction | Land | Credit ratio | Price ratio | Construction ratio | N° of unit per block

  Quality 1 and 2 | % of apartment by Census | Log of public valuation
- The algorithms (CV=5):

  Random Forest | Support Vector Machine | Logit | Kneighbors
- The winner = Kneighbors

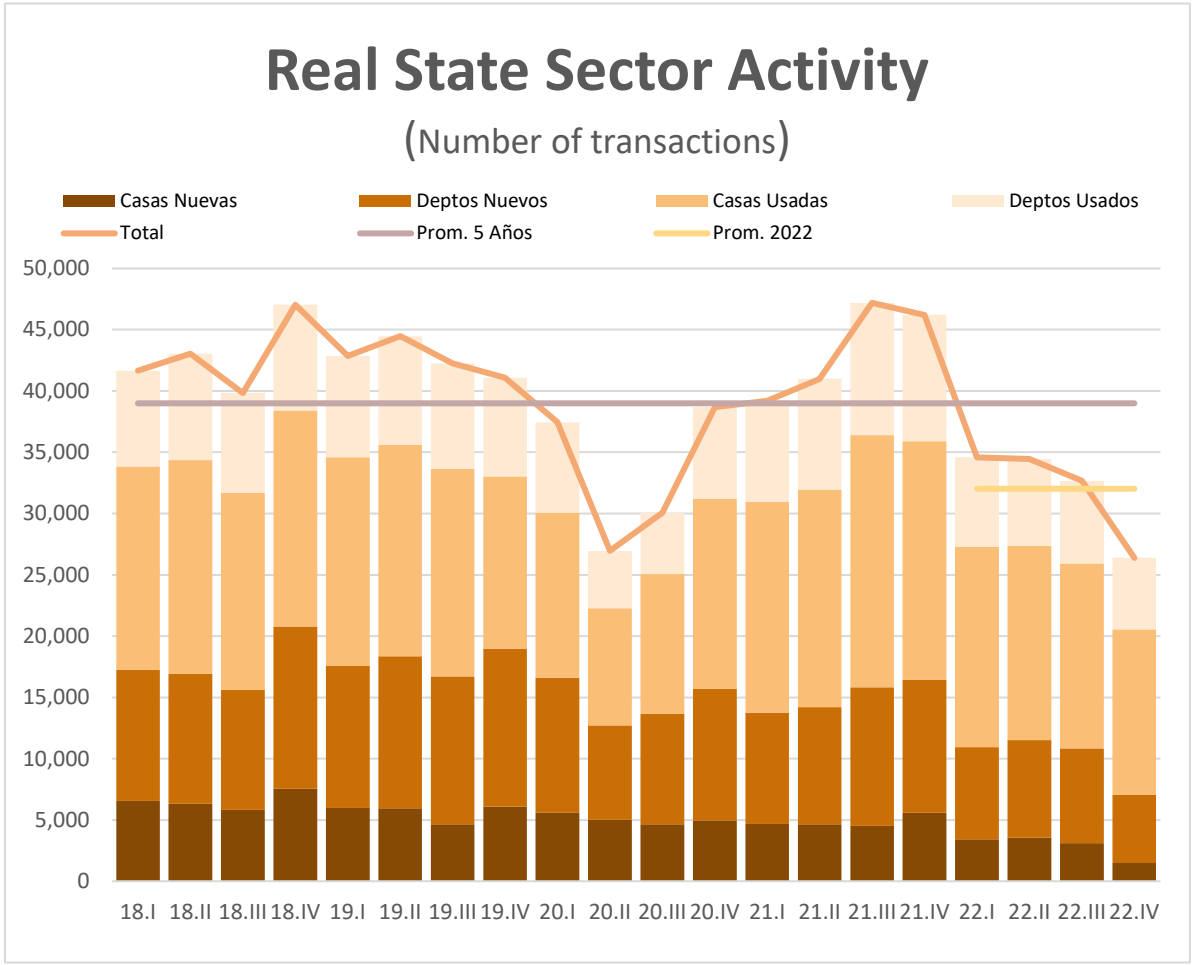
  Neighbors = 13 | weights = ‘distance’ | p = 1 | leaf size = 15

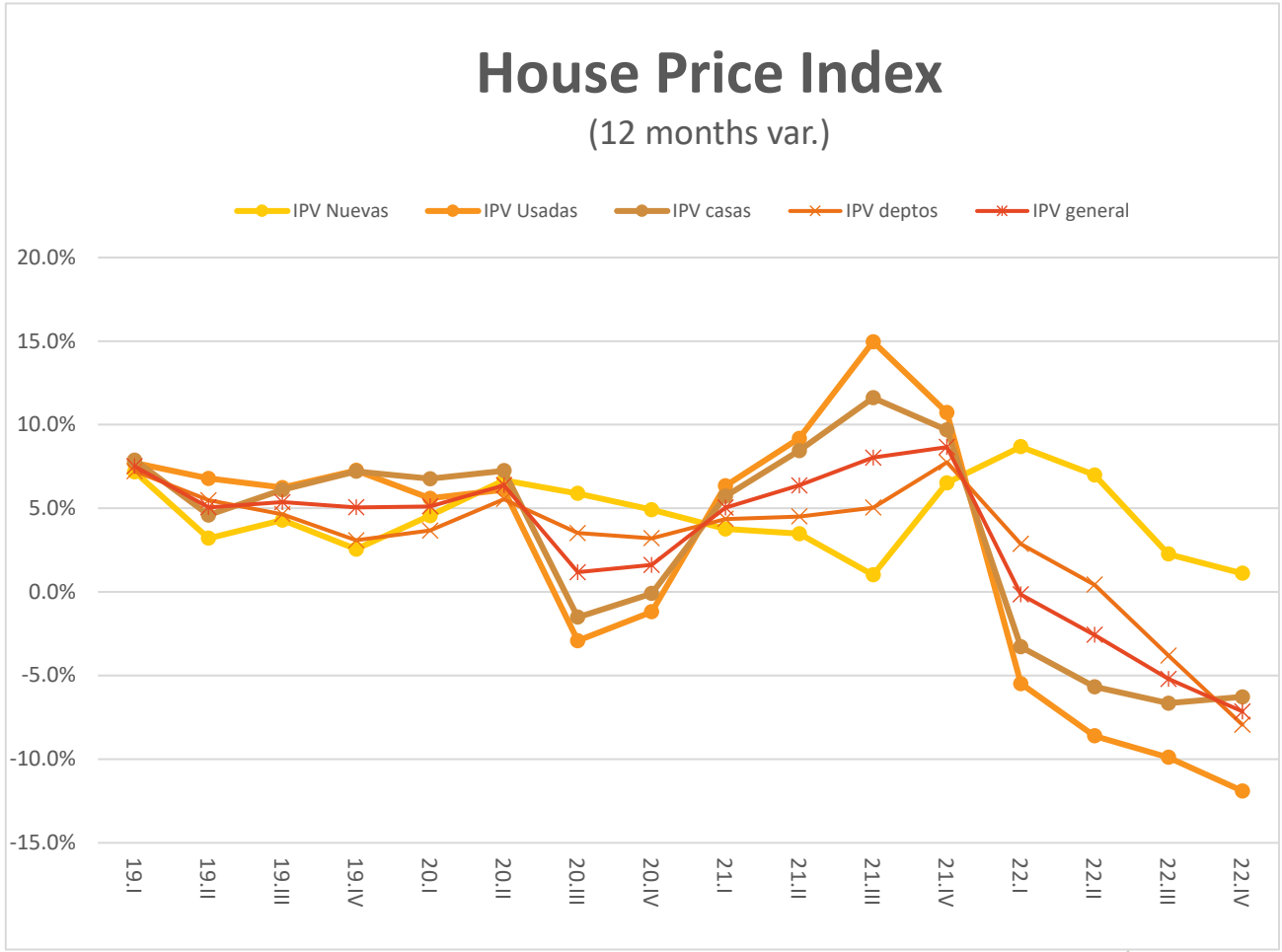
Results

Bottom Line

- The improved process is parallelized and takes 30 minutes
- The process reduces the working hours from several days to one morning with a level of accuracy very close to 0.
- The algorithm predicts with an accuracy of 99.2% in-sample and 98.9% out-of-sample
- The process improve with each iteration through the hand-classified database (250.000 vs 7.000 dif)

Uses of the data





banco central Chile

Conclusion



Available toolbox for everyone

- Simple process that only requires a training dataset and preferably some expert criteria for pre-selection of variables.
- Process usable for any type of binary problem (extendable to multicategory problems).
- Evolutionary and re-tunable process.

If it is possible don't focus on the specific problem, focus on how to solve the entire family of problems, and share your findings.

Property classification with administrative data: The case of the Chilean House Price Index

Department of Microdata

Statistics Division , July 2023

Annex 1: Stratification or "Composition Adjustment"

What is the stratification method?

- The stratification method, also known as "composition adjustment," is based on the creation of strata or groups of transactions that aim to neutralize variations in the composition of properties sold.
- Stratification involves breaking down the total sample of transactions into a series of sub-samples or strata. Then, for each stratum, the price index based on the mean or median is obtained to obtain the aggregated housing price index as a weighted average of the indices of each stratum. That is:

$$P^{0t} = \sum_{m=1}^{M} \omega_m^0 P_m^{0t}$$

- Where P_m^{0t} is the index of the stratum m that compares the average price of the period t is the period with base 0, and ω_m^0 denotes the weight of the stratum m.
- The formation of the strata is based on certain variables, such as geographic location, type of housing, housing condition, property size, etc. In this way, changes in the composition of housing between different periods are controlled, but not within each stratum.

banco central Chile



Thank you

Juan José Balsa F.
jbalsa@bcentral.cl

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Estimating effective inflation using scanner data[1]

Younghwan Lee,
Bank of Korea

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Estimating Effective Inflation Using Scanner Data

Younghwan Lee[1]

Abstract

This paper presents a cost of living index using Korean scanner data collected between February 2017 and October 2022. As most existing official statistics of consumer price indices measure the cost of living based on changes in the prices of a few selected varieties collected by survey methodology, this can be costly and may mislead policy makers and economists. By contrast, the new index accounts for changes in consumption patterns, such as substitution effects and changes in preferences, as well as changes in the market environment, such as the entry and exit of product varieties, by exploiting the information in the scanner data. Although scanner data are a valuable source of information for improving the consumer price index, there have been few implementation examples in practice owing to technical difficulties. In this paper, I propose some theoretically founded solutions with an implementation example. The proposed solutions can be implemented with manageable computing costs Empirically, the new index not only closely correlates with existing official statistics but it also statistically leads the existing indices, thus serving as a leading indicator for inflation dynamics.

Keywords: cost of living index, scanner data, chain index, Törnqvist index

Table of Contents

Estimating Effective inflation Using Scanner Data 1

1. Introduction 2

2. Methodology 3

   2.1 Cost of Living Index 4

   2.2 Chain Index 4

   2.3 Törnqvist Formula 6

   2.4 Rolling Window GEKS 7

3. Empirical Result 9

4. Concluding Remarks 12

References 13

[1] Data Research Section, Bank of Korea

E-mail: yhlee@bok.or.kr

Disclaimer: The views expressed in this article do not necessarily reflect those of Bank of Korea.

1. Introduction

Traditionally, many national statistical institutions (hereafter, NSIs) measure inflation using the fixed-base method which assumes that the consumption basket remains constant over time. However, this unrealistic assumption of consumer behavior and the market environment can mislead us measuring the price levels faced by households in their daily lives; thus, it is ineffective to capture economically meaningful price movements. This study presents a method for constructing an effective price index using scanner data. The proposed method allows for changes in the consumption basket by considering substitution effects, changes in preferences, and the entry-exit of items. This approach enables a more accurate measurement of effective prices.

Economists have long acknowledged the bias stemming from the use of traditional fixed-base methods, which are commonly linked to three primary sources: substitution bias, changes in preference, and entry-exit bias. Braun and Lein (2021) quantified the extent of this bias by measuring the inflation originating from these sources. Their findings indicate that overlooking changes in preference and entry-exit bias results in an average bias of 2.6% in the annual inflation rate, which is economically significant. The following three motivating questions illustrate how the fixed-base method can yield biased measurements of price levels in relation to the identified sources of bias.

Q. (Substitution bias) Suppose there is a product with a regular price of $100, but it is available at a 50% discount every Monday, Wednesday, and Friday. What is the effective price of this product?

Assuming rationality among consumers and the storability of the product, consumers will not purchase the product on Tuesdays, Thursdays, Saturdays, or Sundays at a price twice as high. In other words, consumers respond to price changes to minimize spending while maximizing welfare. Therefore, assuming a fixed consumption basket can lead to overestimating the price levels.

Q. (Changes in preference) In 1976, the price of the Apple 1 was $666.66, and in 2014, it was sold for $0.9 million. What is the inflation rate of computer prices between 1976 and 2014?

Although computing power of the Apple 1 lags significantly behind that of modern computers, its price has increased substantially over the last 40 years. This is primarily due to changes in consumer preferences for Apple 1. When Apple 1 was first introduced in the market, consumers purchased it because of its computational ability. However, over time, consumers perceive it as an iconic piece of art rather than a computer. Therefore, interpreting the change in the price of the Apple 1 as indicative of changes in the price levels of computers is generally misleading.

Q. (Entry-exit bias) Did the price of stone axes increase during the COVID-19 period?

The price of stone axes can be a significant indicator of the cost of living during the Stone Age. However, it is irrelevant to track the cost of stone axes as a measure of the cost of living for modern households because we no longer use stone axes. Advances in technology, deterioration of profitability, and changes in government regulations influence the variety of products available in the market. Thus, if such

changes in the market environment are not considered when constructing a consumption basket, the measurement of price levels can become distorted.

The primary reason why the fixed-base method is preferred by many National Statistical Institutions is the difficulty in obtaining timely access to consumption quantities. To update the consumption basket to reflect changes in household consumption, data on consumption amounts should be available. However, collecting such information and updating the consumption basket directly in a timely manner can be costly. NSIs, such as Statistics Korea, routinely update the consumption basket indirectly through household surveys.

To address this problem, I propose a method to construct an effective price index using scanner data. Scanner data provides detailed records of real transactions in the market, including the sales price of specific items and the quantities sold. Assuming that the timing of transactions and consumption are sufficiently close, scanner data can shed light on the evolution of household consumption behavior. Specifically, the effective price index is defined as a chain index that operates under the Törnqvist formula derived from the household consumption problem. Based on an economically founded model, the proposed method accounts for changes in consumption behaviour, thereby enabling a more accurate measurement of price levels. Additionally, the Rolling Window GEKS method is employed to manage the measurement error problems associated with the chain method, ensuring index stability.

The remainder of this paper is structured as follows. Section 2 outlines the methodology in detail. Section 3 applies the proposed method is applied to Korean scanner data to construct an effective price index and examines its empirical properties alongside the official consumer price index. Finally, Section 4 concludes the paper.

2. Methodology

The effective price index in this study is defined as the cost of living index (COLI), an economically founded concept. It is defined as a minimum level of expenditure required to sustain a constant level of utility. The Törnqvist formula and the chain identity are derived from the properties of COLI. From the definition of COLI, the consumption basket determined by Hicksian demand which evolves according to the changes in prices and preferences of consumers. In addition, the fact that the COLI has chain identity theoretically supports the use of the rolling windows GEKS (RWGEKS) method which is applied to suppress the chain drift problem in this study.

This section proceeds as follows. First, I derive the COLI from the economic model and investigate how changes in prices and preferences of consumers interact with COLI. Specifically, the COLI and Hicksian demand are derived from the cost minimization problem to shed light on possible sources of bias in the fixed-base method. Second, the theoretical and empirical support for the use of the chain index is provided. Third, the Törnqvist formula is derived as a COLI measurement. This implies that applying Törnqvist formula can alleviate the bias caused by the substitution effect and changes in preferences. Fourth, the RWGEKS method is introduced to suppress measurement error accumulation.

2.1 Cost of Living Index (COLI)

Assume an environment where a representative household minimizes the cost of living for each period t while achieving the level of utility $\bar{U}$. Let the household's utility at t be $u(c_t \mid \alpha_t)$ where the price vector of items and the parameter vector of preference of household are denoted by p_t and α_t, respectively. Then the Hicksian demand $c^*(\bar{U}, p_t|\alpha_t)$ is a cost-minimizing consumption bundle that is the solution to the following cost-minimization problem:

$$c^*(\bar{U}, p_t|\alpha_t) = \text{argmin}_c \, p_t^T \cdot c \quad \text{s.t.} \quad \bar{U} = u(c_t|\alpha_t).$$

Then, $\Pi_{0,t}$, the COLI between period 0 and t is derived as follow:

$$\Pi_{0,t} = \frac{p_t^T \cdot c^*(\bar{U}, p_t|\alpha_t)}{p_t^T \cdot c^*(\bar{U}, p_0|\alpha_0)}. \qquad (1)$$

From the above, we observe that COLI monitors changes in living expenditures through a weighted sum of prices, where the weights are determined by Hicksian demand. Two noteworthy observations were made. First, assuming that the other factors remain constant, households respond to changes in their preferences by increasing their consumption of preferred goods. In other words, as a product becomes preferable while its price remains constant, consumers can minimize their cost of living by increasing their consumption of these preferred goods and reducing their consumption of comparatively less preferred goods. This implies that the consumption basket varies with changes in the preferences.

Second, an increase in product price does not necessarily lead to an equivalent increase in COLI. Suppose that the price of a product variety increases while other factors remain constant, and the utility function is sufficiently regular. Thus, it is more advantageous for households to shift their consumption from the product variety with increased prices to other, relatively cheaper products. Specifically, as long as the marginal utility of consuming products decreases, households can maintain their level of utility with reduced expenditures by reallocating their budgets to relatively less expensive products. This finding demonstrates that the consumption basket responds to price changes, which can be a potential source of substitution bias that can overestimate consumer expenditure.

2.2 Chain Index

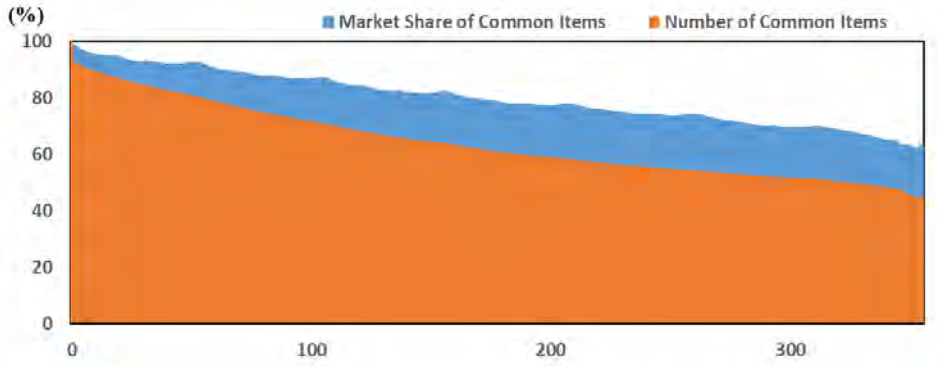
The chain index measures the change in price levels between two distant periods by accumulating changes in price levels over smaller time intervals. This allows us to update the composition of the consumption basket frequently while keeping it realistic. This method is justified by the chain identity property of the COLI, which is a consequence of equation (1). Specifically, the following relationship can be derived directly from equation (1):

$$\Pi_{0,t} = \Pi_{0,1} \times \Pi_{1,2} \times \cdots \times \Pi_{t-1,t}.$$

The theoretical implication of chain identity is whether to directly measure changes in the price level between two distant periods or indirectly calculate it by accumulating changes in smaller time intervals, the chain links, the results are equivalent.

Attrition Rate[1]

Figure 1.



[1] The vertical axis represents the percentage share of common items available in both measurement periods, while the horizontal axis represents the time interval between the two measurement periods in weeks.

Source: Author's calculation

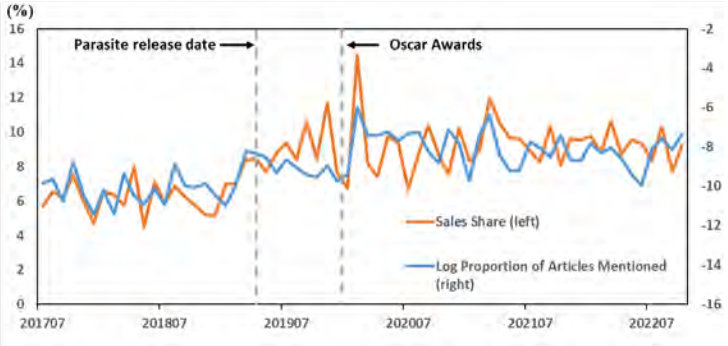
Although the chain index requires more computation and theoretically provides the same results as the direct comparison method, working with it offers two advantages. First, it can help alleviate entry-exit bias. This is because, as the measurement points become less distant, more common product varieties become available for comparing price levels. Figure 1 illustrates the proportion of common products in terms of the number of product varieties and their sales share relative to the time interval between the two measurement periods in weeks. This shows that the closer the time interval is, the more common the available product varieties are. Specifically, on average, approximately 20% (7% in terms of sales amount) of items disappear from the common product baskets within 1 year, and 45% (26% in terms of sales amount) of items disappear within 5 years. The fact that the proportion of items that disappeared from the common set in terms of the number of items is larger than that in terms of sales amount implies that the entry and exit of product varieties is not a random phenomenon. Rather, it suggests that relatively popular items tend to survive longer. This survival bias can also exacerbate the bias resulting from the entry and exit of product varieties when measuring the COLI.

Second, it is unrealistic to assume the same consumption basket for two distant periods because the composition of the consumption basket is subject to change for various reasons, such as the marketing activities of firms or seasonality. Figure 2 illustrates how the composition of a consumption basket can be altered for various reasons. For example, the Oscar-winning movie "Parasite" features a scene that includes the preparation of "Jjapagetti," one of the representative instant noodle brands in Korea. After the movie's release, the sales share of "Jjapagetti" began to increase and reached its peak after the film won the Oscar. Prior to the release of "Parasite," the average market share was approximately 6.4% in the instant noodle market. However, the average share increased to 9.3% immediately after the Oscar awards, which is economically significant. Interestingly, the proportion of economic

news articles mentioning " Jjapagetti" showed a similar increasing pattern during the same period. One plausible hypothesis is that consumers changed their preferences as the product gained popularity because of the movie, leading to adjustments in their consumption baskets. This example illustrates how quickly the composition of a consumption basket can change, even in a short amount of time.

Change in Sales Share of "Jjapagetti" [1]

Figure 2.



[1] The left vertical axis represents the market share of "Jjapagetti" in the instant noodle market, while the right vertical axis represents the logarithmically scaled proportion of economic news articles that mention "Jjapagetti."

Source: Author's calculation

2.3 Törnqvist Formula

The Törnqvist index is the weighted geometric mean of the change in the price of product varieties in a household's consumption basket, where the weight is defined as the average expenditure weight of the product. In addition, it is derived from the definition of the COLI under certain mathematical conditions. The Törnqvist formula is as follows:

$$\Pi_{s,t} = \prod_{i \in I_{s,t}} \left(\frac{p_{i,t}}{p_{i,s}}\right)^{\frac{w_{i,s}+w_{i,t}}{2}}.$$

Here, $I_{s,t}$ denotes the set of available product varieties available both at time s and t, and $p_{i,t}$ and $w_{i,t}$ are price and expenditure share of product i at time t, respectively. Assume following Cobb-Douglas utility function of the household:

$$u(c_t|\alpha_t) = \prod_i c_{i,t}^{\alpha_{i,t}},$$

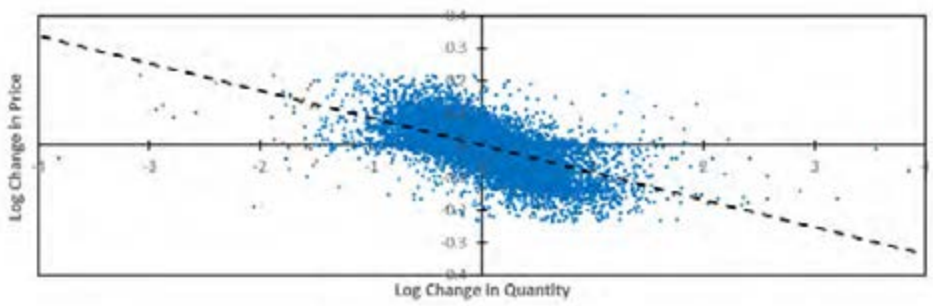
where $\sum_i \alpha_{i,t} = 1$ and $\alpha_{i,t} > 0 \; \forall$ i. In this case, the COLI is equal to the following:

$$\Pi_{s,t} = \prod_{i \in I_{s,t}} \left(\frac{p_{i,t}}{p_{i,s}}\right)^{\alpha_{i,t}}.$$

In equilibrium, each $\alpha_{i,t}$ equal to the expenditure share of product i at time t. The above COLI is equal to the Törnqvist formula if we substitute $\alpha_{i,t}$ by $\frac{w_{i,s}+w_{i,t}}{2}$. In other words, if we approximate the expenditure share at times s and t by their average expenditure share, the Törnqvist index becomes the COLI.

Relationship between Change in Price and Price in Sales Quantity of "Instant Noodle" [1]

Figure 3.



[1] Each point in the figure shows the weekly log change in price and quantity sold for 'Instant noodle' category items at 90 supermarkets from July 2017 to Oct. 2022.

Source: Author's calculation

The Törnqvist index formula, derived as a COLI that operates under Hicksian demand, effectively addresses the bias stemming from substitution effects. Figure 3 illustrates the statistically significant negative correlation between changes in the sales quantities and prices. This indicates that households adjust their consumption baskets in response to price shifts, decreasing the demand for higher-priced items. Neglecting this substitution effect by assuming a fixed consumption basket could potentially overstate inflation because an increase in the prices of certain product varieties may be misinterpreted as an increase in the aggregate price level. However, the Törnqvist index formula consider changes in the consumption basket owing to substitution effects, allowing for a more precise measurement of the cost of living.

2.4 Rolling Window GEKS (RWGEKS)

The fundamental challenge encountered when working with a chain index is the tendency of measurement errors in each chain link to accumulate and amplify rather than cancel out. This issue is akin to the unit-root problem in a time-series analysis. To address this issue, the Rolling Window GEKS (RWGEKS) method proposed by Ivancic et al. (2011) was employed. Recognizing how the accumulation of measurement errors can lead to excessive volatility in price indices is essential. Denote $\ln \Pi_{s,t}$ by $\pi_{s,t}$. The chain index is defined as follows:

$$\pi_{0,t} = \sum_{s=1}^{t} \pi_{s-1,s}.$$

However, it is important to note that we cannot directly observe $\pi_{s-1,s}$. Instead, the best we can obtain is its noise-contaminated estimate, denoted as $\tilde{\pi}_{s-1,s}$ where $\tilde{\pi}_{s-1,s} = \pi_{s-1,s} + \varepsilon_{s-1,t}$. The error in $\tilde{\pi}_{0,t}$ as calculated by the chain index increases as t advances, as shown below:

$$\begin{aligned}\tilde{\pi}_{0,\mathrm{t}} - \pi_{0,\mathrm{t}} &= \sum_{\mathrm{s}=1}^{\mathrm{t}} \tilde{\pi}_{s-1,s} - \sum_{\mathrm{s}=1}^{\mathrm{t}} \pi_{\mathrm{s}-1,\mathrm{s}} \\ &= \sum_{\mathrm{s}=1}^{\mathrm{t}} \varepsilon_{\mathrm{s}-1,s}.\end{aligned}$$

This implies that the price index becomes more uncontrollably erratic as the chain grows longer. To mitigate this noise, a is required to suppress it.

The RWGEKS method addresses this issue by identifying multiple potential measurements of the chain link $\pi_{s-1,s}$ using the chain identity. It then computes a proxy $\hat{\pi}_{s-1,s}^{GEKS}$ for $\pi_{s-1,s}$ by averaging these measurements. Specifically, for each $\mathrm{l} > 0$, we can obtain a noisy measurement $\tilde{\pi}_{s-1,s}^{l}$ for $\pi_{s-1,s}$ as follow:

$$\begin{aligned}\tilde{\pi}_{s-1,s}^{l} &= \tilde{\pi}_{l,s} - \tilde{\pi}_{l,s-1} \\ &= \pi_{s-1,s} + \varepsilon_{s-1,s}^{\mathrm{l}},\end{aligned}$$

where $\varepsilon_{s-1,s}^{l} = \varepsilon_{l,s} - \varepsilon_{l,s-1}$. Therefore, based on the definition of the RWGEKS chain link $\hat{\pi}_{s-1,s}^{GEKS}$, it becomes clear that the bias caused by measurement error can be controlled by selecting the appropriate value for L as follows:

$$\begin{aligned}\hat{\pi}_{s-1,s}^{GEKS} &= \frac{1}{L}\sum_{l=1}^{L} \tilde{\pi}_{s-1,s}^{l} \\ &= \pi_{s-1,s} + \frac{1}{L}\sum_{l=1}^{L} \varepsilon_{s-1,s}^{l}.\end{aligned}$$

The second equation shows that the size of error can be regulated by controlling for the number of averaged samples L. That is, the larger the L, the smaller the error term. However, using a larger L can reduce the sensitivity of capturing changes in the price level. Thus, it is important to determine the proper value of L.

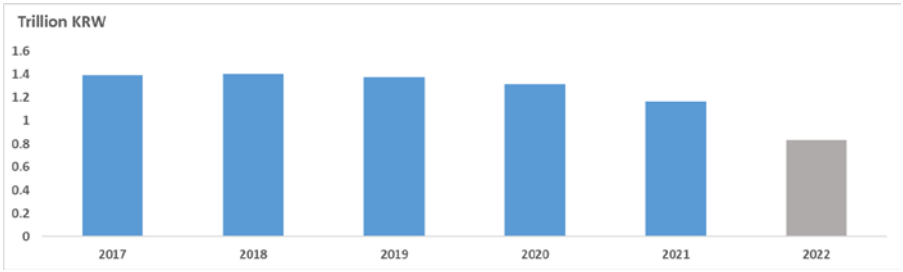
In this study, I constructed a price index with different values of L: 4, 13, and 52 weeks, which correspond to one month, one quarter, and one year, respectively. The empirical results suggest that the price index exhibits the best balance between stability and sensitivity when L is 13 weeks. Therefore, 13 weeks was chosen for this study.

3. Empirical Results

I applied the new method to scanner data collected from 90 supermarkets in Korea from July 2017 to October 2022 to test the empirical validity of the effective price index (EPI). Figure 4 shows that the data covers an average annual sales amount of approximately 1.4 trillion KRW, which is sufficiently large to assess the validity of the new index.[2] The data contains store-level information about the weekly sales amount and quantity of each product variety identified by its corresponding barcode. The barcode values for each product variety are universal across all stores, and their product names and categories are also identifiable. As the source of the scanner data is supermarkets, the variety of items included in the scanner data is limited to those typically sold at supermarkets. Approximately 80% of sales come from food items, and 20% come from living necessities.

Total Sales Amount of Sample Stores [1]

Figure 4.



[1] As data cover sales until Oct. 2022, the sales amount of 2022 is incomplete.

Source: Author's calculation

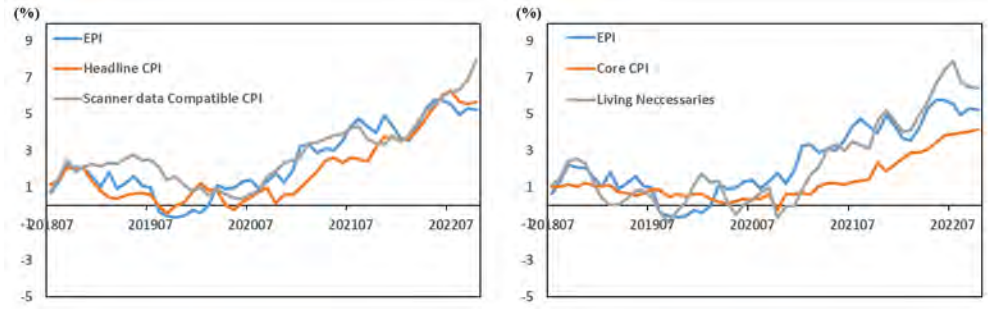
To investigate the empirical properties of the EPI, I compared it with various official CPIs, including the headline CPI, core CPI, living necessity CPI, and their variant, the scanner data-compatible CPI. The scanner data-compatible CPI is designed to serve as a benchmark for assessing the validity of the EPI because the new index covers a subset of the consumption basket owing to the limited coverage of scanner data. It is calculated using CPI values at the subcategory level, along with their respective weights in the CPI. These subcategories align with the items included in the scanner data.

Figure 5 plots the EPI alongside the official CPIs and the scanner data-compatible CPI. The correlation coefficient between the EPI and the headline CPI, core CPI, living necessity CPI, and the scanner data compatible CPI turn out to be 0.9, 0.88, 0.81, and 0.88, respectively. Notably, the correlation between the EPI and the headline CPI is slightly higher than that of the scanner data-compatible CPI, even though the EPI does not cover all product varieties included in the headline CPI. This could be attributed to indirect influences from changes in prices of other categories, such as oil prices, which are not directly reflected in the EPI but are captured through adjustments in consumption behavior and price changes in some product varieties in the scanner data, given the large number of product varieties in the EPI.

[2] The data set has been generously provided by The Korea Chamber of Commerce and Industry.

EPI and official Consumer Price Indexes [1]

Figure 5.



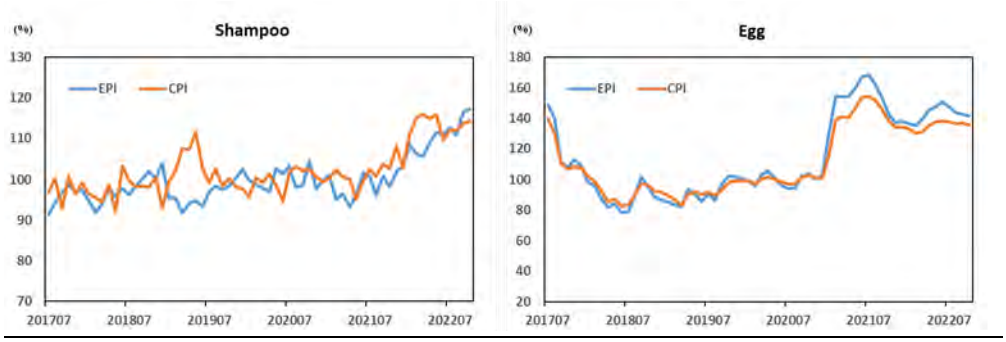
[1] All indexes are year-over-year based.

Source: Author's calculation, Statistics Korea

Where do these differences originate? One hypothesis is that the EPI reflects a competitive market structure different from the official CPI. The fundamental difference between the methods used to measure the EPI and official CPIs is that the EPI allows consumption baskets to evolve dynamically, whereas the price data of all product varieties in the scanner data are included. In contrast, official CPIs consider only changes in the prices of a few representative varieties. The strategy of relying on a few representative product varieties is effective when the product varieties within the category are highly homogeneous, and the market is close to perfect competition. In such cases, changes in a few representative product varieties provide sufficient information to describe price level changes.

EPI and CPI for "shampoo" and "egg" [1]

Figure 6.



[1] The average price level of 2020 is set to 100.

Source: Author's calculation, Statistics Korea

However, if the product varieties within a category are highly heterogeneous and the market exhibits monopolistic characteristics to some extent, tracking the prices of a few product varieties to measure the overall price level of the product category can be misleading. Figure 6 illustrates this point. It plots the EPI and CPI for the "shampoo" and "egg" categories. The "shampoo" category is presumed to be more heterogeneous than the "egg" category. Consequently, the deviation of the

"shampoo" EPI tends to be more pronounced compared to the official CPI than the "egg" EPI. In contrast, in the case of "egg," the EPI and the official CPI are closely aligned for the entire period. This serves as evidence to support the hypothesis.

Result of Granger Causality test (Z = EPI) [1]

Table 1.

Y	N of Obs.	Null hypothesis	
		Z ≠> Y	**Y ≠> Z**
Headline CPI	49	4.15	1.87
		(0.25)	(0.6)
Scanner data Compatible CPI	49	10.39**	1.5
		(0.02)	(0.68)
Core CPI	49	20.3***	2.95
		(1.5E-4)	(0.4)
Living Necessity CPI	49	5.22	2.91
		(0.16)	(0.41)

[1] The values in the third and fourth columns of the table represent the Wald statistics, with the corresponding p-values displayed in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

Source: Author's calculation

To assess whether the EPI can help predict the direction of inflation, a Granger causality test is conducted and the results indicate that the EPI is not only highly correlated with the official CPI but also significantly leads to the core CPI. Specifically, the model is assumed as follows:

$$y_t = \beta_0 + \sum_{m=1}^{M} \beta_m y_{t-m} + \sum_{n=1}^{N} \gamma_n z_{t-n} + \varepsilon_t.$$

The maximum time lags M and N are both set to 3. Using this model, whether z_t Granger cause y_t was tested under the null hypothesis that z_t does not cause y_t. That is, for example, to test whether EPI causes core inflation; z_t is set to EPI, and y_t is set to core inflation. Table 1 summarizes the results. This shows that the EPI Granger cause core inflation significantly while there is no sufficient evidence for the converse. As shown in Table 2, an alternative test was conducted to see if the headline CPI can also predict core inflation, but there is insufficient evidence to support it. This finding implies that the new index provides unique information that allows us to closely monitor and predict the future direction of inflation. To confirm the robustness of the empirical results, the same procedure was applied under alternative specifications, where M and N are set to 1,2, and 3, and the results were consistent over all specifications.

Result of Granger Causality test (Z = Headline CPI) [1]

Table 2.

Y	N of Obs.	Null hypothesis	
		Z ≠> Y	Y ≠> Z
EPI	49	1.87	4.15
		(0.6)	(0.25)
Scanner data Compatible CPI	49	6.3*	11.71***
		(0.1)	(0.01)
Core CPI	49	4.69	2.4
		(0.2)	(0.49)
Living Necessity CPI	49	4.74	2.83
		(0.19)	(0.42)

[1] The values in the third and fourth columns of the table represent the Wald statistics, with the corresponding p-values displayed in parentheses. *, **, and *** denote the 10%, 5%, and 1% significance levels, respectively.

Source: Author's calculation, Statistics Korea

4. Concluding Remark

Thus far, a method for constructing an effective price index using scanner data has been introduced, and its empirical properties have been investigated. This method allows us to accurately measure the inflation households face in their daily lives by operating under an economic model describing household consumption behavior. The empirical investigation shows that the EPI conveys information similar to that of existing official CPIs while leading the Core CPI. Thus, it can help monitor and forecast inflation closely.

References

Braun, R., & Lein, S. M. (2021). Sources of bias in inflation rates and implications for inflation dynamics. *Journal of Money, Credit and Banking,* 53(6), 1553-1572.

Ivancic, L., Diewert, W. E., & Fox, K. J. (2011). Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics*, 161(1), 24-35.

Estimating Effective Inflation Using Scanner Data

Younghwan Lee

Bank of Korea

Questions

- Suppose there is a product with a regular price of $100, but it is available at a 50% discount every Monday, Wednesday, and Friday. What is the effective price of this product?
- In 1976, the price of the Apple I was $666.66, and in 2014, it was sold for $0.9 million. What is the inflation rate of computer prices between 1976 and 2014?
- Did the price of stone axes increase during the COVID-19 period?

Ideas

- Challenges:
  - The fixed base index with a fixed consumption basket cannot take consumer behavior into account when measuring the effective price index.
  - The lack of quantity data is the root of the problem.

- Strategy:
  - Utilize **scanner data** and the **Tönqvist formula** to construct a **chain index** that captures changes in consumer behavior.
  - Using **RWGEKS** of Ivancic et al. (2011), to mitigate the amplification of measurement noise in the chain index.

Cost of Living Index (COLI)

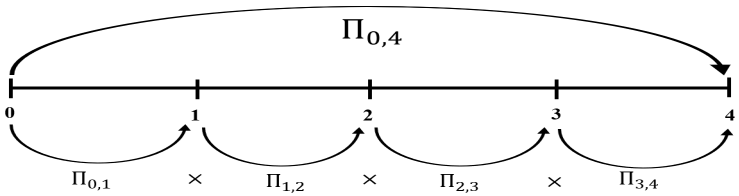
- The **Hicksian demand** $\mathbf{c}^*(\bar{U}, \mathbf{p}_t, |\alpha_t)$ that achieves the utility $\bar{U}$ given the price vector $\mathbf{p}_t$ and the time-varying preference α_t is the solution of the following cost minimization problem:

$$\mathbf{c}^*(\bar{U}, \mathbf{p}_t, |\alpha_t) = \underset{\mathbf{c}}{\operatorname{argmin}}\, \mathbf{p}_t^T \cdot \mathbf{c} \qquad \text{s.t.} \qquad \bar{U} = u(\mathbf{c}|\alpha_t).$$

- Accordingly, $\Pi_{0,t}$, the COLI inflation measure between time 0 and t, is defined as the change in expenditure required to sustain the household's level of utility, as follow:

$$\Pi_{0,t} = \frac{\mathbf{p}_t^T \cdot \mathbf{c}^*(\bar{U}, \mathbf{p}_t, |\alpha_t)}{\mathbf{p}_0^T \cdot \mathbf{c}^*(\bar{U}, \mathbf{p}_0, |\alpha_0)}.$$

Implication of COLI: Chain Identity



- Formally, the Chain Identity implies the following:

$$\Pi_{0,t} = \Pi_{0,s} \times \Pi_{s,t} \qquad \forall s.$$

- Thus, we can measure the COLI inflation by measuring a sequence $\{\Pi_{s-1,s}\}$ instead of directly measuring $\Pi_{0,t}$.

Why Chained Index?: Entry-Exit Bias

- The number common items that exist in both period decreases in the time difference between the measurement points.

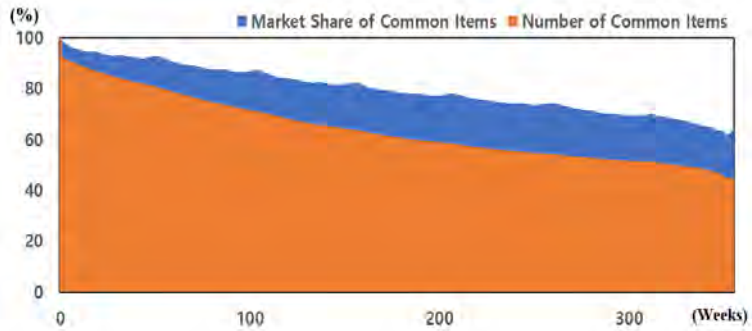


Figure: Attrition Rate

Why Chained Index?: Change in Consumer Preferences

- Consumer preferences can change over time...

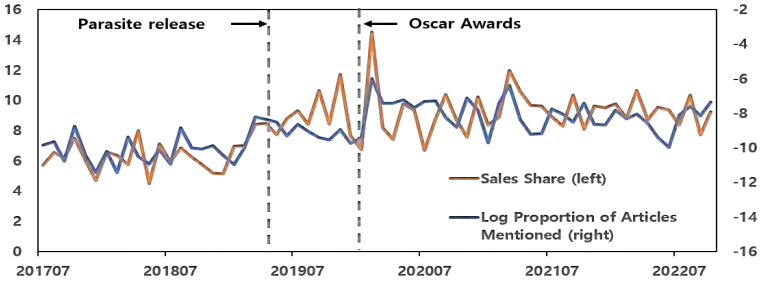


Figure: Sales Share of Jjapagetti

Implication of COLI: Törnqvist Formula

- The Törnqvist formula is defined as follows:

$$\Pi_{s,t}^{\text{Törn}} = \prod_{i \in I} \left(\frac{p_{i,t}}{p_{i,s}} \right)^{\frac{w_{i,s}+w_{i,t}}{2}},$$

  where $w_{i,t}$ is the sales share of the item i at t.

- Assume that the utility function is a Cobb-Douglas function as follow:

$$u(\mathbf{c}_t|\alpha) = \prod_{i \in I} c_{i,t}^{\alpha_i},$$

  where $\alpha_i = \frac{w_{i,s}+w_{i,t}}{2}$. Then, the COLI inflation between s and t coincides with the Törnqvist index.

Why Törnqvist formula?: Substitution bias

- Usually, an increase in price decreases the sales quantity.

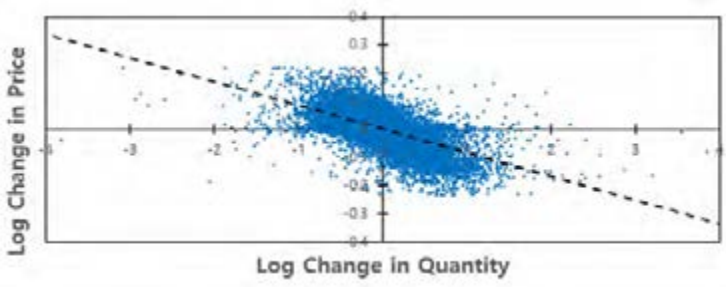


Figure: Substitution Effect in Instant Noodle Market

Why RWGEKS?

- Chain index is vulnerable to noise if the measurement error of each chain link is not suppressed. Let $\pi_{s,t}$ be $\ln \Pi_{s,t}$ and let $\hat{\pi}_{s,t} = \pi_{s,t} + \varepsilon_{s,t}$ be a noisy measurement of $\pi_{s,t}$. Then,

$$\begin{aligned}\sum_{s=1}^{t} \hat{\pi}_{s-1,s} &= \sum_{s=1}^{t} \pi_{s-1,s} + \varepsilon_{s-1,s} \\ &= \pi_{0,t} + \sum_{s=1}^{t} \varepsilon_{s-1,s}.\end{aligned}$$

- The idea of RWGEKS(Vancic et al., 2011) is to suppress the noise of each chain links by imposing chain identity. i.e. since $\pi_{t-1,t} = \pi_{l,t} - \pi_{l,t-1}$ for all l, RWGEKS chain link $\tilde{\pi}_{t-1,t}$ is defined as follow:

$$\tilde{\pi}_{t-1,t} = \frac{1}{|L|} \sum_{l \in L} \left(\hat{\pi}_{l,t} - \hat{\pi}_{l,t-1} \right).$$

Scanner Data

- Collected from 89 supermarkets in Korea from July 2017 to October 2022, the dataset covers an average of 80,000 items on a weekly basis.

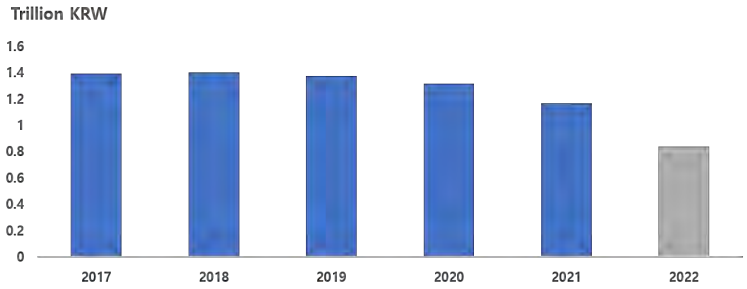


Figure: Total Sales Amount per Year

Comparison with Official CPI (1)

- The new index evolves similarly to the official Consumer Price Index (CPI).

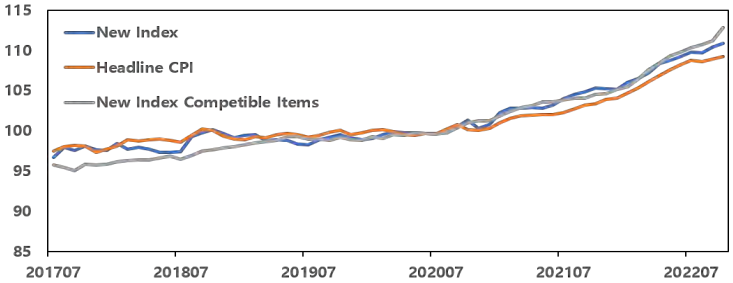


Figure: New Index vs Headline CPI (Level)

Comparison with Official CPI (2)

- Interestingly, the new index statistically more similar with headline CPI.

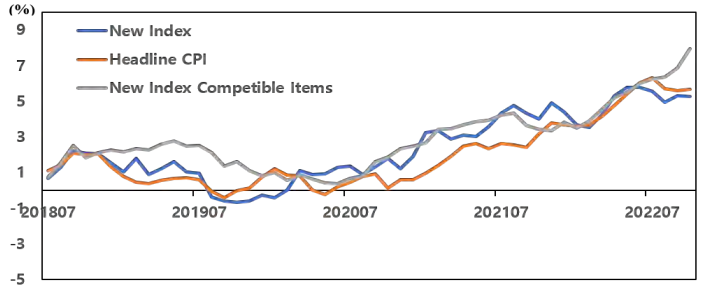


Figure: New Index vs Headline CPI (YoY)

Comparison with Official CPI (3)

- Statistically, new index lead the Core CPI significantly.

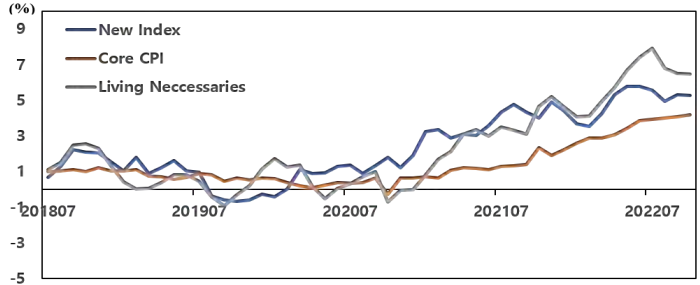


Figure: New Index vs Core Inflation and Living Necessaries (YoY)

Implication

- The fact that the new index evolve similar to and lead official CPI implies that granular data sets can help us to track and predict the official CPI.

- Remaining question is: Where do the differences originate from?

Implication (Cond.)

- The difference of monopolistic power or product heterogeneity could be a possible source of difference.



Figure: New Index and Official Core CPI (YoY)

- As the fixed base method does not take into account the change in consumption pattern, it could mislead the measurement of cost of living.

Future Work

- One promising direction of research is utility-based inflation measurement.
- Redding and Weinstein (2020) propose a method to measure the cost of living in an environment where consumer preferences change, assuming a CES utility function.
- In a similar context, Braun and Lein (2021) decompose the sources of bias in the fixed base method and find that the entry-exit of items and changes in preferences account for 2.6% of the bias.
- During the COVID-19 period, the degree of competition may have also changed.
- It is necessary to verify whether the change in market structure is a new source of the recently surging inflation.

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Financing imbalances and heterogeneous effects of monetary policy[1]

Jorge Antonio Pozo Sanchez and Youel Rojas Zea,
Central Reserve Bank of Peru

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Financing Imbalances and Heterogeneous Effects of Monetary Policy *

Jorge Pozo† & Youel Rojas‡

September, 2023

Abstract

This work documents the existence of a heterogeneous banking channel in Peru, where monetary policy actions have a differentiated effect on the credit market and which depends on the level of leverage of financial institutions. First, we demonstrate that the heterogeneity of the banking channel may not be attributed to asymmetric monetary policy shocks. Using aggregate data, we find a symmetric aggregate response of the credit market to expansive or contractive monetary policy shocks. Second, we characterize the heterogeneity of the banking channel based on the financing structure of financial institutions. Using financial entity level data, we demonstrate that the external financing component of the debt structure, measured by the level of leverage, primarily determines a heterogeneous monetary policy lending channel. Meanwhile, the internal banking network for mobilizing deposits and credit plays a less significant role. To show causal effects of this channel of financing imbalances, we use microdata at the branch level, together with an econometric strategy that focuses on local credit markets and a measure that considers the geographic distribution of financing imbalances. We show that the dependence on external funding at the branch level determines the heterogeneity in the credit channel, for the monetary policy transmission, with two important margins: an amplifying effect and another attenuating effect, associated with a lower or higher level of leverage, or with a greater or lesser need for external funding, respectively. To rationalize our findings we build a theoretical model. In the model, the financing structure of a bank generates effects that amplify or attenuate the effects of monetary policy.

Keywords: Regional heterogeneity, monetary policy shocks, leverage, regional imbalances.

JEL Classification: E5, G21

*Jorge Pozo Sanchez and Youel Rojas Zea thank the Association of Banks of Peru (Asbanc), which through its research contest in economics and finance Cátedra Renzo Rossini 2021 provided financial support to carry out this research. We also thank seminar participants at the Regional Seminar BCRP-Trujillo, the 2023 Conference of the Latin American Journal of Central Banking, the BoC-IFC satellite seminar on "Granular data: new horizons and challenges for central banks". The opinions expressed in this working paper correspond solely to those of the authors.

†Monetary Statistics Department at the Central Reserve Bank of Peru. Jr. Santa Rosa 441-445, Lima-1, Peru. Email: jorge.pozo@bcrp.gob.pe

‡Macroeconomic Modelling Department at the Central Reserve Bank of Peru. Jr. Santa Rosa 441-445, Lima-1, Peru. Email: youel.rojas@bcrp.gob.pe

Introduction

The financial system serves as the primary intermediary for facilitating the flow of funds between savers and borrowers, with banks playing a crucial role in reducing financial frictions in this process. Thus, the credit market serves as the main source of financial services for firms and families that carry out investment, saving and consumption activities. Moreover, it not only facilitates economic activity but also acts as a relevant channel for the transmission of monetary policy.

Monetary policy actions influence the credit market in two ways. On the one hand, by altering the cost of funding, they affect the credit supply decisions made by financial entities, which is called the bank loan channel (Bernanke and Blinder, 1988; Bernanke and Gertler, 1995). On the other hand, they also affect the demand for loans by altering the solvency of the firm equity, known as the balance sheet channel (Bernanke and Gertler, 1995; Bernanke, Gertler and Gilchrist, 1999; Kiyotaki and Moore, 1997). In equilibrium, the repercussion of a monetary policy shock on the credit market corresponds to a credit contraction (See for example Gertler and Karadi (2015)). But, the determination and final magnitude of the equilibrium depend on the relative importance of the supply and demand factors. Pozo and Rojas (2020b, 2021a) show that credit demand and supply factors are related in equilibrium, and a correct characterization of them is important for the design of policies. In this paper, we focus on the supply credit channel of monetary policy or the lending channel of monetary policy. In particular, using credit Peruvian data we characterize the heterogeneous equilibrium of the credit market, adequately controlling for supply and demand factors, and considering the credit and deposit structure.

We seek to answer the following questions regarding the effects of monetary policy on the credit market: Are credit market adjustments non-linear or do they depend on the contractionary or expansionary stance of monetary policy? More importantly, are the effects of monetary policy on the credit market heterogeneous? What are the implications of the credit and deposit structure in the transmission of monetary policy? By answering these questions, we can determine the causes behind the heterogeneity of the macro adjustment of the credit market. In addition, it is possible to specify the differentiated impacts and describe the operating margins of the monetary policy actions in the credit market.

We answer these questions using Peruvian data. In Peru, the capital market is still underdeveloped; therefore, the depository financial system mainly mobilizes financial resources from savers to borrowers. As a result, considering that financial frictions are more pronounced in Peru compared to more advanced economies, the impact of monetary policy is expected to be more pronounced via credit markets. This characteristic provides a favorable environment for our study.

Specifically, we focus on the lending channel and examine the heterogeneities in the

evolution of credit associated with changes in the monetary policy position. To do this, we use a two-pronged strategy. We first estimate empirical models at the bank and bank branch level, using publicly available data on loans and deposits for the period 2003:M1-2019:M12. Second, to rationalize our results we developed a simple model with heterogeneous banks or branches on the financing structure of the loans.

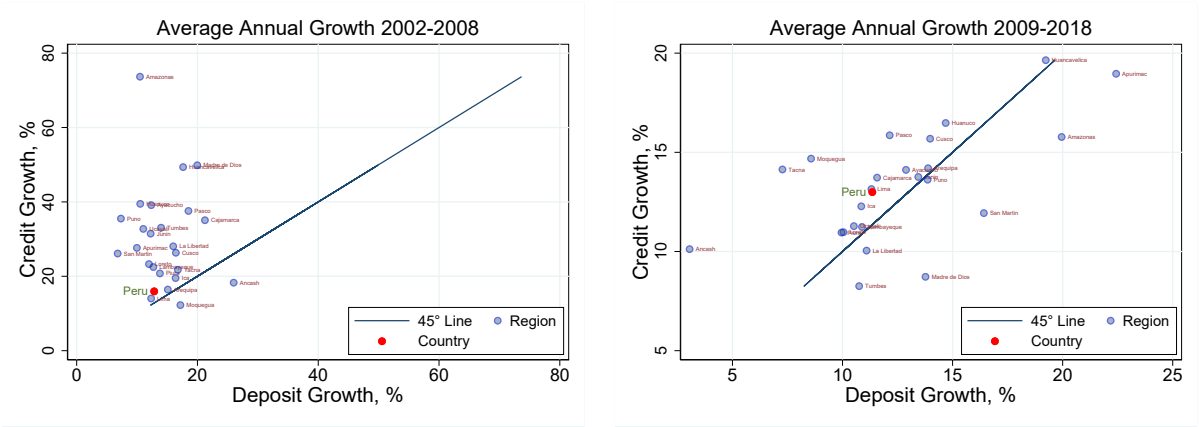
In our empirical exploration, to achieve identification, and to overcome the endogeneity problem of the monetary policy response to credit markets, first, we use a new time series of monetary policy shocks constructed by Pozo and Rojas (2022) for the Peruvian economy. As part of our initial results, at the aggregate level, we extend the results of Pozo and Rojas (2022) and show that contractionary or expansionary monetary policy shocks do not, on average, have asymmetric effects on the credit market. Second, similar to Pozo and Rojas (2021a, 2022), we use data at the bank branch level to compare the differential responses of branches to a monetary policy shock, in a specification that allows us to control for demand factors within a financial institution. An important assumption of our strategy is to consider that changes in the monetary policy rate affect the cost and/or the sources of credit financing and, therefore, the opportunity cost of deposits, which are the main source of funding for banks. However, at the local level, deposits, made for example in a district, do not necessarily coincide with credit financing needs. It is then that the reallocation of funds across different locations becomes important.

The mobilization of funds to finance credit does not occur in the same way in all regions of Peru. Figure 1 shows the differences in deposit and credit growth rates across regions for two periods: one before the international financial crisis, 2002-2008, and one after it, 2009-2018. In the 2002-2008 period, the acceleration of aggregate credit occurred at a faster rate than the expansion of deposits (with an average annual rate of 16% for loans, but 12.8% for deposits). In general, in all regions a greater acceleration of credit was observed in relation to deposits. In addition, a greater heterogeneity is documented in the dynamics of regional credit with respect to the dynamics of deposits. While average credit growth rates range from 12.3% to 73.7%, deposit rates only range from 6.8% to 26.0%.

However, after the great international financial crisis, in the 2009-2018 period, the expansion of aggregate credit has moderated, with an average annual growth rate of 13%, but deposits grew at a slower pace, at a rate of 11.4% per year, which is similar to that of the 2002-2008 period. However, although there is less growth dispersion between deposits and loans, there is still great heterogeneity between them at the regional level. This heterogeneity in the credit financing structure is a relevant characteristic that would explain the differentiated impacts of monetary policy on the credit market.

In the remainder of the analysis we proceed as follows: First, we use segregation metrics used in sociology (Duncan and Duncan, 1955; White, 1983; Aguirregabiria, Clark

Figura 1: *Regional dynamics of deposits and loans*



Note: Source: Superintendency of Banking, Insurance, and Private Pension Fund Administrators (SBS).

and Wang, 2019) to descriptively document the existence of an unbalanced geographic structure between deposits and loans, both for individual banks and for all banks located in a district. We show that some districts have a higher share of total loans in the country, but a low share of total deposits, and vice versa for other districts. Although there is no specific geographic pattern, these distributions remain relatively constant over time. At the level of shares of branches of a bank in a district, there is a high degree of inequality in both the distribution of credit shares and the distribution of deposit shares, where the shape of the loan share distribution has a tail that is fatter than that of the distribution of deposit shares. 50% of the district branches of financial entities have shares below 1% (0.2% for credit shares, and 0.8% for deposit shares). However, 1% of bank branches hold shares of more than 65% and 77% of a bank's total deposits and loans, respectively. To capture this differentiation in the geographic financing structure of financial entities, we compute a geographic imbalance index.

Second, we estimate the heterogeneous effect of monetary policy at the bank level. This effect would be a function of the financing structure, measured by leverage and the geographic imbalance index. The first measure considers the overall balance sheet structure of a financial institution, and therefore characterizes its external financing needs. In general, in this work we call "external financing" the funding that does not come from local deposits. The second one measures the structure of deposit and loan networks within a financial entity. Our results indicate that the leverage is the most important variable to determine the monetary policy power in the credit market, with an amplifying margin and another that attenuates monetary policy shocks. In particular, the greater the leverage (or less dependence on external financing), the less positive it will be the impact of an expansionary monetary policy shock. On the other hand, we find that, although a higher imbalance index amplifies the impact of monetary policy, these effects are not significant, and probably converge with that of leverage. Additionally, our specification at the bank level does not allow us to identify the loan channel since we cannot differentiate this

supply effect from demand factors.

Third, we exploit the geographic variation in the credit-to-deposit ratio induced by differences in the concentration of local credit and deposit markets. At the branch level of a financial institution, the credit-to-deposit ratio is an indicator of leverage or needs for external financing. To identify the imbalance channel, we need the variation of this indicator to be independent of the lending opportunities of the financial institution or the demand factors that influence the financial institution's decisions. To obtain this variation, we compare the loans between branches of the same financial entity located in different districts.[1] Under this strategy, we can control for loan demand opportunities that a bank faces, and we can identify the sensitivity induced by changes in monetary policy on the credit supply strategy of a bank branch. Our results indicate that the sensitivity of credit growth is negatively related to the credit-to-deposit ratio. This is indicative of the heterogeneity in the monetary policy transmission that depends on the relative size of external funding at the branch level and that it works in the expected direction from a given threshold.

Finally, we build a theoretical model to rationalize the results obtained about the non-linear monetary policy impact on credit at the bank and branch levels. According to the model, this is because the impact of the policy rate on bank profits and hence on the probability of bankruptcy, for a given level of credit, depends on the level of leverage or the financing structure.

In the model, the greater the leverage, the less positive impact of an expansive monetary policy. This is due to the greater leverage or greater dependence on local deposits amplifies the positive impact of a lower deposit rate, which leads to a stronger reduction of bank or branch default probability, ceteris paribus. In the general equilibrium, this reduces bank incentives to issue excessive loans. In addition, the heterogeneous impacts are better identified at the bank branch level due to the greater leverage variability by definition at the branch level.

The rest of this document is organized as follows: section 1 presents the literature review of the related research work. Section 2 presents aggregate evidence of the impact of the monetary policy stance on credit, and also discusses the importance of asymmetric effects. Section 3 describes data at the bank and branch levels. The section 4 shows descriptive evidence of the geographic imbalance in the structure of loans and deposits, at the level of local markets and banks. The section 5 shows estimates of the heterogeneous effect of monetary policy, but at the level of financial institutions. Section 6 presents the identification strategy and the estimates with granular information, and robustness

[1]Notice that we collapse information at the bank-district level, since it is not possible to track information of more than one bank branch in a district. The districts of Peru are the third-level country subdivisions of Peru. They are subdivisions of the provinces, which in turn are subdivisions of the regions.

exercises. Section 7 discusses the theoretical model with banks and with heterogeneity in the credit financing structure in order to rationalize the empirical results. Finally section 8 concludes.

1 Literature Review

This document is related to the bank lending channel literature as in ?Kashyap and Stein (1994). In this literature, the bank lending channel depends on the supply of required reserves, which is under the control of the central bank, and through this, the central bank controls the size of bank balance sheets. However, as the monetary policy framework shifted from control of aggregates to control of the interest rate, the role of reserve requirements diminished (Galí, 2015). However, as Kashyap and Stein (200); Jiménez et al. (2017); Drechsler et al. (2017) shows, the central bank's market operations alter the opportunity cost of financing and influence the bank's lending decisions. Our main contribution to this literature includes and explanation of heterogeneity of the lending channel. Also, we provide a set of empirical estimates on this channel in an emerging market economy where the monetary policy framework involves changes in the interest rate and the reserve requirement. We also provide theoretical results to rationalize our findings.

Quispe (2001), Carrera (2011), Viladegut and Cabello (2011), Bustamante et al. (2019) and Pozo and Rojas (2021a), are previous empirical studies that have documented the lending channel of monetary policy for Peru. However, none of these previous studies have characterized the heterogeneous equilibrium of the credit market, adequately controlling for supply and demand factors, and considering the credit and deposit structure.[2] Moreover, our work formalizes and extends this previous evidence of a lending channel in Peru.

Our work is also related to empirical research that studies the impact of conditional monetary policy on the financing structure of credit. For their part Dell'Ariccia et al. (2016) using loan information at the bank level for the period 1997-2011 in the United States, find that the impact of monetary policy on bank risk taking is conditional at the level of bank capitalization. In particular, the less capitalized a bank is (or the higher its leverage level), the less negative the impact of the contractive monetary policy on risk taking. Unlike Dell'Ariccia et al. (2016) we focus on the credit volume and not only at the bank level but also at a bank branch level.

Recently, Li (2022) using monthly data for the 1994-2017 period, finds that monetary policy in the United States is more effective in reducing financing costs, when financial

[2]Exceptions are Bustamante et al. (2019) and Pozo and Rojas (2021a) that also differentiate between demand and supply factors of monetary policy, but the first seeks to explain the determinants of loan growth and the second documents the risk-taking channel.

intermediaries are well capitalized (or have low leverage). This allows for a greater boost in actual activity. Likewise, the authors show that the lower effectiveness when in a state of nature with a low capital ratio would not necessarily be associated with the lower bound limit of the policy rate.

Bräuer and Rünstler (2014) in an empirical study for 10 Eurozone countries in the 1982-2017 period, find that the credit channel of monetary policy is more important in periods where the bank leverage is higher. This is due to the relatively more important impact on bank equity. Similarly, De Groot (2014) in a general equilibrium model finds that highly leveraged banks experience a high change in equity after a monetary policy shock and therefore decide to adjust the supply of credit more aggressively. Unlike these authors, this paper explores the excessive risk-taking channel, and we find evidence at bank and bank branch levels that the leverage or dependence on local funding (deposits) could diminish the monetary policy impact.

The theoretical model presented in this paper is related to the literature that models the interaction of limited liability and deposit insurance to explain excessive bank risk-taking (see, p. eg, Agur and Demertzis, 2012, 2019; Collard et al., 2017; De Nicolò et al., 2012; Pozo, 2019). Unlike Collard et al. (2017) and similar to Pozo (2019), the bank default probability is determined in general equilibrium, which allows us to measure the policy rate impact on bank excessive risk-taking' decisions.

In addition, this article is related to the literature that studies the different channels through which monetary policy can affect not only credit but also bank risk-taking (see, p. eg, Agur and Demertzis, 2012, 2019; Dell'Ariccia et al., 2014, 2016). Two main channels are suggested: the *profit channel* (or bank profit channel) and the *leverage channel* (or leverage channel). According to the *profit channel*, a lower policy rate increases bank profits in good states of nature and thus reduces bank incentives to take risk. The *leverage channel* suggests that a lower rate makes leverage less expensive. This means that the bank internalizes its risk-taking less and its incentives to assume more risks increase. Dell'Ariccia et al. (2014) conclude that when leverage is endogenous, low domestic interest rates lead to greater risk-taking by banks. However, if the level of leverage is exogenous, the effect depends on the level of leverage as follows: the higher the leverage, the higher the probability that a lower rate will reduce bank risk-taking. Our model also captures these two channels, but in a different setup. Thus, the monetary policy impact is non-linear on bank leverage, due to the *profit channel*, which has an impact not only on risk-taking incentives but also, unlike Dell'Ariccia et al. (2014), on the loan level. In this case, the higher the banks' leverage, the lower the positive impact on credit of an expansive monetary policy.

2 How does the credit market adjust to a monetary policy shock?

First, before documenting the heterogeneous effects of changes in monetary policy at the micro level, we present results documented in Pozo and Rojas (2022) and extend their estimates to differentiate between possible asymmetries of changes in monetary policy, between an expansive or contractive stance. In particular, in this section we demonstrate that the heterogeneity of the banking channel may not be attributed to asymmetric monetary policy shocks.

In particular, we adopt Pozo and Rojas (2022)'s estimation strategy that considers a flexible form that imposes few restrictions on the data to establish causal effects on the aggregate variables. However, since the monetary policy rate is endogenous to the economic cycle and credit market, an exogenous monetary policy shock is necessary. Therefore, from now on, we use the new series of monetary policy shocks constructed for the Peruvian economy by Pozo and Rojas (2022).[3]

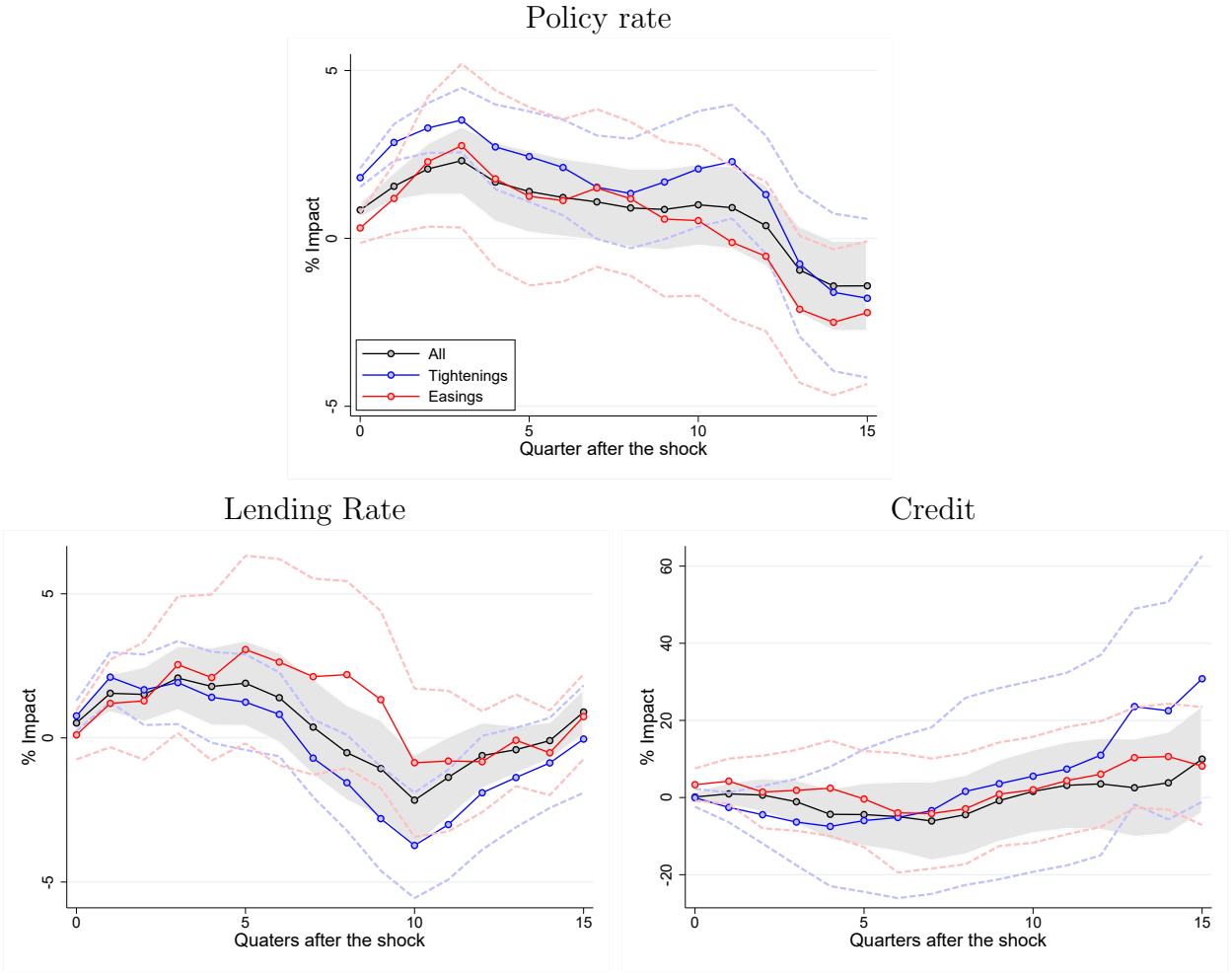
Second, we use local projections (Jordà, 2005) to model the dynamic equilibrium response of credit and the lending rate after the monetary policy shock. Specifically, let y_t be the outcome variable at time t, we estimate a series of the following regressions:

$$y_{t+h} - y_{t-1} = \alpha_h + \beta_h \, \epsilon_m^{MP} + \sum_{k=1}^{K} \gamma_k^h X_{t-k} + u_t^h \,, \tag{1}$$

where $h = 1, 2, ..., 16$. The coefficients β_h indicate the percentage change in the h horizon in response to a 100 basis point monetary policy shock. X_t denotes a vector of controls. Following the same specification as in Romer and Romer (2004), it includes three years of lagged values of the monetary policy shock and two years of lagged values of the dependent variable. In addition, we added a dummy variable for the period after the first quarter of 2009. After the Global Financial Crisis, the main macroeconomic variables began to show lower growth rates and higher volatility. It is important to note that the specification (1) leaves the contemporaneous response unrestricted, and differs from a typical Cholesky identification generally used in VAR specifications. Following Ramey and Zubairy (2018), local projections allow us to estimate non-linear effects using the same specification (1) but with dummies variables.

[3] Pozo and Rojas (2022) construct a new time series of monetary policy shocks for Peru using the same approach as Romer and Romer (2004). To do this, they use annual GDP growth and inflation forecasts published by the Central Bank of Peru, and complement them with projections from professional forecasters, as in Holm et al. (2021), when the Central Bank's forecasts are not available. See Pozo and Rojas (2022) for a detailed explanation of the construction of these monetary policy shocks.

Figura 2: *Monetary Policy and Credit Market Shock*



Note: This figure shows the impulse response functions to a monetary policy shock. Base: Indicates the impact of a contractionary monetary policy shock of 100 bps as reported in Pozo and Rojas (2022). Expansive: indicates the impact of an expansionary monetary policy shock of 100 bps, where the specification (1) considers only negative monetary policy shocks. Contractionary: indicates the impact of a contractionary monetary policy shock of 100 bps, where the specification (1) considers only positive monetary policy shocks. Confidence bands are of 68%, using Newey and West (1987) standard errors. Quarterly frequency for the 2003:Q4-2019:Q4 period.

Results

The local projections, in specification (1), are estimated with quarterly frequency data for the 2003:Q4-2019:Q4 period. We consider the impact of a contractionary monetary policy shock of 100 bps on the credit market at the aggregate level: total credit and the average lending rate[4].

The impulse responses of credit market outcomes after the tightening monetary policy shock are shown in Figure 2. Considering the base result reported in Pozo and Rojas

[4]The data was obtained from the repository of the Central Bank of Peru, BCRPData, available in https://estadisticas.bcrp.gob.pe/estadisticas/series/. Total credit (PN00505MM) measured by Total credit to the private sector of deposit institutions, in millions of S/. end of quarter. The average lending rate is constructed as a credit-weighted average between the series of flow interest rates in local currency (PN07808NM) and flow interest rates in foreign currency (PN07828NM).

(2022), an increase in lending rates is observed after a 100 bps increase in the monetary policy rate: the lending rate rises and reaches a maximum increase of 2% after two quarters and remains around this level a year later. There is also a negative, persistent and economically significant response in aggregate credit after a monetary policy shock: credit fell by around 3% four quarters after the shock, and reached a drop of around 5% in the next year. What is most important is that when considering different monetary policy regimes there is no considerable evidence to indicate asymmetric effects in credit markets after an expansionary or contractionary monetary policy stance. Based on these results, from now on, when we study the heterogeneous effects of monetary policy at the micro level, we do not distinguish between positive or negative monetary policy shocks.

This aggregate empirical evidence confirms that the supply-side adjustment mechanism in credit markets is important for credit market equilibrium after a monetary policy shock. Motivated by this added evidence, in the following sections we analyze the channel of imbalances between credits and deposits or external leverage at the branch level as a credit supply-side mechanism that determines the equilibrium after a monetary policy shock.

3 Data

In general, we use two levels of information. First, following the related literature, we use data at the bank level. Second, we use information at the branch level of a financial institution of the financial system[5] that is not very common to see in the literature, due to the limited availability of such information. The latter is more granular and hence allows us to better control demand characteristics that can bias our results. Both sets of information are publicly available on the website of the Superintendence of Banking, Insurance and Private Pension Fund Administrators (SBS).[6]

In the next subsection, we describe the time-bank level data set, the empirical model, and the results of our first bank-level regression. And in subsection 3.2 we focus on the more granular data that will be part of the input of the main estimates presented in the following sections.

[5]The financial system in Peru is divided in five financial groups: commercial banks (banks), *empresas financieras*, municipal credit and saving institutions (CMACs by its Spanish acronym), rural credit and savings institutions (CRACs by its Spanish acronym) and small business and microenterprises development institutions (EDPYMEs by its Spanish acronym). However, we do not include EDPYMES in our analysis since they are not allowed to capture deposits.

[6]Available in https://www.sbs.gob.pe/estadisticas-y-publicaciones/estadisticas-/sistema-financiero_

3.1 Data at the bank level

We explore data at the bank level to assess the heterogeneous effects of monetary policy on credit.[7] In particular, we study its conditional effect on the bank leverage level (financing structure). Information on loans and other indicators of banking companies are provided by the (SBS). The variable we seek to explain is credit growth, Δy (calculated as the deviation of the logarithm of the bank's loan balance). We use the monetary policy shock measure identified for the Peruvian economy, ϵ_t^{MP}, presented in section 2.

In the analysis at the bank level, we include some variables to control for individual bank characteristics that vary over time, such as return on assets (ROA) and relative bank size or loan market share (SIZE). In addition, we include the risk-weighted asset-to-equity (RWA) ratio. The RWA ratio aims to control for the bank's resilience after a financial crisis, and for risk-taking incentives.[8]

Another control that we include is the non-performing loan (NPL) ratio defined as the ratio of loans in arrears to total loans under the same criteria defined by the Peruvian financial regulator, SBS.[9] It provides information of the non-performing loans at the institution-time level.

We use monthly data and the time period analyzed spans from 2003:M1 to 2019:M12. We start in 2003 to consider only the inflation targeting period in Peru and stop in December 2019 to avoid the Covid-19 pandemic period.

Table 1 presents the descriptive statistics. In our study period, we have an average of 15 banks and around 2700 observations. Average monthly credit growth is 1.2% with a relatively high dispersion. The risk-weighted assets to capital ratio is 6.8 and a relatively low dispersion considering. The average NPL ratio is 3.3% and there is a high dispersion. The average loan market share (SIZE) is 6.2%, with a relatively high dispersion. The return on assets (ROA) is the one with the highest relative dispersion. We also use a measure of the imbalance of deposits and loans at the bank level (GI). [10] Finally, we show the average of the monetary policy shock, very close to zero, with a standard deviation of 0.10%.

[7]The banking system represents on average 88% of total credit in the financial system in our study period.

[8]See Agur and Demertzis (2012, 2019) and Dell'Ariccia et al. (2014) for a detailed explanation of the effect of bank leverage on risk-taking (i.e., the *leverage channel*). Intuitively, the higher the leverage, the lower the share of the owners' wealth in financing the bank's investment activities. For the time being, the lower the losses from bank duels if the banks fail (due to limited liability) and, therefore, the higher the preference for taking on more risk.

[9]The definition of the non-performing loans ratio would be the following:

$$\frac{\text{arrears loans (big (15d), small (30d), mortgage (30d), personal (90d))}}{\text{Total Credit}} 100. \tag{2}$$

Big firms correspond to corporate, large-sized and mid-sized firms. Small firms refer to small-sized and micro-sized firms.

[10]The construction of this index will be seen in the 4 section.

Tabla 1: *Descriptive statistics of the observations at the bank-time level*

Variables	N° Obs.	Mean	Standard deviation	Min	Max
Δy_{bt}	2933	0.012	0.06	-0.62	0.80
RWA_{bt}	2786	6.800	1.66	0.80	10.24
NPL_{bt} (%)	2786	3.300	3.42	0.00	34.28
SIZE_{bt} (%)	2933	6.176	8.63	0.00	32.55
ROA_{bt} (%)	2748	1.711	1.99	-12.51	8.06
GI_{bt}	2784	0.270	0.21	0.00	0.78
ϵ_t^{MP} (%)	196	0.000	0.10	-0.23	0.35

Source: SBS. Own Elaboration. SIZE_{bt} = (total loans of bank b at time t)/(total credit at time t). We omit extreme values and keep $-1 < \Delta y_{bt} < 1$. We drop 10 observations.

3.2 Data at financial institution branch level

Table 2 presents the descriptive statistics of the main variables of the analysis. It shows averages of time series and cross-sectional data. The empirical analysis uses the loan variations at the branch level. **Indeed, in this paper branch level data is built collapsing information of all branches of a bank at the district level, since it is not possible to track information of more than one branch in a district.** There are 3,682 branches located in 451 districts and 189 provinces.

Tabla 2: *Descriptive statistics*

	All		Low C/D		High C/D	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Panel A. Province Characteristics						
Population (thousands)	203.2	765.9	332.1	1,070.0	74.2	49.2
Area (sq. km.)	6,915.1	12,489	7,451	15,128	6,379.0	9,196.9
Formal Employment (%)	2.6	3.7	4.4	4.4	0.9	1.5
Cred-Dep ratio (%)	187584	1.1e+06	155	56	375013.5	1.6e+06
Obs.(Provinces)	154		77		77	
Panel B. Branch Characteristics						
Credit (thousands S/)	114.1	949.8	160.0	1,091.7	68.1	779.5
Credit growth (%)	8.0	24	7.8	26	8.2	20.9
Deposits (thousands S/)	111.4	826.2	199.6	1,114.4	22.9	326.6
Deposit growth (%)	10.9	32	7.7	24	14.1	37.5
Cred-Dep ratio (%)	1,072.8	3,786	83	59	2,065.2	5,170.2
Obs.(Branch × month)	91,390		45,744		45,646	

Note: This table provides summary statistics at the province and branch levels. All panels provide a breakdown by high and low loan-to-deposit (C/D) ratio, using as a break-point the median. Panel A presents the characteristics for all provinces with at least one branch. The underlying data is from the 2017 census. Data on employment comes from SUNAT (Peru's tax authority). Panel B presents data on total credit and loan and deposit growth at the branch level. Source: SBS. Own elaboration.

All panels in table 2, for presentation purposes, provide a breakdown by high and low Credit/Deposit (C/D) ratio using the median as the threshold.Panel A of table 2 shows data statistics for all provinces with at least one branch. It is notable that the low C/D ratio provinces are larger and have more formal employment than the high C/D provinces. The average population in the low C/D ratio provinces is 333.7 thousand compared to 72.7 thousand in the high C/D ratio provinces. The formal employment rate in markets with a low C/D ratio is more than 4 times that of markets with a low C/D ratio: 4.4 vs. 0.9.

Summary statistics at the branch level are shown in Panel B of 2 table. Branches in low C/D ratio provinces are larger. The average branch with a high C/D ratio has loans and deposits of 160 thousand soles and 199.6 thousand soles, respectively, which are well above 68 thousand soles and 22.8 thousand soles, respectively, from low C/D ratio branches. However, on average high C/D ratio branches show a higher rate of credit and deposit growth relative to high C/D ratio branches.

4 Evidence of Geographic Imbalances between Loans and Deposits

In this section, we provide evidence of how spatially unbalanced deposits and loans are. As in Aguirregabiria, Clark and Wang (2019), we use segregation metrics used in sociology (Duncan and Duncan, 1955; White, 1983) to identify variations in the geographic distributions of deposits and loans, both for individual banks and for all banks.

Figure 3 shows the geographic distribution of districts as net borrowers or net lenders, measured by the district-level geographic imbalance index $GI_{S_L-S_D}$, which is computed by the difference between a district's share of the country's deposits, dep_{dt}/Dep_t, and its share of the country's loans, cre_{dt}/Cre_t:

$$GI_{S_L-S_D} = \frac{dep_{dt}}{Dep_t} - \frac{cre_{dt}}{Cre_t}.$$

For each year (2002, 2008, 2014, or 2019), returns the distribution intervals divided by the 10th, 50th, and 90th percentiles of $GI_{S_L-S_D}$. From lightest to darkest red, these boundaries form four groups of districts: i) Districts belonging to the bottom 10 percentiles, with very low loan shares relative to deposit shares, ii) districts between the 10th and 50th percentiles, with low loan-to-equity deposit ratios; iii) districts between the 50th and 90th percentiles, with percentages of loans above their percentages of deposits and iv) districts belonging to the top 10 percentiles, with percentages of loans that are very high in relation to their percentages of deposits.

Figure 3 shows evidence of deposit and loan imbalance. Some districts present higher participation of total loans of the country, but a low level of participation of the total

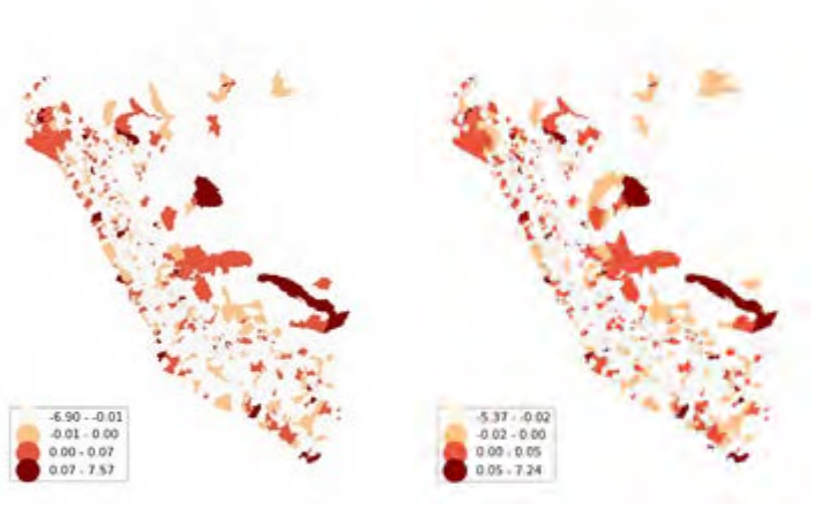
deposits, and the opposite is true for other districts. There is no regular regional pattern regarding the districts' concentration with net debtor position or net lender position. And it is interesting to see important changes over time. In 2002, mainly along the coast of the country were the net debtor districts. Consequently, the other districts financed these loans and concentrated the largest share of the deposits. In 2008, the figure changed and the imbalance map becomes more heterogeneous throughout the country. However, between 2008 and 2019 the geographical distribution of net borrowing and net lending districts remains relatively constant.

Figura 3: *Geographic distribution of districts as net borrowers or net lenders*

A. 2002 B. 2008



C. 2014 D. 2019

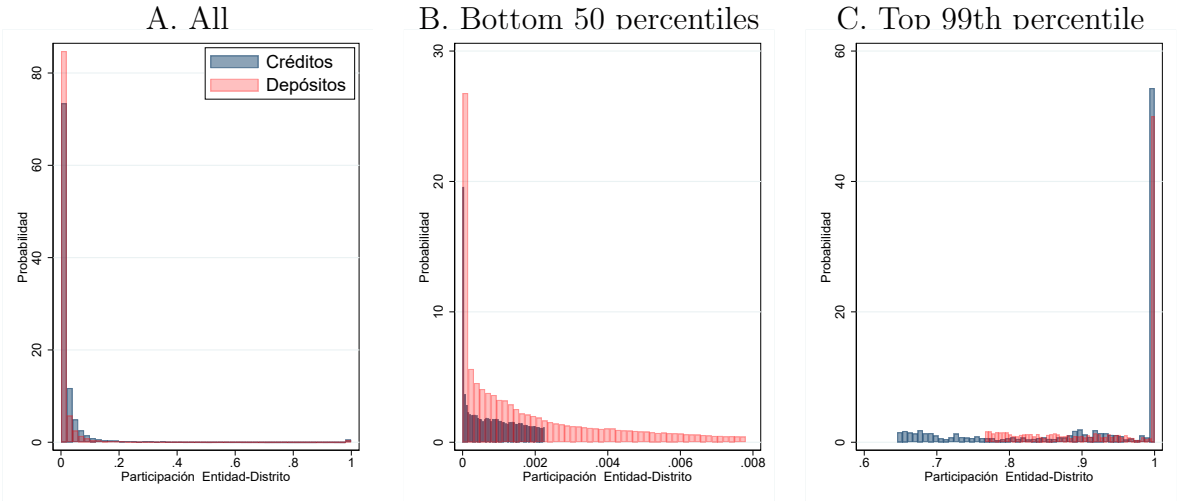


Note: The figure shows the geographic distribution of districts' position as net borrowers or net lenders as measured by the district-level geographic imbalance index $GI_{S_L-S_D}$, which is computed as the difference between the share of loans and their share of deposits, both of the national total. For each year (2002, 2008, 2014, or 2019) it shows the distribution intervals divided by the 10th, 50th, and 90th percentiles of $GI_{S_L-S_D}$. From lightest red to darkest red, these boundaries form four groups of districts: i) Districts that are in the bottom 10 percentiles, with very low loan rates relative to deposit rates, colorboxmyredlightii) districts between the 10th and 50th percentiles, with a low share of loans relative to share of deposits; iii) districts in the 50th to 90th percentiles, with loan percentages above their deposit percentages and iv) districts in the top 10 percentiles, with loan percentages very high in relation to deposit percentages.

To capture the impact on the behavior of the financial system of this regional imbalances between loans and deposits, using monthly data from all financial institutions from the 2002:M1-2019:M12 period figure 4 show that the distributions of deposit and loan shares in each district in a financial institution have wide tails. The credit and deposit

shares are calculated as the ratio of the deposits or loans of an entity-district over the amount to the total loans or deposits of the financial institution in a given month-year.

Figura 4: *Distribution of the credit and deposit shares of financial institutions in a district*



Note: This graph shows the distribution of shares of deposits and loans of each financial entity in a district. The panel on the left shows the complete distribution. The middle panel widens the distribution of shares of the bottom 50 percentiles. The panel on the right zooms in on the top 99th percentile of shares. The participation of credits and deposits is calculated as the ratio of the amount of deposits or loans of an entity-district over the total amount of loans or deposits of the financial institution in a given month-year. The shares are shown grouped for all banks and months-years.

Panel A shows a high degree of inequality between the credit and deposit share distributions. The loan share distribution has a wider tail. Panel B and C confirm this high inequality. 50% of the district branches of financial entities have shares below 1% (0.2% for credit shares, and 0.8% for deposit shares). However, 1% of bank branches hold shares of more than 65% and 77% of total deposits and loans, respectively. Figure 8 in Appendix A shows that the shape of these distributions has not changed much over the years.

To account for this heterogeneity, we go further and compute an index that captures the imbalance of deposits and loans at the bank level, as in the sociology literature (Duncan and Duncan, 1955; White, 1983):

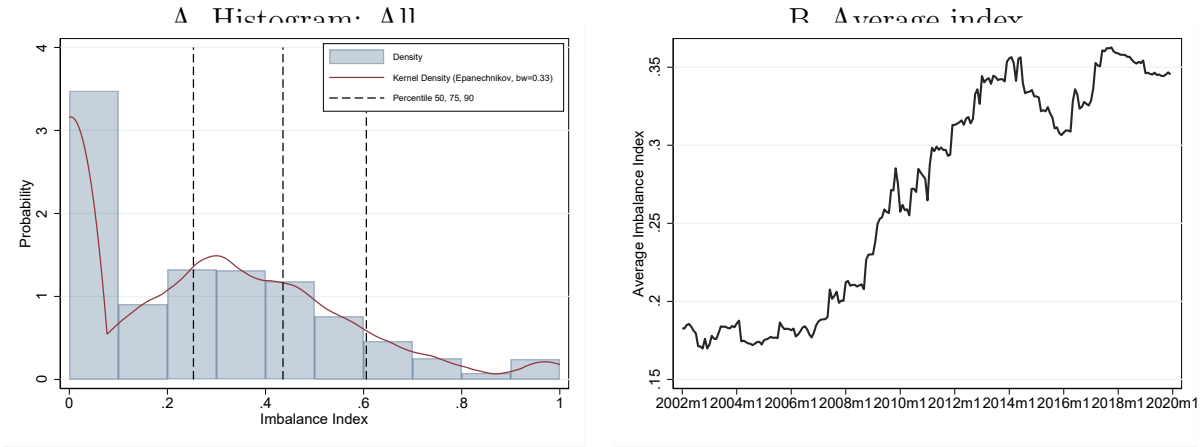
$$GI_{bt} = \frac{1}{2}\sum_{d=1}^{M}\left|\frac{dep_{dtb}}{Dep_{bt}} - \frac{cre_{dbt}}{Cre_{bt}}\right| \tag{3}$$

where dep_{dbt} and cre_{dbt} represent the amount of deposits and loans, respectively, in district d and month t for a bank b, and Cre^{d}_{bt} and Dep_{bt} represent the total amounts of the bank's deposits and loans. This ratio is a measure of a bank's deposit and loan imbalance or, alternatively, a measure of the bank's internal bias. For example, an index value of zero

represents an extreme case of perfect internal balance at the district level, that is, the geographic distributions of bank loans and deposits are identical. At the other extreme, a ratio equal to one means that a bank takes all its deposits in markets where it does not make loans, and lends only in markets where it does not make deposits, which would be an extreme case of geographic segmentation of loans.

Figure 5 in Panel A shows that a significant number of financial entities have a strong local district bias, and they allocate loans mainly funded by the deposits they obtain in these same districts. In each period there is a significant mass of financial entities with an index close to zero. These are entities that operate exclusively in very few districts. In addition, most financial institutions transfer funds across geographic locations. At the other extreme, we find some entities with very high imbalance indices. In fact, the index is greater than 0.5 for more than 20% of financial entities. Figure 9 in Appendix A shows that over time this distribution has moved further to the right, with an increase in the mass of financial entities with a higher index. Panel B of Figure 5 precisely shows that since 2008 more financial entities have become involved in moving deposits and credits through geographic locations. But, as of 2014, these dynamics have stopped, and the average index has remained relatively constant. Figure 10, in Appendix A, shows that these trends are maintained or even more pronounced if we *empresas financieras*.

Figura 5: *Distribution of the imbalance index between credits and deposits at the level of financial entities*



Note: Panel A shows the bank-level histogram of the credit and deposit imbalance index, measured by the GI_{bt} specification in the equation 3, for all banks and all sample periods. 2002:M1-2019:M12. Panel B shows for each period the average imbalance index of all financial entities.

5 Leverage and imbalance channel at the bank level

In this section we seek to characterize the heterogeneous effect of monetary policy at the bank level, and that it depends on the financing structure, measured by leverage and the geographic imbalance index. The first measure considers the overall balance sheet

structure of a financial institution, and therefore characterizes its external financing needs. The second measure measures the structure of the financing network and internal loans of a bank.

We empirically estimate the following model at the bank level for the 2003:M1-2019:M12 period using monthly frequency of the monetary policy shock and Ordinary Least Squares (OLS):

$$\Delta y_{bt} = \alpha + \beta_1 RWA_{bt-1}\epsilon_t^{MP} + \beta_2 GI_{bt-1}\epsilon_t^{MP} + \beta_3 \epsilon_t^{MP} + \beta_4 RWA_{bt-1} + \beta_5 GI_{bt-1} + \beta_6 X_{bt} + \epsilon_{bt}, \tag{4}$$

where the subscript b refers to a bank, the subscript t refers to month, Δy_{bt} is the monthly credit growth and ϵ_t^{MP} is the monetary policy shock. If $\epsilon_t^{MP} > 0$, there is an unexpected contractionary policy; while if it is negative, the opposite holds.

The specification considers the interaction of the monetary policy shock with the level of leverage (RWA) and the geographic imbalance index (GI). These variables enter the specification with a lag to reduce endogeneity problems. The purpose of these variables is to measure the monetary policy impact conditional on the financing structure of bank loans.

Likewise, controls are incorporated to isolate unobservable bank characteristics that are invariant over time (bank fixed effects). Similarly, time-fixed effects can be incorporated. The latter allows controlling for fluctuations, for example, in the exchange rate, and the aggregate economy. Similarly, we control for some time-varying bank characteristics. These control variables are grouped in X_{bt} and include the return on assets (ROA), the non-performing loan ratio (NPL) and the bank's market share (SIZE). Finally, ϵ_{bt} is a random error that has a normal distribution.

By interacting the geographic imbalance index with the monetary policy shock, we seek to capture whether the impact of monetary policy may depend on the imbalance level, regardless of whether it is due to a greater or lesser relative presence of credits or deposits in a district.

Consistent with the results in aggregate analysis in section 2 and with the theoretical model that we discuss in section 7, it is expected that a contractionary monetary policy (a positive realization of ϵ_t^{MP}) also a negative impact on credit at the bank level. That is, we expect $\beta_1\overline{RWA} + \beta_2\overline{GI} + \beta_3 < 0$.[11]. Also, in line with the theoretical model's results, due to the excessive risk-taking channel, it is expected that the leverage level is associated with a less effective monetary policy. That is, the positive impact on credit of policy rate reduction is less the higher the leverage (or the less dependence on external funding), i.e., it can be expected to find $\beta_1 > 0$.

Table 3 reports the results of the estimations with Ordinary Least Squares, for differ-

[11]The upper bar indicates the average across both time t and banks b.

ent specifications. In columns 1-3 the geographic imbalance index is not yet considered, while in columns (4)-(6) we do. Likewise, estimates without fixed effects are shown, columns (1) and (4), with bank fixed effects, columns (2) and (5), and time-fixed effects, columns (3) and (6).

According to Table 3 the monetary policy shock is statistically significant. In addition, considering the averages of 6.8 and 0.27 of the leverage (RWA) and the demographic imbalance index (GI) respectively, see table 1, we find that the shock has the expected direction. This is $\beta_1 \overline{RWA} + \beta_2 \overline{GI} + \beta_3 < 0$.

The novel result is to find that the interaction coefficient of the monetary policy shock with the lagged leverage is statistically significant and has the expected sign, in all the specifications, columns (1)-(6). This might suggest that the excessive risk-taking channel is at least statistically significant and might dominate other leverage-related channels. The coefficient can take values from 0.012 to 0.017. These results hold if we also add as control the interaction of the monetary policy shock with the lagged geographic imbalance index, which controls for the internal structure of the banks' deposit and loan network. Interestingly, if we do not exclude the extreme values of credit growth, this coefficient is still statistically significant, but quantitatively more important, so it could take values between 0.032 and 0.041 (see table 9 in Appendix B).

As a robustness exercise, table 4 shows the results including controls for some bank characteristics that vary over time, such as the NPL ratio, return on assets, and market share. Qualitatively, the results hold, although quantitatively it is found that the coefficients associated with the monetary policy impact and its interaction with leverage are larger. This time this interaction coefficient takes values between 0.020 and 0.025. Considering for example the interaction coefficient of 0.02 (see column 4) and the average leverage of 6.8, a negative monetary policy shock of 100 bps, without considering the effects of the structure of the loan and deposit network captured by the geographic imbalance index, implies an increase in credit growth of 2.5 bps. Note that if we do not consider the differential impact of leverage (i.e., leverage is zero), credit growth would be 16.3 bps. This indicates that the leverage level is quite important in determining the monetary policy power in the credit market.

To measure the importance of monetary policy heterogeneity induced by the leverage level, we apply the same change in the monetary policy shock to different values of the leverage ratio: one equal to one standard deviation below the mean (5.1) and the other to one standard deviation above the mean (8.5). This comparison in the distribution of leverage ratios shows increases in credit growth of 5.9 bps and -0.9 bps, respectively. Holding other characteristics constant, banks with less leverage (or with greater dependence on external financing) increase their credit to a greater extent in response to expansionary monetary policy shocks. An entity with a level of leverage with one standard deviation

below the average increases credit by 3.4 bps more than an average bank. In contrast, banks with very high levels of leverage, one deviation above the average, increase their credit growth by 3.4 bps less than an average bank. These results are consistent with the theoretical model that we present in section 7.

If we consider the coefficients associated with the effects of the structure of the loan and deposit network captured by the geographic imbalance index, these are not statistically significant, and indicate that the internal financing structure of a financial institution would not affect the impact of monetary policy. However, if we consider columns (4) and (6) of Table 4 that does not control for the constant heterogeneity between banks, it shows that banks that tend to mobilize more funds between districts, or with a broader and more heterogeneous branch network (a higher imbalance index), amplify the impact of monetary policy. Although these impacts are not significant, since they probably converge with the leverage, in addition to the fact that the index does not distinguish whether the imbalance is due to a greater or lesser participation of credit and/or deposits in a district. Although at the bank level the imbalance between credits and deposits may not be significant, as we will see later, at the branch level, we will have a leverage indicator. Implicitly, it will indicate that greater or smaller participation of deposits in the funding structure will have different implications on the impact of monetary policy.

Although these results show that a monetary policy shock has the expected effect on the credit market at the bank level, these are limited. A monetary policy shock has an impact on both sides of the credit market, and we want to study the lending channel that only considers supply effects. But, in the previous specifications that we present, we cannot separate the supply effects from those of demand. Therefore, from now on, we use an identification strategy at the bank branch level to obtain a clearer signal of the heterogeneous effect caused by the leverage level when it interacts with the monetary policy shock.

Tabla 3: *Regression at bank level, monthly frequency, estimated with OLS*

	Dependent variable: Δy_{bt}					
	(1)	(2)	(3)	(4)	(5)	(6)
$GI_{bt-1}\epsilon_t^{MP}$				0.00233	0.0235	-0.00638
$RWA_{bt-1}\epsilon_t^{MP}$	0.0128*	0.0156**	0.0167**	0.0127*	0.0147**	0.0169**
ϵ_t^{MP}	-0.106**	-0.128***		-0.105**	-0.129***	
RWA_{bt-1}	-0.00177**	-0.00653***	-0.00365***	-0.00201***	-0.00610***	-0.00387***
GI_{bt-1}				0.00674	-0.0467**	0.00596
Obs.	2,680	2,680	2,680	2,680	2,680	2,680
R^2	0.005	0.086	0.122	0.005	0.088	0.122
F test (ρ-value)	0.00664	7.48e-09	3.28e-07	0.0192	6.62e-09	3.23e-06
Bank FE		✓			✓	
Time FE			✓			✓

*p<0.1, **p<0.05, ***p<0.01.

Tabla 4: *Bank level regression with additional controls, monthly frequency, estimated with OLS*

	Dependent variable: Δy_{bt}					
	(1)	(2)	(3)	(4)	(5)	(6)
$GI_{bt-1}\epsilon_t^{MP}$				-0.00903	0.00577	-0.0209
$RWA_{bt-1}\epsilon_t^{MP}$	0.0201***	0.0237***	0.0243***	0.0203***	0.0235***	0.0248***
ϵ_t^{MP}	-0.164***	-0.193***		-0.163***	-0.193***	
RWA_{bt-1}	0.00323***	0.00151	0.00246***	0.00266***	0.00168	0.00194**
GI_{bt-1}				0.0164***	-0.0187	0.0153**
NPL_{bt}	-0.00158***	-0.00215***	-0.00118***	-0.00206***	-0.00203***	-0.00170***
ROA_{bt}	0.000681	-0.000593	-1.50e-05	0.000162	-0.000351	-0.000467
$SIZE_{bt}$	-8.11e-05	0.00322**	-6.74e-06	-7.92e-05	0.00330**	-1.13e-05
Obs.	2,642	2,642	2,642	2,642	2,642	2,642
R^2	0.025	0.087	0.121	0.027	0.087	0.124
F test (ρ-value)	0	4.81e-10	8.66e-07	0	2.85e-09	4.49e-07
Bank FE		✓			✓	
Time FE			✓			✓

*p<0.1, **p<0.05, ***p<0.01.

6 Transmission of Monetary Policy at Bank Branch Level: The Imbalances Channel

6.1 Identification Strategy

We exploit the geographic variation in the credit-to-deposit ratio induced by differences in the concentration of local credit and deposit markets. At the branch level of a

financial institution, the credit-to-deposit ratio is an indicator of leverage or the need for external financing. This is, the higher the credit-to-deposit ratio, the greater the need for external financing. As we will show later, this ratio allows us to capture the heterogeneous impact of monetary policy induced by differences in leverage ratios or external financing needs. An important assumption in our analysis is that credit markets are local. In general, the process of granting loans is an information-intensive activity, where a financial entity must filter among potential borrowers, due to information asymmetries. In general, geographic proximity reduces the costs of transmitting and processing information (Nguyen, 2019). Therefore, the relationships between clients and local financial entities become crucial.

We use the credit-to-deposit ratio as a measure of the composition of balance sheet liabilities or leverage within a bank branch (Agénor, Alper and Pereira, 2018). The key idea is that this ratio, which captures the bank branch leverage, is important for bank lending strategies (Gertler and Karadi, 2011). Although banks can internally allocate funds between branches, this is not a costless process between districts. In particular, we rely on the administrative cost assumption to allocate resources across the district, which implies that it is preferable for a bank branch to collect local deposits to fund loans. If a bank branch has a high credit-to-deposit ratio, financing additional liquidity on top of deposits is expensive. The increase in the funding cost comes from changes in the allocation of assets and liabilities within a bank and costly liquidity transactions in the intrabank and/or interbank markets. This is a credit supply channel induced by the imbalance structure between credits and deposits within a bank.

To identify this channel of imbalances, it is important to have a variation in the composition of liabilities that is independent of the bank's lending opportunities or the demand factors that influence the bank's decisions. To obtain that variation, we compare the loans between branches of the same bank located in different districts. Under this strategy, we can control for loan demand opportunities that a bank faces, and we can identify the sensitivity induced by monetary policy changes on the credit supply strategy of a bank branch.

6.2 Reduced form estimation: Sensitivity of loans of a bank branch

Before any assumptions of causal identification, we first characterize the behavior of the credit change of all bank branches. In particular, we run the following time series regression for each branch j:

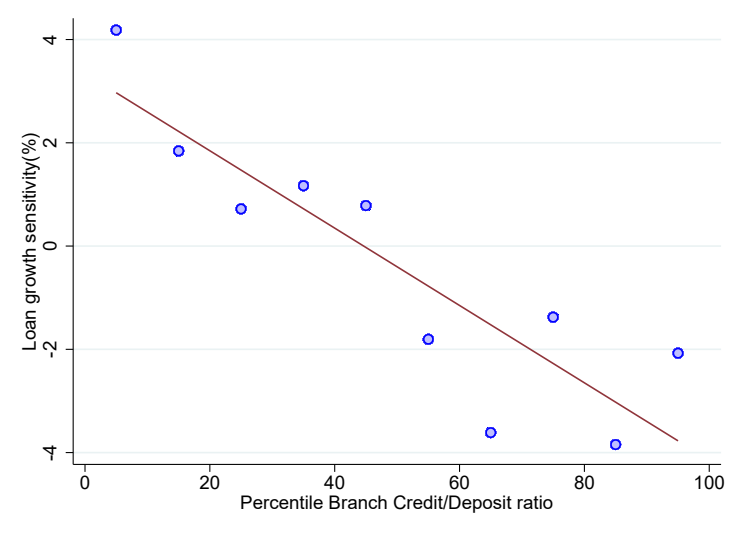
$$\Delta y_{jt} = \alpha_j + \beta_j \Delta i_t + \epsilon_{jt}\,, \tag{5}$$

where Δy_{jt} is the absolute deviation of the log of credit supplied by branch j at time t, Δi_t is the contemporaneous variation of the interbank interest rate, which is our measure of the monetary policy rate, β_j measures the sensitivity of the credit of branch j to changes in monetary policy. This is a branch credit beta.[12].

Next, we relate the credit beta, β_j, to the district's credit-to-deposit ratio. The latter measures the composition of external financing observed in a given district. We average the credit betas of all branches within a district, winsorizing at 1% to minimize the influence of outliers. We then classify all the branches into 10 equally sized partitions, according to the external financing composition (credit-to-deposit ratio)

Figure 6 shows that the sensitivity of credit growth is negatively related to the credit-to-deposit ratio. This is indicative of heterogeneity in monetary policy transmission depending on the relative size of external funding at the branch level and working in the intended direction from a given threshold.

Figura 6: *Credit growth sensitivity and the credit-to-deposit ratio*



Note: This figure shows the relationship between the credit-to-deposit ratio and the sensitivity of credit growth to changes in the monetary policy rate. This figure is built in two steps. In the first step, we estimate a time series regression for each branch $\Delta y_{jt} = \alpha_j + \beta_j \Delta i_t + \epsilon_{jt}$, where Δy_{jt} is credit growth and Δi_t is the policy rate change. In the second step, we average the credit betas of all branches within a district, winsorizing at 1 percent. We then rank all branches into 10 equal-sized bins by external funding composition and calculate the bin averages. Sample period 2003:M1-2019:M12. This includes all branches of banks, *cajas* and other loan-issuing branches.

On the one hand, the sensitivities to credit growth of branches that have a credit-to-deposit ratio above the 50th percentile become more negative and, therefore, amplify the impact of a change in the monetary policy stance on aggregate credit growth. On average, the credit beta is -2.75 in this group.

[12]Following the same spirit of the capital asset pricing model (CAPM) literature, this beta measures whether credit flows (credit change) at a given branch j they move in the same or opposite direction as the aggregate market, which is led by changes in the monetary policy rate.

On the other hand, those branches that have a loan-to-deposit ratio below the 50th percentile have positive or near-zero beta credit factors and, for example, when faced with a tight monetary policy stance, make more loans, rather than to reduce loans. These branches mitigate any impact of monetary policy on the credit market. These branches have an average credit beta factor of 1.74 percent.

6.3 Branch level estimation

Equation 6 shows the main specification at the branch level in a specific district that implements our identification strategy. The dependent variable, Δy_{bdt}, is the growth rate of loans granted by the branches of the bank b in district d at time t. $\left(\frac{Cred}{Dep}\right)_{bd,t-1}$ is the credit-to-deposit ratio and is our measure of funding frictions at the branch level. $\mathbb{1}_{\frac{Cred}{Dep}}$ is an indicator variable that takes 1 if the credit-to-deposit ratio is greater than the district median. This threshold captures our previous analysis of the sensitivity of monetary policy at the branch level. With this variable we intend to capture the heterogeneity in the credit response between banks due to supply frictions. ϵ_t^{MP} is the monetary policy shock. All these regressors (except the policy shock) enter the regression with a delay to control for endogeneity issues. We also include fixed effects (FE) that control for trends at the branch and local market factor level, at the region or district level, represented by α_{bt} and α_{rt} or α_{dt}. We also include district-bank fixed effect, α_{bd}, to capture unobserved time-invariant district-bank relationships.

$$\begin{aligned}\Delta y_{bdt} = \alpha_{bt} + \alpha_{rt} + \alpha_{dt} + \beta_0 \left(\frac{Cred}{Dep}\right)_{bdt-1} \epsilon_t^{MP} \\ + \beta_1 \left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbb{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP} + \gamma_x X_{bdt-1} + \epsilon_{bdt}\end{aligned} \tag{6}$$

The bank-time fixed effects, α_{bt}, are key in our identification as it absorbs all time differences between banks, and allow us to control for the lending opportunities that a particular bank faces. So, we compare between branches of the same bank. $\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbb{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP}$ captures the channel of imbalances and/or external financing of the monetary policy at the bank branch level. Consequently, if $\beta_1 < 0$, after an expansionary monetary policy shock, the branches that operate with a high credit-to-deposit ratio grant more loans than branches with lower external financing needs.

To control for initial credit conditions or observation-specific variance, we add a lag of all loans made by a branch, y_{bdt-1}. We also test different sets of controls in X_{bdt-1}, such as the lags of the dependent variables and the lags of the monetary policy shock that interact with the credit-to-deposit ratio at the branch level, to control for the mean reversion property of the endogenous variable at the branch level and the shock diffusion, respectively. We also include interactions and intermediate components to avoid

misspecification issues, such as $\left(\frac{Cred}{Dep}\right)_{bdt-1}$, $\mathbb{1}_{\frac{Cred}{Dep}}$, $\left(\frac{Cred}{Rip}\right)_{bdt-1} \times \mathbb{1}_{\frac{Cred}{Dep}}$.

Results

The table 5 shows the results of the estimation of the equation (6), which formally implements our identification strategy. The sample includes all branches of all financial, banking and non-banking companies. The quarterly sample covers the 2003:Q1-2019:Q4 period. Credit data was winsorized at 2% to control for biased results due to outliers. In general, we intend to control for variables that could influence the lender's decision and that may also be correlated with finance company ownership, financial institution size, regional economic conditions, management decisions, or initial trends of the branches.

Columns (1)-(4) show various specifications for adding or removing controls. Column (1) is our preferred specification, as it includes fewer restrictions on the data. Column (1) shows that the monetary policy channel of leverage at the bank branch level is statistically and economically significant. It confirms that loan sensitivity to monetary policy changes is non-linear and increases as the credit-to-deposit ratio of a bank branch is higher, even within a financial company. Numerically, after a 100 bps expansionary monetary policy shock, financial institutions with a loan-to-deposit (C/D) ratio below its median increase on average loan growth rates by only 0.29 bps. Meanwhile, for the group of branches with C/D ratios above its median, after the same policy shock, loan growth rates increase on average by 273 bps more. For an aggregate perspective, this latter response is numerically higher; however, it might be explained since this average effect is not weighted by the relevance of branch credit or deposits.

Similar to the bank level analysis (section 5), to measure the importance of monetary policy heterogeneity induced by the C/D ratio, we apply the same monetary policy shock to a different value of the leverage ratio but to only the group of branches with a C/D ratio above its mean: one standard deviation above the leverage mean (72.35). It results in an increase of credit growth of 955 bps by more.

Column (2) omits region-time fixed effects, but considers district-time fixed effects, so identification occurs in a smaller geographic space, by comparing only bank branches that operate within the same district. Column (3) includes longer lags for monetary policy shocks and the dependent variable to account for more persistent mean reversion and propagation effects. Column (4) is like the specification in column (3) but omits bank-time fixed effects. The leverage channel coefficients in all these columns remain similar to those in column (1). These results confirm that the sensitivity of loans to changes in the monetary policy rate increases with the loan-to-deposit ratio, but they are significant and more important for branches with very high loan-to-deposit imbalances. These results are also complementary to the monetary policy risk channel documented

in Pozo and Rojas (2020a, 2022). In general, as can be seen in Table 2, branches with a high C/D ratio are those that tend to display similar characteristics to branches that operate in riskier markets.

Tabla 5: *Estimation results at the level of banking branches*

	Dependent variable: Δy_{bdt}			
	(1)	(2)	(3)	(4)
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \epsilon_t^{MP}$	-0.00354	0.0686	0.119	-0.0585
	(0.111)	(0.145)	(0.216)	(0.135)
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbb{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP}$	-0.132*	-0.174*	-0.238*	-0.306**
	(0.0777)	(0.0893)	(0.145)	(0.146)
y_{bdt-1}	-0.588***	-0.693***	-0.0275	-0.769
	(0.0763)	(0.0949)	(0.208)	(0.624)
R^2	0.642	0.676	0.581	0.508
Obs.	86093	81974	58220	58544
Controls	✓	✓	✓	✓
Control Long lags			✓	✓
Bank-Time FE	✓	✓	✓	
Region-Time FE	✓		✓	
Distric-Time FE		✓		
Branch FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓
Time FE				✓

Note: It shows the estimated effects of Peruvian monetary policy shocks on credit growth at the branch level. The dependent variable is the log change in credit at the branch level, Δy_{bdt}. Loan growth is the logarithmic change in credit at the branch level. Quarterly Sample: 2004:Q1 to 2019:Q12. The sample includes only banks with branches in two or more provinces. Cred/Dep measures the degree of resource dependency of a bank branch. ϵ_t^{MP} is the monetary policy shock. Fixed effects are described at the bottom of the table. Standard errors clustered at time and bank level in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Robustness

Table 6 shows that the result of our base specification is robust to sample selection problems or omitted variables, such as bank concentration or market risk, which are documented in Pozo and Rojas (2021a), and that can bias the results. The effect of the leverage channel or financing imbalance continues to be economically and statistically significant in all specifications.

Tabla 6: *Branch level estimation: Robustness I*

	Dependent variable : Δy_{bdt}				
	(1) After 2006Q1	(2) Before GCF	(3) After GCF	(4) Control HH, Mor	(5) Ctrl. HH, Mor + Cluster
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \epsilon_t^{MP}$	-0.0162	-1.848**	-0.147	0.0123	0.0672
	(0.108)	(0.865)	(0.110)	(0.113)	(0.185)
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbb{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP}$	-0.144*	0.0538	-0.222***	-0.139*	-0.198**
	(0.0755)	(0.528)	(0.0760)	(0.0777)	(0.0862)
$\text{HH}_p \times \epsilon_t^{MP}$				-2.935	-3.501
				(2.559)	(3.893)
$\text{NPL}_p \times \epsilon_t^{MP}$				-0.0272	-0.0611
				(0.0435)	(0.0545)
y_{bdt-1}	-0.646***	-5.093***	-0.660***	-0.588***	-0.231
	(0.0794)	(1.307)	(0.0872)	(0.0762)	(0.259)
R^2	0.661	0.698	0.674	0.639	0.526
Obs.	78918	6696	65358	85881	87963
Control lags	✓	✓	✓	✓	✓
Bank-Tiempo FE	✓	✓	✓	✓	✓
Region-Time FE	✓	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓

Note: This table estimates the effect of Peruvian monetary policy shocks on credit growth, Δy_{bdt}. Loan growth is the logarithmic change in credit at the branch level. Quarterly Sample: 2004:Q1 to 2019:Q12. The sample includes only banks with branches in two or more provinces. ϵ_t^{MP} is the monetary policy shock. Fixed effects are described at the bottom of the table. Cred/Dep measures the degree of resource dependency of a bank branch. After the first quarter of 2006, the sample removes the observations from 2004 to 2005. Before the GCF (Great Financial Crisis) sample, the sample corresponds to the period from the first quarter of 2004 to the fourth quarter of 2008. The sample after the GCF corresponds to the period 2009Q4-2019Q4. NPL measures the market risk in the province where a branch is located (see Pozo and Rojas (2021a)). HH is the standard Herfindahl index of the province where a branch is located. HH and NPL are obtained by averaging all the years for each province (p). In Cluster, standard errors are clustered at the time and bank level. *p<0.1, **p<0.05, ***p<0.01.

Column (1) removes observations for the period before 2006. The early years of our sample have measurement problems and are underrepresented in aggregate credit dynamics. The result does not show a significant difference from our base estimate.

In columns (2) and (3) we try to control for the effects of variations in the amount of international liquidity, low international interest rates and the expansive stance of monetary policy. In particular, before the Great Financial Crisis (GCF) the monetary policy rates, both local and international, were not low. Likewise, in this previous period the rates of economic growth in Peru were also higher. After the GFC, with an expansionary monetary policy stance by almost all the central banks of developed countries,

international liquidity has been high and interest rates have been lower. Consistently, the monetary policy rates also in Peru, on average, have been lower during this period. In this context, the behavior of the level of financing imbalance could be the result of international macroeconomic trends. In column (2) we only consider the sample period before the Great Financial Crisis (GFC). While in column (3) we restrict the sample to the period after the GFC, starting in 2019:Q4. Results holds for the sample period after the GCF, but no for the sample period before the GFC. After the GCF, this channel has been concentrated in those branches with higher levels of leverage, above the median.

Column (4) controls for other factors of the local markets where the financial branches operate. Pozo and Rojas (2021a,b) show that the degree of banking concentration or competition in the credit market or market risk influence the credit supply policies of financial institutions. For this purpose we use similar measures as in Pozo and Rojas (2021a): we use the standard Herfindahl (HH) index at the province level, to measure the degree of market concentration, and as a measure of risk at the local financial market level we use the non-performing loan (NPL) ratio at the province level as in Pozo and Rojas (2021b). HH and NPL ratio are obtained by averaging these measures for all years and for each province. These market characteristics could be part of our results, since more leveraged branches could operate in markets where they take more risk and where there is more concentration. The results in columns (4) and (5) show that the degree of leverage at the level of financial institution branch is still significant and very similar to our baseline estimate, even after controlling for these local market factors, column (4), or by controlling for the correlation of loans at the bank level and time, column (5).

Table 7 shows the robustness of the main specification in the equation (6) for different samples of financial institutions and geographic areas. In particular, we show evidence of the heterogeneity of the imbalance or leverage channel at the branch level for different types of financial institutions and local credit markets. Column (1) considers the branches of all financial institutions, column (2) restricts the sample to bank branches, column (3) considers branches of the four largest banks in Peru, and column (4) considers branches of non-banking institutions. Panel A reports the results for branches from all Peruvian local markets; while Panel B shows the results excluding branches operating in the metropolitan area, which accounts for most of the credit in the country.[13]

Panel A of table 7 shows that the base specification results hold in the same direction for the sample of banks, but with a doubling of sensitivity. The impacts are economically and statistically significant at the top of the distribution of C/D ratios. This differentiated impact of the monetary policy effect conditional on the size of the external leverage dependency, or high C/D ratio, between branches is even much higher for the sample of the largest banks. For the four largest banks in Peru, which have a more extensive

[13] Pozo and Rojas (2021b) show that branches in the metropolitan area (regions of Lima and Callao) represent around 75% of total credit, and is mainly concentrated in credit to corporate and large firms.

geographic network of branches, covering almost all of the country's provinces, this sensitivity is almost 7 times. In this sense, the largest banks are sensitive to changes in monetary policy, via the imbalance channel, given that they have to mobilize a greater volume of funds to cover the various credit financing needs of their different branches.

Panel A of table 7 also indicates that for the sample of non-banks this imbalance channel is not present. However, Panel B, which analyzes more closely the sample of financial institutions and the difference between geographic areas that do not include the metropolitan area, shows that the channel of monetary policy imbalances is weak for the sample of banks, but significant and statistically significant. significant for the sample of non-banks. The metropolitan area represents approximately 75% of the credit in the sample and is mainly concentrated in credit for safer borrowers, such as corporations and large companies (Pozo and Rojas, 2021b) where banks tend to have more participation and whose branches tend to be more unbalanced. However, non-bank entities that are concentrated in geographical areas outside the metropolitan area tend to be less unbalanced and therefore have low credit-to-deposit ratios. In these bank branches, the imbalance channel is located in the lower part of the distribution of the need for leverage ratio.

Tabla 7: *Branch level estimation: Robustness II*

	Dependent variable: Δy_{bdt}			
	(1) All	(2) Banks	(3) Large banks	(4) No banks
Panel A: Full sample				
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \epsilon_t^{MP}$	-0.00354	0.0456	0.0976	-0.133
	(0.111)	(0.172)	(0.265)	(0.0814)
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbb{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP}$	-0.132*	-0.312**	-0.910*	0.0186
	(0.0777)	(0.152)	(0.521)	(0.0406)
y_{bdt-1}	-0.588***	-0.441***	-0.0492	-1.067***
	(0.0763)	(0.133)	(0.335)	(0.0503)
Observations	68969	30965	17703	34244
R^2	0.346	0.342	0.311	0.427
Panel B: Non-Metropolitan Area sample				
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \epsilon_t^{MP}$	-0.113	0.0295	0.0620	-0.522**
	(0.101)	(0.168)	(0.172)	(0.256)
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbb{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP}$	-0.0456	-0.0678	0.520	0.0375
	(0.0665)	(0.118)	(1.087)	(0.0490)
y_{bdt-1}	-0.733***	-0.761**	-0.433	-0.940***
	(0.0723)	(0.277)	(0.406)	(0.217)
Observations	59115	26457	14722	32514
R^2	0.599	0.601	0.606	0.618
Controls	✓	✓	✓	✓
Bank-Time FE	✓	✓	✓	✓
Region-Time FE	✓	✓	✓	✓
Branch FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓

Note: It shows the estimated effects of Peruvian monetary policy shocks on credit growth at the branch level. The dependent variable is the log change in credit at the branch level, Δy_{bdt}. Quarterly Sample: 2004Q1 to 2019Q12. The sample includes only banks with branches in two or more provinces. Cred/Dep measures the degree of resource dependency of a bank branch. ϵ_t^{MP} is the monetary policy shock. Fixed effects are described at the bottom of the table. The large bank sample considers only the four largest banks in Peru. Non-banking includes CAMCs, CRACs and financial companies. Non-Metropolitan Area Sample: excludes Lima and Callao, which form the largest metropolitan area in Peru. Standard errors clustered at time and bank level in parentheses. *p<0.1, **p<0.05, ***p<0.01.

These findings support the robustness of the imbalance or leverage channel at the branch level, and demonstrate that it operates across all financial institutions but across various sections of the C/D ratio distribution. The effects of monetary policy are precisely stronger and more significant in those branches that have a greater need for financing,

and in particular for the sample of banks, and to an even greater extent for the largest banks.

However, the local markets where financial institutions operate matter. The financial imbalance channel operates in non-bank entities only for branches that are located in the lower half of the C/D ratio distribution and outside the metropolitan area.

6.4 Discussion

In principle, these results, both in a reduced form (section 6.2) and in the causal estimation (section 6.3) show that there is a heterogeneous banking channel in Peru, where the Monetary policy actions have an amplifying and a dampening effect on the credit market. The importance of these effects depends, at a granular level, on whether a bank branch is more or less dependent on external funding, defined as others than local deposits.

However, at the aggregate credit response (section 2), the amplifying margin for credit-to-deposit ratio heterogeneity must be stronger than the attenuating margin. This is also consistent with the results of the empirical model at the bank level (section 5). Also, at the bank level, we find that the impact of monetary policy may be conditional on the level of leverage in the same direction found in the branch analysis.

Note that our results are based on the fact that there are heterogeneous local credit markets across the country. Therefore, the monetary policy supply channel varies between regions, and depends on the actions of financial institutions, which depend on funding and opportunity costs. In particular, a relevant variable for local credit decisions could be the credit-to-deposit ratio. This variable could determine the heterogeneity of the monetary policy transmission mechanism toward the credit market. Thus, in principle, those branches that depend less on local deposits as a source of funding (that is, have a high credit-to-deposit ratio) may be less exposed to an increase in the interest rate of the monetary policy that is transferred to higher deposit interest rates. However, these same branches require more external funding. As a result, results suggest that the cost of this external funding increases much faster than the cost of local funding (local deposits). Therefore, the greater the dependence on external funding sources, the more positive the impact of the policy rate on the credit funding cost.

7 Theoretical model

The motivation of the theoretical model is to rationalize the heterogeneous effects of monetary policy on credit at the aggregate bank level and bank branch level. To do this, we propose a theoretical framework with specifications for each approach: aggregate and

micro.

Whereas, in the first specification, the model seeks to shed light on the response of credit to the monetary policy rate conditional to the leverage level of the financial institution (credit to capital ratio); in the second specification, it seeks to understand the impact of the monetary policy conditional to the credit financing composition of a bank branch. Therefore, the specifications can complement each other, since the first seeks to help understand what happens across heterogeneous banks, while the second would help us understand what happens across heterogeneous branches of the same bank.

For both cases we develop a two-period banking model, where banks or banking branches grant loans to firms, which invest in a risky project. Banks or branches finance their investments with local deposits and other funds. To simplify the model, it is assumed that all agents are risk neutral. In addition, we assume that banks have limited liability and deposits are insured by government, through lump-sum taxes on households.

Motivated by the research question and empirical results, the model focuses on characterizing the heterogeneity that may exist in banks regarding their level of leverage. Although this measure reflects the capacity of a bank to absorb negative shocks, it can also capture bank incentives to take risk or excessive risk.

7.1 Banks with different levels of leverage

We develop a two-period model with a continuum of heterogeneous banks, firms, and households (depositors). Both firms and households are identical. Banks are heterogeneous in their exogenous level of initial net worth (capitalization).[14] Households own banks and firms. They also make bank deposits and supply labor to firms. Banks use their net worth and deposits to finance loans. Due to the nature of the business activity, loans are risky. Firms use loans to buy capital that, in combination with labor, is used in the production of goods. All agents are risk neutral and the unit of measurement is given by the final good.

It is assumed that loan risk is systematic (aggregate). It can be understood that it is the risk of a loan portfolio after having eliminated all idiosyncratic risk. Finally, it is assumed, for simplicity, that a bank is associated with a single company and vice versa.

In period t=0, firm i buys capital, k_0^i, financed entirely by bank loans, l_0^i=k_0^i. For simplicity and because it is not the purpose of this research to study the role of firm leverage in monetary policy, we assume that firm net worth is zero.[15] Firms use capital

[14]We assume in this two-period model that equity (capital of banks) is exogenous without abstracting too much from reality, since we know that raising new capital is a long-term process. For example, if the bank will face higher capital requirements, it will mainly reduce lending rather than increase its level of equity.

[15]In the same spirit of understanding the transmission mechanism of monetary policy Pozo and Rojas (2020b) present empirical evidence for Peru on the risk-taking channel of monetary policy. In this case,

and labor, h_1, the latter demanded at t=1, for the production of goods, y_1^i, using a Cobb-Douglas technology,

$$y_1^i = Z_1(k_0^i)^{\alpha}(h_1)^{1-\alpha},$$

where $0<\alpha<1$, and Z_1 is the multiplicative aggregate productivity shock. Z_1 is assumed to have a lognormal distribution, $ln(Z_1) \sim N(\mu_z, \sigma_z^2)$. F and f are the cumulative and probability density function of Z_1, respectively. The profits of the firm i at t=1 are:

$$\Pi_1^i = (1-\delta)k_0^i + y_1^i - r_1^{l,i} l_0^i - w_1 h_1,$$

where δ <1 is the capital depreciation rate, $r_1^{l,i}$ is the loan interest rate. To capture the risk associated with the firm's activities, it is assumed that the interest rate $r_1^{l,i}$ is contingent to the realization of Z_1. Firms' demand for loans is obtained by maximizing discounted future earnings at t=0, $\mathbb{E}_0\{\beta\Pi_1^i\}$, where l_0^i=k_0^i. The first order condition for k_0^i is,

$$0 = \mathbb{E}_0\left\{\beta\left(r_1^{k,i} - r_1^{l,i}\right)\right\},$$

where $r_1^{k,i}$=$1-\delta+\alpha Z_1(k_0^i)^{\alpha-1}(h_1)^{1-\alpha}$ is the marginal productivity of capital. In the period t=1 the level of productivity is known and companies demand labor until the marginal product of labor is equal to the wage, that is, w_1=$(1-\alpha)Z_1(k_0^i)^{\alpha}(h_1)^{-\alpha}$. A non-negative condition is assumed for realized benefits, that is, $0 \leq \Pi_1^i$. This implies that the first-order condition for capital is:

$$r_1^{k,i} = r_1^{l,i}.$$

For simplicity, it is assumed that households inelastically supply a unit of labor. So, in equilibrium, the demand curve for bank loans yields,

$$r_1^{l,i} = 1 - \delta + \alpha Z_1(l_0^i)^{\alpha-1}.$$

The problem of the representative household is simple. Since it is risk neutral, the household maximizes utility, U_0=$C_0 + \beta E_0\{C_1\}$, where C_0 and C_1, respectively, denote the levels of consumption of the household at t=0 and t=1, subject to budget constraints at t=0, C_0=$Y_0 - D_0 + \Pi_0$, and at t=1 , $C_1 = r_0 D_0 + \Pi_1 + w_1 - T_1$, where β, Y_0, D_0, r_0, Π_0, Π_1, and T_1, respectively, represent the exogenous household discount factor, the exogenous initial endowment, deposits in banks, the gross rate of return on deposits, dividends from banks in the period t=0, dividends from banks at t=1, and government lump sum taxes.[16]

attention is not paid to monetary policy conditional on the level of leverage of financial institutions, but to bank incentives to offer loans between portfolios located in regions that might have different levels of risk.

[16]In equilibrium, Π_0 captures the net cash flow at t=0. In particular, at t=0 banks invest in risky

Since deposits are fully insured, these are risk-free and therefore their gross rate of return will be the same as the risk-free interest rate, i.e. $r_0 = \frac{1}{\beta}$. It follows that the supply of deposits faced by banks is perfectly elastic at the interest rate of r_0.

In the case of financial intermediaries, they are heterogeneous in their exogenous initial level of equity. A bank i finances its risky loans, l_0^i, with deposits, d_0^i, and with an exogenous initial equity, n_0^i. That is,

$$l_0^i = d_0^i + n_0^i. \tag{7}$$

As the literature suggests, the interaction of limited liability and deposit insurance leads to excessive risk taking (see, for example, Collard et al. 2017; Pozo 2019).[17]

To clearly explain this distortion, we first abstract from limited liability. The equilibrium thus leads to the socially efficient allocation. The future earnings of the bank are the difference between the income received from the loans and the payments for their deposits:

$$r_1^{l,i} l_0^i - r_0 d_0^i. \tag{8}$$

If future earnings are negative, $n_1^i < 0$, the bank transfers negative dividends to its owners (households); otherwise, it transfers positive dividends.[18] The bank optimally chooses the level of deposits to maximize $\mathbb{E}_0\{\beta n_1^i\}$, where β is the household discount factor, subject to the equation (7). The first order condition yields,

$$1 - \delta + \alpha \bar{Z} (l_0^i)^{1-\alpha} = r_0, \tag{9}$$

where $\bar{Z}{=}\mathbb{E}_0\{Z_1\}$. This result, equation (9), is not new. In equilibrium, the return on loans (marginal product of capital) is equal to the marginal cost of borrowing, r_0. Clearly, the interest rate (risk-free rate) directly affects the marginal cost of borrowing. As a result, a lower rate increases the incentives of banks to lend due to the assumption of diminishing marginal returns to capital.

When the bank faces limited liability, bank profits (dividends) cannot take negative values, that is,

$$N_1 = max\{0, r_1^{l,i} l_0^i - r_0 d_0^i\}.$$

For a given level of loans l_0^i, there will be a loan return r_1^{L*} such that bank profits will be zero,

$$0 = r_1^{l,i*} l_0^i - r_0 d_0^i, \tag{10}$$

projects the amount of K_0 (negative cash flow) and receive internal deposits, D_0 (positive cash flow). In other words, in equilibrium $\Pi_0 = -K_0 + D_0$.

[17] In both works, bank risk taking involves the volume of credit and not the type of credit. This is in line with the current literature, in particular the monetary policy literature which commonly views excessive bank risk-taking in terms of the aggregate volume of credit (see, p. eg, Borio and Zhu, 2012).

[18] Negative dividends means the bank owner needs to put up his own money to pay off all deposits

where in equilibrium,

$$r_1^{l,i*} = 1 - \delta + \alpha Z^{i*}(l_0^i)^{\alpha-1}. \tag{11}$$

This means that if $Z_1 \geq Z^{i*}$, bank i does not default; otherwise, it cannot fully repay deposits and, consequently, default. It follows that the endogenous probability of bank default is,

$$p_0^i = F(Z^{i*}),$$

where for convenience using (10) and (11), we rewrite Z^{i*} as,

$$Z^{i*} = \left[r_0 \frac{d_0^i}{l_0^i} - (1-\delta) \right] \frac{1}{\alpha} (l_0^i)^{1-\alpha}. \tag{12}$$

Since we are interested in the case that the probability of bankruptcy is positive, we parameterize the model such that in equilibrium $Z^{i*} > 0$ holds. From (12) it is easy to verify that, *ceteris paribus*, bank leverage increases bank default probability. That is, the greater d_0^i/l_0^i, the greater $Z^{i,*}$. Also, *ceteris paribus*, the higher the risk-free interest rate, the higher the probability of bank default. This time the expected present value of future earnings is,

$$V_0 = \mathbb{E}_0 \left\{ \beta \left(max\{0, r_1^{l,i} l_0^i - r_0 d_0^i\} \right) \right\}. \tag{13}$$

With limited liability the bank only cares about the states of nature where its income is greater than its liabilities. Since the returns on bank deposits are insensitive to risk due to deposit insurance, the bank does not internalize the effects of its decision on the risk-taking. The bank seeks to maximize (13) subject to (7). For convenience we rewrite (13) as,[19]

$$V_0 = \int_{r_1^{l,i*}}^{+\infty} \beta (r_1^{l,i} l_0^i - r_0 d_0^i) dF^{r^l}(r_1^l), \tag{14}$$

The first-order condition for d_0^i yields,

$$\int_{r_1^{l,i*}}^{+\infty} \beta (r_1^{l,i} - r_0) dF^{r^l}(r_1^l) + \beta (r_1^{l,i*} l_0^i - r_0 d_0^i) F^{r^l}(r_1^{l,i*}) \frac{\partial r_1^{l,i*}}{\partial d_0^i} = 0.$$

Using (10), the optimality condition becomes,

$$\int_{r_1^{l,i*}}^{+\infty} \beta (r_1^{l,i} - r_0) dF^{r^l}(r_1^l) = 0,$$

and then it is held that, in equilibrium,

$$1 - \delta + \alpha \bar{Z}^{i,+} (l_0^i)^{\alpha-1} = r_0, \tag{15}$$

[19] where F^{r^l} is the cumulative distribution function of r_1^l, which inherits the distributive properties of Z_1.

where $\bar{Z}^{i,+}=\mathbb{E}_0\{Z_1|Z_1 > Z^{i,*}\}$.

From the equations (9) and (15), we see that the interaction of limited liability and deposit insurance leads banks to overestimate the marginal benefit of borrowing, since $\bar{Z}^{i,+} > \bar{Z}$. In other words, since banks do not absorb negative losses and the deposit interest rate is risk insensitive, banks take excessive risk leading to inefficiently high loan allocation.

From the equation (15) *ceteris paribus*, the probability of bank default increases bank incentives to take excessive risk and hence to excessively lend. We call the impact on risk taking and credit through the bank default probability as *excessive bank risk taking channel*.

Our measure of inefficiency or measure of excess bank risk, which involves the volume of credit, is the relative difference in loans under unlimited liability (UL) and under limited liability (LL), that is, $l_0^{i,LL}/l_0^{i,UL}-1$.

Under unlimited liability, as suggested by (9) a higher interest rate increases the marginal cost of deposits and reduces bank lending. Under limited liability and deposit insurance, according to (15), in addition to the direct impact there is also an indirect impact of the interest rate on the marginal benefit of loans, through the probability of bank default, in addition to the general equilibrium effect of credit on the probability of default.

Regarding the indirect impact, on the one hand, for a given level of loans, a lower interest rate reduces bank obligations, increases bank profits and, therefore, reduces the probability of bank default; while, on the other hand, higher credit (caused by the direct impact of the rate on the marginal cost of borrowing) produces the opposite effect on the probability of bank default.

According to equation (12), for the same level of loans l_0^i, the negative impact of a lower risk-free rate (funding cost of deposits) in the probability of bank default, it will be more important the greater the share of deposits as a source of credit funding, or the greater the leverage, which we can define as the ratio d_0^i/l_0^i. In other words, due to the excessive risk-taking channel, according to equation (12), the impact of expansionary monetary policy (captured by a reduction in the risk-free interest rate) on credit should be relatively less positive in those banks with relatively high levels of leverage.

The intuition is that lower financing costs increase profits for banks (*profit channel*). Therefore, banks, in order to guarantee these higher profits, take less risk. So, the greater the share of deposits, the incentives for less risk in the face of a reduction in the risk-free rate are stronger. What amplifies the negative impact on the granting of credit through the channel of excessive risk taking.

Next, we parameterize the model (with limited liability and deposit insurance) and

present numerical solutions since analytical solutions are not straightforward. The parameter of capital participation in production, α, is set at 0.33, which is a standard value in the literature. We set our quarterly capital depreciation rate at 20%, which is higher than the literature, to avoid an explosive credit response.[20] The household's discount factor, β, is set at 0.98, which is a relatively standard value. We say that $\mu_z = -0.5\sigma_z^2$ so the unconditional mean of productivity is zero. Therefore, σ_z will mainly affect productivity volatility and help us to set the size of the bank default probability.

The other two parameters, $\{n_0, \sigma_z\}$, are established so that the bank leverage ratio (l_0/n_0) and the annual bank default probability are 6.0 and 1.5, respectively.[21] The level of leverage is slightly below the levels observed in Peru. In general, in many countries bank leverage has been reduced, albeit slowly since the 2008 global financial crisis. The level of default probability is a conservative, considering the probability of a banking crisis in emerging economies built with the data of Laeven and Valencia (2020). The table 8 summarizes the values of the parameters.

Tabla 8: *Parameters*

Description	Parameter	Value	
Capital share	α	0.33	Standard value
Capital depreciation rate	δ	0.20	
HHs discount factor	β	0.984	Standard value
			Target
Initial bank net worth	n_0	0.353	Bank leverage of 6.0
Volatility of $\ln(Z_1)$	σ_z	0.724	Annualized default probability 1.5%

The figure 7 shows us the numerical results of the impact of the monetary policy (changes in r_0) on the level of leverage, credit and bank default probability. To see the impact of the monetary policy conditional on the leverage level, results are reported for three different levels of leverage (low, medium and high).[22] According to the figure, 7 we see that at a higher leverage, the impact of the expansionary policy on credit is less positive. However, the difference in terms of the heterogeneous response to monetary

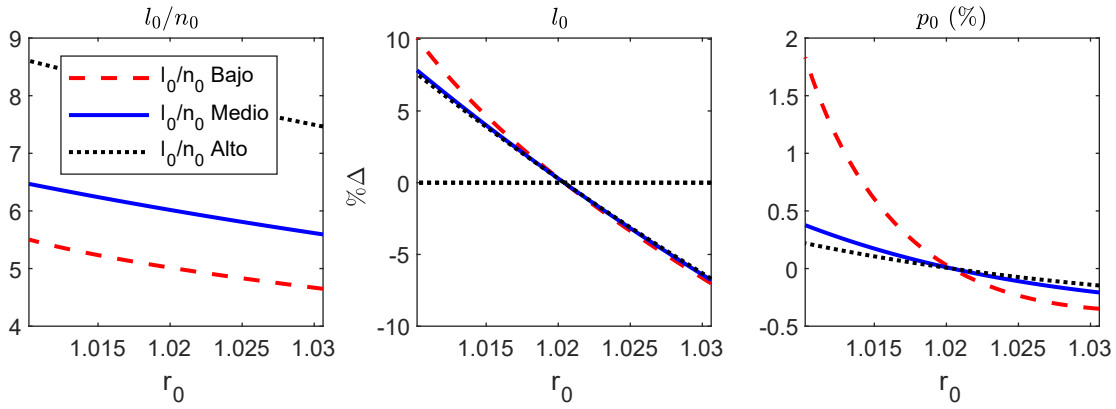
[20]The lower δ, the more explosive the credit response to the rate of interest. If the principal does not fully depreciate, one part of the loan payments is risk-free, $(1-\delta)l_0^i$, and the other part is risky, $\alpha Z_1 (l_0^i)^{\ alpha}$. Capital that does not depreciate reduces the relative importance of venture bank income. Technically, it means that after a change in bank liabilities (driven by interest rate changes) large adjustments of Z^* are required to keep earnings unchanged (or equal to zero). As a result, the probability of bank and credit default is more sensitive to changes in the interest rate. To guarantee a positive probability of default, in the worst state of nature, bank revenues must be less than bank liabilities, that is, $\delta > 1 - r_0(1 - n_0^i/l_0^i)$. This produces a lower bound for δ. In fact, the lower bound for our annualized δ turns out to be higher than that observed in the data.

[21]The quarterly bank default probability p_0 of the model is found in the following equation $3.5\% = p_0\left[1 + (1-p_0) + (1-p_0)^2 + (1-p_0)^3\right]$.

[22]The medium level is the one used in the baseline parameterization. Every time the target level of leverage is changed, the parameters n_0 and σ_z are recalculated.

policy is not very significant. The latter may be associated with the simplicity of the model, and the difficulty involved in its parameterization.

Figura 7: *Monetary policy conditional to the bank leverage level*



%Δ: Percentage change with respect to the base level of the interest rate $r_0 = 1/\beta = 1/0.98 = 1.02$.

7.2 Results at bank branch level

Although the previous results are associated with the heterogeneous response of credit at bank level, under some assumptions, we can extrapolate them to also conclude the same at the branch level. Therefore, in what follows we focus on branches of the same bank. In that case, i now refers to a branch.

We assume that lending decisions are made at the branch level. In this specification, a bank branch i that can take local deposits as d_0^i and issue loans as l_0^i, requires external financing as $b_0^i = l_0^i - d_0^i$. If $l_0^i > d_0^i$, the branch requires positive external financing; otherwise, the branch is lending rather than borrowing funds. As at the bank level, the profit of the branch i is:

$$max\left\{0, r^{l,i} l_0^i - r_0 d_0^i\right\}.$$

As before, we can get:

$$Z^{i*} = \left[r_0 \frac{d_0^i}{l_0^i} - (1-\delta)\right] \frac{1}{\alpha} (l_0^i)^{1-\alpha}, \tag{16}$$

and,

$$1 - \delta + \alpha \bar{Z}^{i,+} (l_0^i)^{\alpha-1} = r_0. \tag{17}$$

An additional difference with the bank-level model is that d_0^i could be greater than l_0^i, so higher levels of heterogeneity on $\frac{d_0^i}{l_0^i}$ could be observed at branch level.

As a result, we could find that for those branches with a relatively high ratio $\frac{d_0^i}{l_0^i}$ (or with a relatively low dependence on external funding), the indirect impact of r_0 on credit, through the branch default probability, might be more important. So, at the branch level, the heterogeneous impact of the monetary policy condition on bank (branch) leverage

should be more marked.

Finally, the model rationalizes the results found in the data at bank level and at branch or district level, that the impact of monetary policy can be conditional on the leverage level. In particular, the greater the branches' dependence on external financing (or the less bank leverage, $\frac{d_0^i}{l_0^i}$) the more positive will be the impact of expansionary monetary policy on credit, amplifying its impact . Consequently, the lower the dependence on external funding (the higher $\frac{d_0^i}{l_0^i}$) we would see an attenuating effect on the impact of monetary policy.

8 Conclusion

In this research work we document the existence of a heterogeneous banking channel in Peru, where monetary policy actions have an amplifying and a dampening effect on the credit market, and depend on the leverage level of financial institutions. To show causal effects of this mechanism, we use a more granular level of data, and we show that monetary policy actions also affect supply decisions in a bank branch, but that effect will depend on the greater or lesser dependence of this branch on external funding that are not local deposits. In the aggregate, it is the amplifying margin, due to the heterogeneity of the funding, that determines the credit market corrections after monetary policy actions.

Our results, to our knowledge, are the first effort to understand how the micro dynamics of the financing structure of financial institutions impacts the transmission mechanisms of monetary policy. However, our work can still be extended to explain the reasons for our findings. For example, it is possible that those bank branches that depend less on local deposits as a source of funding (that is, they have a high ratio of loans to deposits) are less exposed to an increase in the interest rate of the monetary policy that is passed on to higher interest rates on deposits. However, these same branches require more external funding and the results we show suggest that the cost of this external financing increases much faster than the cost of funding local deposits. Therefore, the greater the dependence on external financing sources, the more positive is the impact of the policy rate on the cost of financing. Although this may be a plausible explanation, it is also necessary to document a deposit channel, and study the behavior of interest rates in the deposit and credit markets together.

Although we study the impacts of monetary policy conditional on a characteristic of the credit market, there are other elements of the Peruvian financial market that could also be important, such as the relatively high level of financial frictions, the relatively high market power of banks and the existence of underdeveloped capital markets. Although these are also important supply channels, we leave their exploration for a future research agenda.

References

Agenor, Pierre-Richard, Koray Alper and Luiz Pereira da Silva. 2018 "External shocks, financial volatility and reserve requirement". *Journal of International Money and Finance* 8, 23-43.

Agur, I. and M. Demertzis. 2012."Excessive Bank Risk Taking and Monetary Policy." ECB WP 1457.
tai and Maria Demertzis, 2015, *Will Macroprudential Policy Counteract Monetary Policy's Effects on Financial Stability?*, IMF WP 283.

Agur, I. and M. Demertzis. 2019. "Will Macroprudential Policy Counteract Monetary Policy's Effects on Financial Stability?" *The North American Journal of Economics and Finance*, 48, 65-75.

Aguirregabiria, Victor, Robert Clark and Hui Wang. 2019. "The Geographic Flow of Bank Funding and Access to Credit: Branch Networks, Local Synergies, and Competition," CEPR Discussion Papers 13741, C.E.P.R. Discussion Papers.

Bernanke, Ben S. and Ilian Mihov, 1998. "Measuring monetary policy." *The Quarterly Journal of Economics*, 113 (3), 869-902.

Bernanke, B. and A. Blinder 1988, "Credit, Money and Aggregate Demand," *American Economic review.*

Bernanke, Ben S., and Mark Gertler. 1995. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives*, 9 (4): 27-48.

Bernanke, B. S., Gertler, M., and Gilchrist, S. 1999. "Chapter 21 The financial accelerator in a quantitative business cycle framework." *Handbook of Macroeconomics*, 1(PART C), 1341-1393.

Borio, C. and H. Zhu. 2012. "Capital regulation, risk-taking and monetary policy: A missing link in the transmission mechanism?," *Journal of Financial Stability*, 8, 236-251.

Bräuer, Leonie and Gerhard Rünstler. 2014. "Monetary policy transmission over the leverage cycle: evidence for the euro area." Working Paper No 2421, BCE.

Bustamante, José, Walter Cuba and Rafael Nivin, 2019. "Determinants of credit growth and the bank-lending channel in Peru: A loan level analysis," BIS Working Papers 803, Bank for International Settlements.

Carrera, Cesar. 2011. "El canal de crédito bancario en el Perú: Evidencia y mecanismo de transmisión". *Revista Estudios Económicos* 22, 63-82.

Collard, F., H. Dellas, B. Diba and O. Loisel. 2017. "Optimal Monetary and Prudential Policy," *American Economic Journal: Macroeconomics*, 9(1) 40-87.

De Nicolò, G., A. Gamba and M. Lucchetta. 2012. "Capital Regulation, Liquidity Requirements and Taxation in a Dynamic Model of Banking". IMF WP 72.

Dell'Ariccia, G., L. Laeven and R. Marquez. 2014. "Real interest rates, leverage, and bank risk-taking." *Journal of Economic Theory*, 149(1), 65-99.

Dell'Ariccia, G., L. Laeven and G. A. Suarez. 2016. "Bank leverage and monetary policy's risk-taking channel: evidence from the United States." *The Journal of Finance*, 72(2), 613-654.

De Groot, Oliver. 2014. "The Risk Channel of Monetary Policy," *International Journal of Central Banking*, 10(2), 115-160.

Drechsler, Itamar, Alexi Savov and Philipp Schnabl. 2017."The Deposits Channel of Monetary Policy", *The Quarterly Journal of Economics*, 132 (4)1819-1876.

Duncan, O. D., and Duncan, B. 1955. "A methodological analysis of segregation indexes." *American Sociological Review* 20: 210-17.

Galí, Jordi. 2015. "Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and its Applications." Princeton University Press, Second Edition.

Gertler, M. and Karadi, P. 2011. "A Model of Unconventional Monetary Policy", *Journal of Monetary Economics* 58, 17-34.

Gertler, M., and P. Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics*, 7 (1): 44-76.

Gertler, M. and Kiyotaki, N. 2015. "Banking, Liquidity, and Bank Runs in an Infinite Horizon Economy." *American Economic Review* 105(7), 2011-2043.

Jiménez, Gabriel, Steven Ongena, José-Luis Peydró, and Jesúus Saurina. 2017. "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say about the Effects of Monetary Policy on Credit Risk-Taking?" *Econometrica*, 82 (2014), 463?505.

Jordà, Oscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95(1):161-182.

Holm, Martin, Pascal Paul and Andreas Tischbirek, 2021. "The Transmission of Monetary Policy under the Microscope," *Journal of Political Economy*, 129, 10, 2861-2904.

Kashyap, Anil and Stein, Jeremy. 1994. "Monetary Policy and Bank Lending", p. 221-261, Book: Monetary Policy. The University of Chicago Press. National Bureau of Economic Research.

Kashyap, Anil and Stein, Jeremy. 2000. "What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?" *American Economic Review*, 90, 407-428.

Kiyotaki, N. and J. Moore. 1997. "Credit Cycles", Journal of *Political Economy*, 105, 211-48.

Laeven, L. and F. Valencia. 2020. "Systemic Banking Crises Database II," *IMF Economic Review*, 68, 307-361.

Li, Zehao. 2022. "Financial intermediary leverage and monetary policy transmission", *European Economic Review,* 144.

Newey, Whitney K. and Kenneth D. West. 1987. "A Simple, positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (3): 703-708.

Nguyen, Hoai-Luu Q. 2019. "Are Credit Markets Still Local? Evidence from Bank Branch Closings." *American Economic Journal: Applied Economics*, 11 (1): 1-32.

Pozo, Jorge. 2019. "Bank Risk-Taking in a Small Open Economy." *Working Paper BCRP*, 2019-16.

Pozo, Jorge and Youel Rojas, 2020a. "Unconventional Credit Policy in an Economy with Supply and Demand Credit Frictions," *Working Paper BCRP*, 2020-14.

Pozo, Jorge and Youel Rojas. 2020b. "The Risk-Taking Channel of Monetary Policy: A New Approach and Evidence from Peru." *Working Paper BCRP*, 2020-19.

Pozo, Jorge y Youel Rojas. 2021b. "Bank Competition and Risk-Taking". Working Papers 2021-007, BCRP.

Pozo, Jorge and Youel Rojas. 2021. "Unconventional Credit Policy in an Economy under Zero Lower Bound." IHEID Working Papers 14-2021, Economics Section, The Graduate Institute of International Studies.

Pozo, Jorge and Youel Rojas. 2022. "Dissecting The Risk-Taking Channel of Monetary Policy: A New Approach and Evidence from Peru". Mimeo.

Quispe, Zenón. 2001. "Transmission mechanisms of monetary policy in an economy with partial dollarization: The case of Peru". BIS papers 8, 210-231.

Ramey, Valerie A. and Sarah Zubairy. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy*, 126(2), pages 850-901.

Romer, Christina D. and David H. Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94(4):1055-1084.

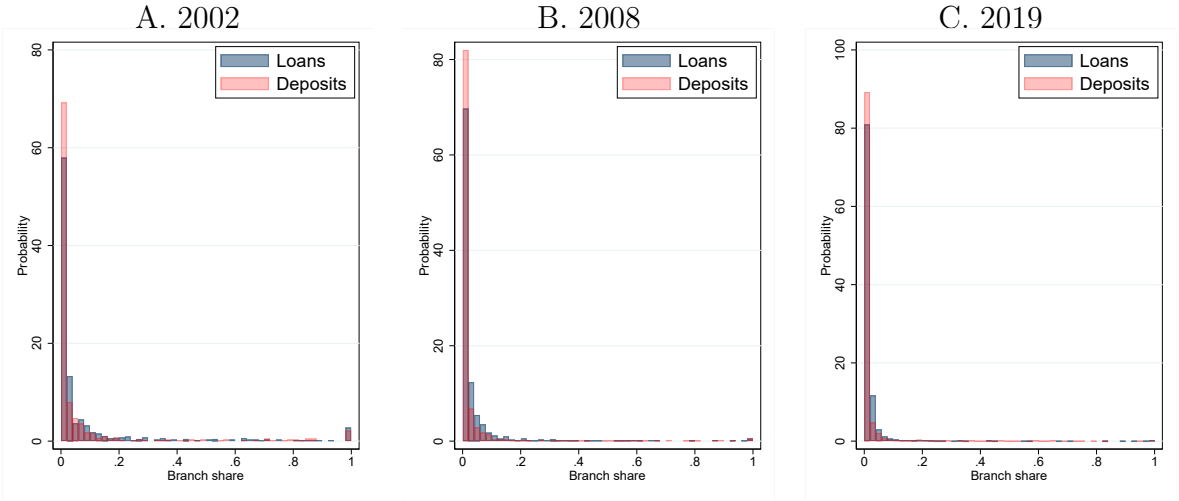
Viladegut, Hugo and Miguel Cabello. 2014. "El canal de crédito en el Perú: una aproximación VAR." *Revista Estudios Económicos*, 27, 51-66.

White, Michael J. 1983. "The Measurement of Spatial Segregation." *American Journal of Sociology* 88 (5): 1008-1018.

Appendices

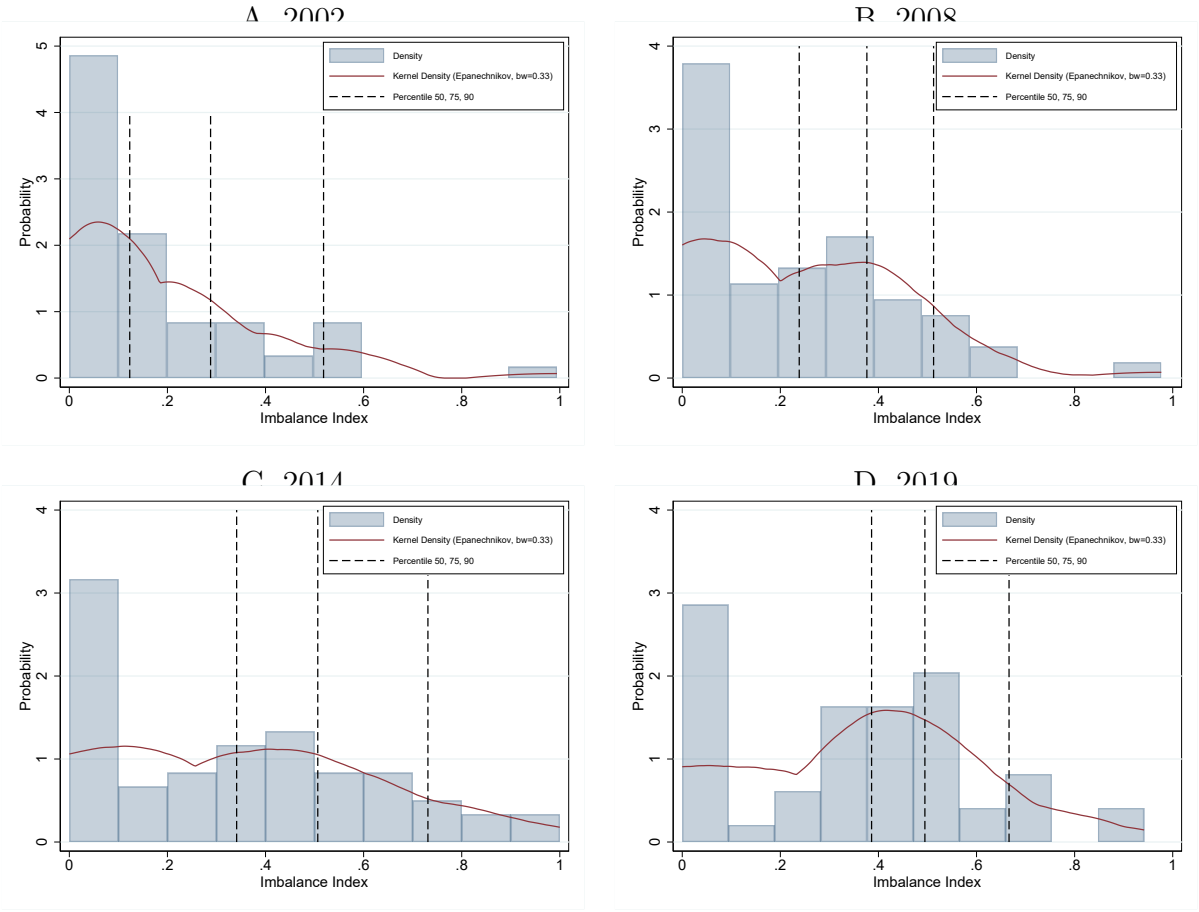
A Other figures

Figura 8: *Distribution of the participation of credits and deposits of financial entities in a district*



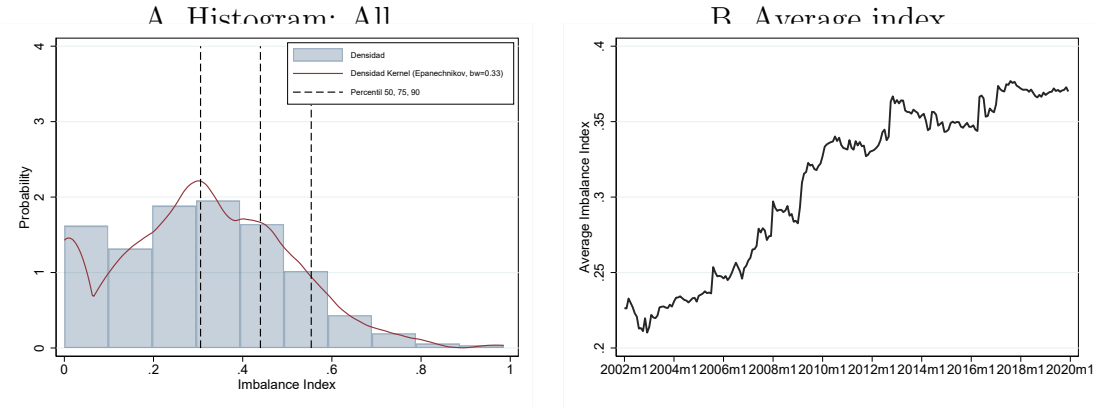
Note: This graph shows the distribution of shares of deposits and loans of each financial institution in a district. The participation of credits and deposits is calculated as the ratio of the amount of deposits or loans of an entity-district over the total amount of loans or deposits of the financial institution in December of each year. The shares are shown grouped for all banks and a given year.

Figura 9: *Distribution of the imbalance index between loans and deposits at the bank level*



Note: The figure shows the histogram at the bank level of the credit and deposit imbalance index, measured by the specification GI_{bt} in the equation 3, for the years 2002, 2008, 2014 or 2019.

Figura 10: *Distribution of the imbalance index between credits and deposits at the level of depository entities*



Note: These graphs consider the sample of depository entities, and therefore exclude financial companies. Panel A shows the bank-level histogram of the credit and deposit imbalance index, measured by the GI_{bt} specification in the equation 3, for all banks and all sample periods. 2002.1-2019.12. Panel B shows for each time period the average imbalance index of all financial entities.

B Robustness: regression at the bank level

Tabla 9: *Regression at bank level, monthly frequency, estimated with OLS: include extreme values of credit growth*

	Dependent variable: Δy_{bt}					
	(1)	(2)	(3)	(4)	(5)	(6)
$GI_{bt-1}\epsilon_t^{MP}$				0.0662	0.135	0.0461
$RWA_{bt-1}\epsilon_t^{MP}$	0.0346**	0.0405***	0.0412***	0.0322**	0.0361**	0.0395**
ϵ_t^{MP}	-0.265**	-0.316***		-0.267**	-0.323***	
RWA_{bt-1}	-0.00336**	-0.0109***	-0.00647***	-0.00447***	-0.0111***	-0.00751***
GI_{bt-1}				0.0314**	0.0191	0.0285**
Obs.	2,686	2,686	2,686	2,686	2,686	2,686
R^2	0.004	0.101	0.101	0.007	0.101	0.102
F test (ρ-value)	0.00746	8.47e-07	6.35e-06	0.00278	4.83e-06	7.82e-06
Bank FE		✓			✓	
Time FE			✓			✓

*p<0.1, **p<0.05, ***p<0.01.

Financing Imbalances and Heterogeneous Effects of Monetary Policy

Jorge Pozo **Youel Rojas**

Central Reserve Bank of Peru

IFC-Bank of Canada Satellite Seminar on “Granular data: new horizons and challenges for central banks” - 2023

The opinions expressed in this working paper correspond solely to those of the authors.

Outline of the paper

Motivation:

- MP has an important supply influence in the credit markets
- MP transmission must be influenced by the financing structure of a bank: Heterogeneous

Research Question:

- How does the financing structure of the credit markets matter for MP transmission?
  - What are the implications of external financing and within financing structure of in the transmission of MP?

Methodology: We develop two approaches

1. Theory: A simple two periods model to discuss mechanisms: After a MP change, higher leverage influence banks profits and its incentives to issue loans.
2. Mainly Empirics:
   - Peruvian data: Financial firm branch-level data & regional credit markets.
   - Causal : Identification strategy (Inequality of credit & deposits shares) + estimated MP shocks

Results:

- Bank-level: Regional financing is not important, but Leverage determines mainly MP response.
- Branch-level: Heterogeneity in the transmission of MP shocks depends on the relative size of external funding $\rightarrow$ Amplification and dampening effects of MP

Aggregate evidence of MP on credit markets (I)

How does the credit market adjust to a MP shock?

- We use a new series of monetary policy shocks constructed for the Peruvian economy by Pozo and Rojas (2022), for the 2003:Q4-2019:Q4 period.
- We prove the validity of the estimated shocks using Local Projections (Jordà 2005) on credit and the lending rate. This estimation strategy follows a flexible form that imposes few restrictions on the data:

$$y_{t+h} - y_{t-1} = \alpha_h + \beta_h \epsilon_m^{MP} + \sum_{k=1}^{K} \gamma_k^h X_{t-k} + u_t^h, \quad (1)$$

where $h = 1, 2, ..., 16$. X includes

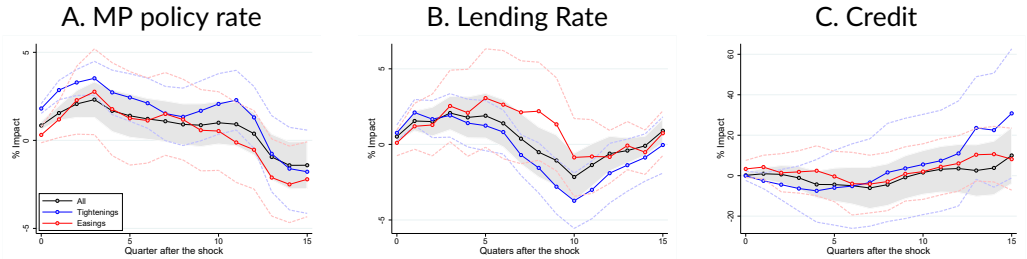
  - 3 years of lagged values of the monetary policy shock
  - 2 years of lagged values of the dependent variable.
  - A dummy variable for post 2009, after GFC.

- We also differentiate between possible asymmetries of changes in monetary policy: expansive or contractionary MP stance.

Aggregate evidence: credit market adjustment after a MP shock (II)

MP shock: Important Supply side adjustment of a credit markets. No considerable evidence of asymmetric effects of MP in credit markets.

Figure: LP Results: Positive Monetary Policy Shock & Credit Market



Note: Impulse response functions to a 1 percent increment in the domestic MP policy rate.

Limits: With aggregate data, specific mechanisms are hard to disentangle.

Granular exploration

- In what follows, we explore granular data to obtain causal effects on financing imbalances credit supply side of monetary policy shock.
- MP heterogeneous **financing structure channel**: Bank's response to MP shocks may depend on:
  1. Needs of external financing: Bank's leverage
  2. Bank within financing network: Bank's geographic imbalance index of credit-deposit
- **Two assumptions** in this paper:
  1. Regional markets: Geographic proximity reduces costs of transmitting and processing information.
  2. Risk varies geographically: By inherent characteristics regions show heterogeneous and persistent riskiness.

A Simple Two-Period Model

The model:

- Economy of identical households, heterogeneous banks and firms. All agents are risk-neutral.
- Banks or bank branches grant loans to risky firms, financed with local deposits and other funds (equity and external loans).
- Key features:
  - Banks face limited liability and deposits are insured by the government.
  - Each bank has one branch in each province.

Results:

1. Banks overestimate the return of their risky loans $\rightarrow$ MP risk-taking channel
2. The positive impact on credit after a expansionary MP diminishes with bank leverage.
3. Effects are amplified at branch level.

Empirical exploration: granular data

Financial Regulator (SBS) data at bank and branch level

- Bank level indicators: RWA, NPL, Size, ROA.
- credit-deposits at branch-level.
- Sample :
  - 2003m1-2019m12
  - 78 financial institutions and 3,682 branches located in 451 districts and 189 provinces.

Table: Descriptive statistics of the observations at the bank-time level

Variables	N° Obs.	Mean	Standard deviation	Min	Max
Δy_{bt}	2933	0.012	0.06	-0.62	0.80
RWA_{bt}	2786	6.800	1.66	0.80	10.24
NPL_{bt} (%)	2786	3.300	3.42	0.00	34.28
$SIZE_{bt}$ (%)	2933	6.176	8.63	0.00	32.55
ROA_{bt} (%)	2748	1.711	1.99	-12.51	8.06
GI_{bt}	2784	0.270	0.21	0.00	0.78
ϵ_t^{MP} (%)	196	0.000	0.10	-0.23	0.35

Granular data: Geographic Imbalances between Loans and Deposits

Geographic distribution of district net borrowers & net lenders:

$$GI_{S_L - S_D} = \frac{cre_{dt}}{Cre_t} - \frac{dep_{dt}}{Dep_t}.$$

Very heterogeneous pattern.

A. 2002 B. 2008 D. 2019

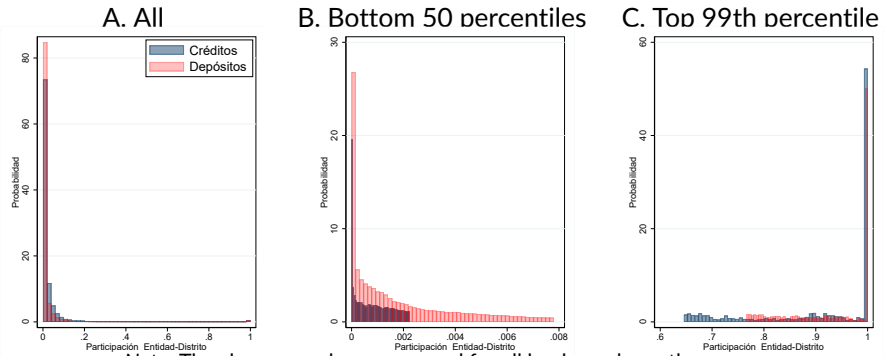


Note: It shows the distribution intervals divided by the percentiles: <10th , 10th-50th , 50th-90th and >90th .

Geographic Imbalances between Loans and Deposits Shares

- 50% of the district branches of financial entities have shares below 1% (0.2% for credit shares, and 0.8% for deposit shares).
- 1% of bank branches hold shares of more than 65% and 77% of total deposits and loans, respectively.

Figure: Distribution of the credit and deposit shares of financial institutions in a district



Note: The shares are shown grouped for all banks and months-years.

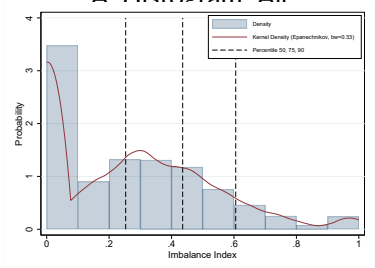
Financing network: Geographic Imbalances index

We compute an index that captures this imbalance of deposits and loans at the bank level:

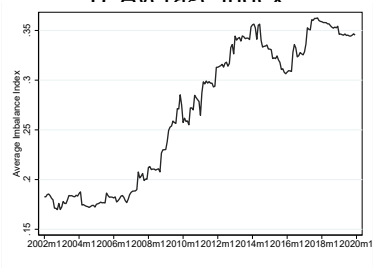
$$GI_{bt} = \frac{1}{2}\sum_{d=1}^{M}\left|\frac{dep_{dtb}}{Dep_{bt}} - \frac{cre_{dbt}}{Cre_{bt}}\right|$$

- It is a measure of the bank's internal bias.
- The index is greater than 0.5 for more than 20% of financial entities.
- Over time the distribution has moved further to the right: i.e more mobilization of deposits and credits across geographic locations.

A. Histogram: All



B. Average index



Note: Panel A shows the bank-level histogram of GI_{bt}. $GI_{bt} = 0$: prefect internal balance; $GI_{bt} > 0$: degrees of loan and deposit markets segmentation. Panel B shows for each period the average imbalance index of all financial entities.

MP & Financing Structure: Bank level

- We characterize the heterogeneous effect of monetary policy as a function of the financing structure: leverage and the geographic imbalance index .
- We estimate the following model at the bank level:

$$\Delta y_{bt} = \alpha + \beta_1 RWA_{bt-1} \times \epsilon_t^{MP} + \beta_2 GI_{bt-1} \times \epsilon_t^{MP} + \beta_3 \epsilon_t^{MP} + \beta_4 RWA_{bt-1} + \beta_5 GI_{bt-1} + \beta_6 X_{bt} + \epsilon_{bt},$$

MP & Financing Structure: Bank level results

- Heterogeneous effects of MP: Leverage more important than within finance network.
- Distribution of leverage matter for MP transmission:
  - Different leverages, $\sigma \pm \overline{RWA}$ = (5.1; 8.5),
  - Different credit growth impacts of MP = (5.9pbs; -0.9pbs)

	Dependent variable: Δy_{bt}					
	(1)	(2)	(3)	(4)	(5)	(6)
$GI_{bt-1}\epsilon_t^{MP}$				-0.00903	0.00577	-0.0209
$RWA_{bt-1}\epsilon_t^{MP}$	0.0201***	0.0237***	0.0243***	0.0203***	0.0235***	0.0248***
ϵ_t^{MP}	-0.164***	-0.193***		-0.163***	-0.193***	
RWA_{bt-1}	0.00323***	0.00151	0.00246***	0.00266***	0.00168	0.00194**
GI_{bt-1}				0.0164***	-0.0187	0.0153**
NPL_{bt}	-0.00158***	-0.00215***	-0.00118***	-0.00206***	-0.00203***	-0.00170***
ROA_{bt}	0.000681	-0.000593	-1.50e-05	0.000162	-0.000351	-0.000467
$SIZE_{bt}$	-8.11e-05	0.00322**	-6.74e-06	-7.92e-05	0.00330**	-1.13e-05
Obs.	2,642	2,642	2,642	2,642	2,642	2,642
R^2	0.025	0.087	0.121	0.027	0.087	0.124
F test (ρ-value)	0	4.81e-10	8.66e-07	0	2.85e-09	4.49e-07
Bank FE		✓			✓	
Time FE			✓			✓

*p<0.1, **p<0.05, ***p<0.01.

Branch-level Identification strategy

Under the financing channel, lending supply should be more sensitive at the branch lending markets.

- A MP shock introduces additional incentives for banks to pay cost for portfolio reallocation and take risk.
- At the branch level: credit-to-deposit ratio is a measure of the composition of balance sheet liabilities or leverage.
- **Key idea**: If a bank branch has a high credit-to-deposit ratio, financing additional liquidity on top of deposits is expensive.

Dealing with omitted variables at bank - level analysis.

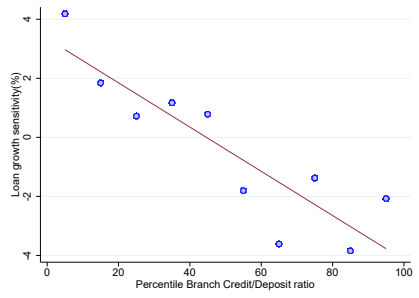
- **Problem**: Need to have variation in the composition of liabilities that is independent of the bank's lending opportunities or demand factor influencing bank's decisions.
- **Solution**: we compare lending across branches of the same bank located in different provinces.

Sensitivity of loans at bank branch to MP

In reduced form, credit-to-deposit ratio determines heterogeneous responses to MP:

- Amplification of MP: branches with credit-to-deposit ratio above the 50th percentile reduce more credit after a tight MP.
- Mitigation of MP: branches that have a loan-to-deposit ratio below the 50th percentile go in the opposite direction after the same MP change.

Figure: Credit growth sensitivity and the credit-to-deposit ratio



Note: i) we estimate a time series regression for each branch $\Delta y_{jt} = \alpha_j + \beta_j \Delta i_t + \epsilon_{jt}$. ii) we average the credit betas of all branches within a district

Within-Bank estimator of the importance of leverage

Our main **specification** for Δy_{bdt}, is the growth rate of loans granted by the branches of the bank b in district d at time t:

$$\Delta y_{bdt} = \alpha_{bt} + \alpha_{rt} + \alpha_{dt} + \beta_0 \left(\frac{Cred}{Dep}\right)_{bdt-1} \epsilon_t^{MP} + \beta_1 \left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbf{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP} + \gamma_x X_{bdt-1} + \epsilon_{bdt} \quad (2)$$

- $\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbf{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP}$ captures the channel of imbalances and/or external financing of the monetary policy at the bank branch level
- **Key set of controls**: α_{bt}, which absorbs all time differences across banks, to control for bank's lending opportunities. So, we compare across branches of same bank.

Branch estimations results

- Loan sensitivity to MP changes is non-linear and increases with credit-to-deposit ratio, even within a financial company.
- Results are robust to specifications, to sample selection problems or omitted variables, such as bank concentration or market risk.

	Dependent variable: Δy_{bdt}			
	(1)	(2)	(3)	(4)
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \epsilon_t^{MP}$	-0.00354	0.0686	0.119	-0.0585
	(0.111)	(0.145)	(0.216)	(0.135)
$\left(\frac{Cred}{Dep}\right)_{bdt-1} \times \mathbf{1}_{\frac{Cred}{Dep}} \times \epsilon_t^{MP}$	-0.132*	-0.174*	-0.238*	-0.306**
	(0.0777)	(0.0893)	(0.145)	(0.146)
y_{bdt-1}	-0.588***	-0.693***	-0.0275	-0.769
	(0.0763)	(0.0949)	(0.208)	(0.624)
R^2	0.642	0.676	0.581	0.508
Obs.	86093	81974	58220	58544
Controls	✓	✓	✓	✓
Control Long lags			✓	✓
Bank-Time FE	✓	✓	✓	
Region-Time FE	✓		✓	
Distric-Time FE		✓		
Branch FE	✓	✓	✓	✓
Province FE	✓	✓	✓	✓
Time FE				✓

Note: *p<0.1, **p<0.05, ***p<0.01.

Discussion & conclusion

- Previous analysis using granular data shows evidence of a heterogeneous banking channel in Peru: leverage matters for MP transmission with two margins: amplification and mitigation.
- However, consistent with aggregate analysis, the amplifying margin must be stronger than the attenuating margin.
- Our results also suggest that the cost of external funding increases much faster than the cost of local funding (local deposits). Therefore, the greater the dependence on external funding sources, the more positive the impact of the policy rate on the credit funding cost.

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

The Eurosystem's granular data on securities, loans and entities – How we use such information today and what we may develop differently next time[1]

Frank Mayerlen,
European Central Bank

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

EUROPEAN CENTRAL BANK

EUROSYSTEM

ECB-PUBLIC

The Eurosystem's granular data on securities, loans and entities -

how we use such data today and what we may develop differently next time

Bank of Canada – IFC Satellite seminar on 'granular data: new horizons and challenges for central banks'



Frank Mayerlen

Head of Division, Directorate General Statistics, European Central Bank

15 July 2023

Overview

1. Background and dimensions
2. Need for an integrated multi-purpose product
3. Challenges and solutions
4. Predefined statistical product
5. What we should address to do better next time
6. The Integrated Reporting Framework (IReF)

www.ecb.europa.eu ©

1. Background and dimensions

WHY



- (also) financial crisis 2009 triggered **new policy questions**
- Information available inadequate: **aggregates were not enough**

→ Decision to **"move beyond the aggregates"** with credit data and securities data

WHAT did we do?



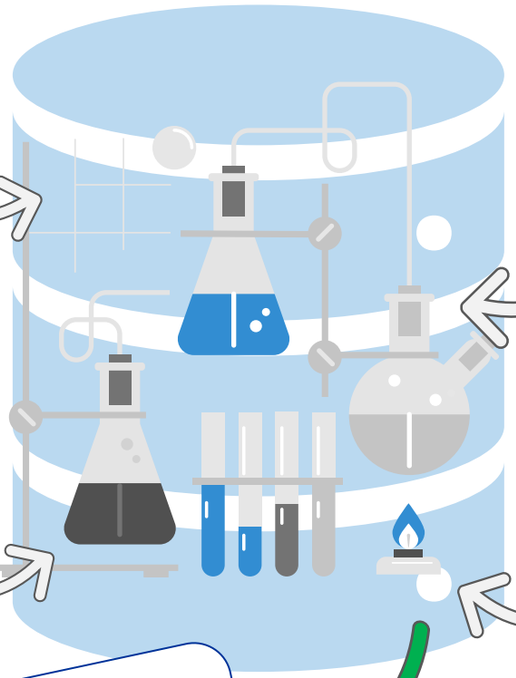
- **Securities issues** data 'Centralised Securities Database' *(CSDB),* since 2009; covering monthly and daily ca. 8 million active securities worldwide with 70+ attributes;
- **Securities holdings** data, 'Securities Holdings Statistics Database' *(SHSDB),* by sector (SHSS) and banking group (SHSG), since 2013
- Data on **banks' loans to legal entities** ('all except private households'), *(AnaCredit),* since 2018, reported by all 20 euro area countries, € 25,000 threshold; 70 attributes, covering monthly and quarterly ca. 30 million active instruments and about 5 million counterparties (mainly corporates)
- **Entity master data** 'register of Institutions and Affiliates' *(RIAD),* about 80 attributes and about 13 million records (this covers issuers, holders, creditors, debtors, etc.)

www.ecb.europa.eu ©

2. Need for an integrated multi-purpose product

Multi-purpose usage:

- Developing new/enhanced statistics (e.g. ad-hoc aggregates, climate)
- Monetary policy (e.g. SME access to finance)
- Research (e.g. effectiveness of monetary policy transmission)
- Financial stability (e.g. sectoral risk exposure, interconnectedness)
- Banking Supervision (e.g. credit exposures, borrower concentration)



CSDB: What are the total issuances of this entity?

RIAD: Who is this entity and in which group does it belong to?

SHSG: Which securities do supervised banks hold?

AnaCredit: Who received loans from this bank?

Integrated: e.g. 'who provides funding via loans or securities to a certain entity or group?'

3. Challenges and solutions

Challenges

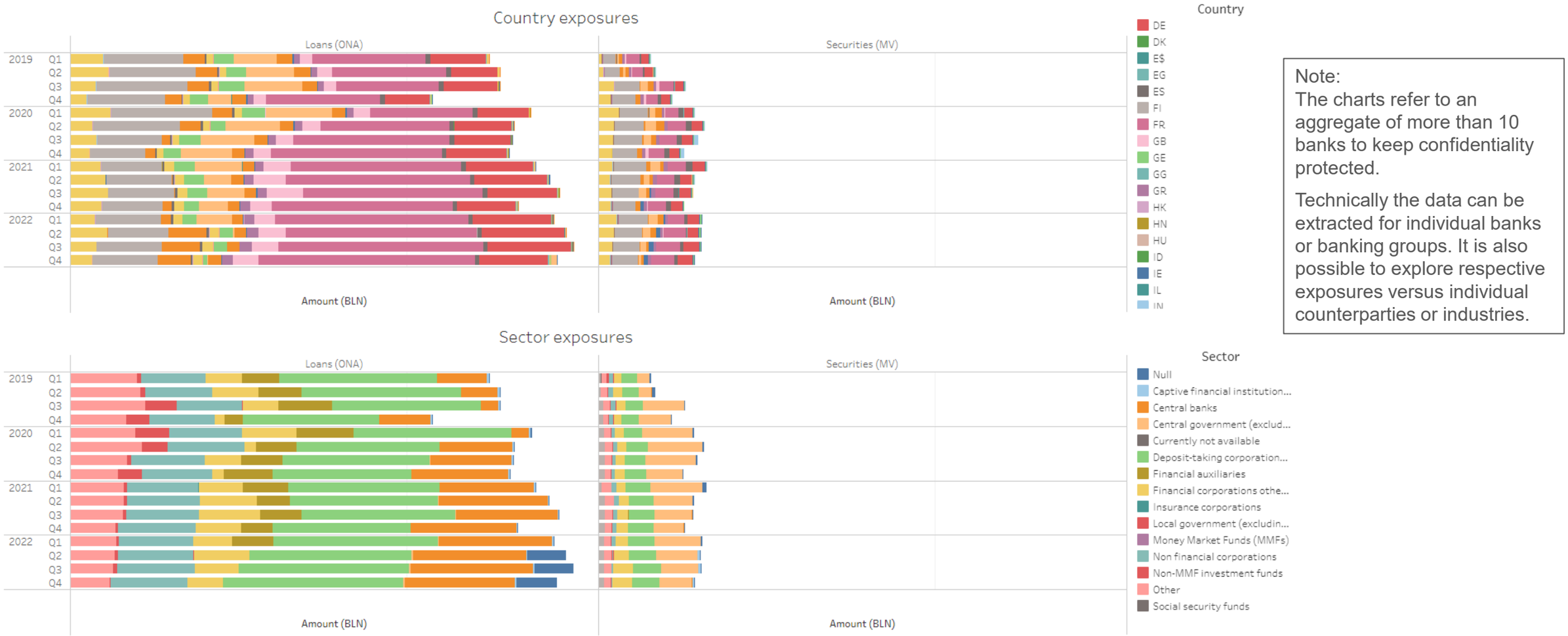
- **Unique identification** of entities across different statistical systems
- **Reconciliation** of differences in reporting population/coverage and level of granularity
- Differences in **frequencies, timeliness** and processing of **revisions**
- Calculations are **computationally heavy** and need to be refreshed frequently
- We need a **single core product**, reliable, replicable, which could be used as a basis for all users, also to connect other, potentially still less developed and standardised data → e.g. to develop climate related indicators

Solutions (see also annex)

- Technical approach via **2 big tables** covering entity reference data and amounts data (by entity), quarterly since 2019
  - ✓ Unique identification of **entities** across time and statistical systems
  - ✓ Establish **group structures** to build on
  - ✓ Establish **amount tables** by nominal value and market value/carrying amount
- Developing an **interface for non-technical users** and **database access for power users**
- Develop **predefined statistical products**, also at aggregated level
- **Collaborate** well across statistical domains!

www.ecb.europa.eu ©

4. Predefined statistical product



Note:
The charts refer to an aggregate of more than 10 banks to keep confidentiality protected.

Technically the data can be extracted for individual banks or banking groups. It is also possible to explore respective exposures versus individual counterparties or industries.

5. What we should address to do better next time

- **Data quality management:** Process requires high level of collaboration and coordination as it is not integrated; so far the underlying statistical regulations are output harmonised and primary reporting may differ among National Central Banks; production processes are not necessarily fully aligned
- **Access monitoring** is so far split and the access to the integrated dataset requires several separate permissions.
- **Gaps or overlaps of requirements which follow slightly different definitions** across different sets of statistics which can hamper integration
- **Increasing reporting burden** with each new requirement for reporting banks and National Central Banks compiling the data

www.ecb.europa.eu ©

6. The **I**ntegrated **Re**porting **F**ramework (IReF)? [1/3]

NCBs defined reporting requirements

Reporting overlaps existed

… One regulation

… Redundancy-free

Standardisation via theIReF

IReF collection layer

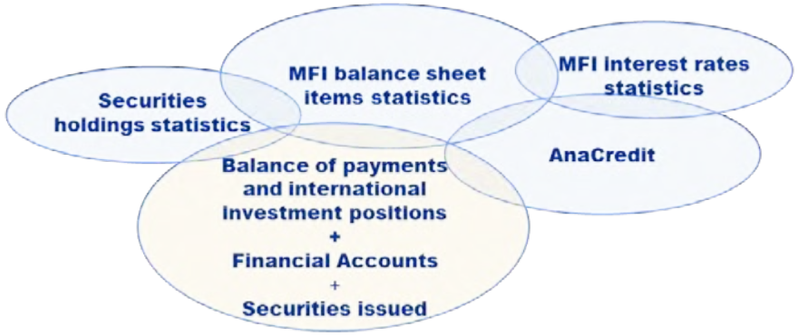


ONE euro area

ONE directly applicable regulation

ONE largely standardised implementation

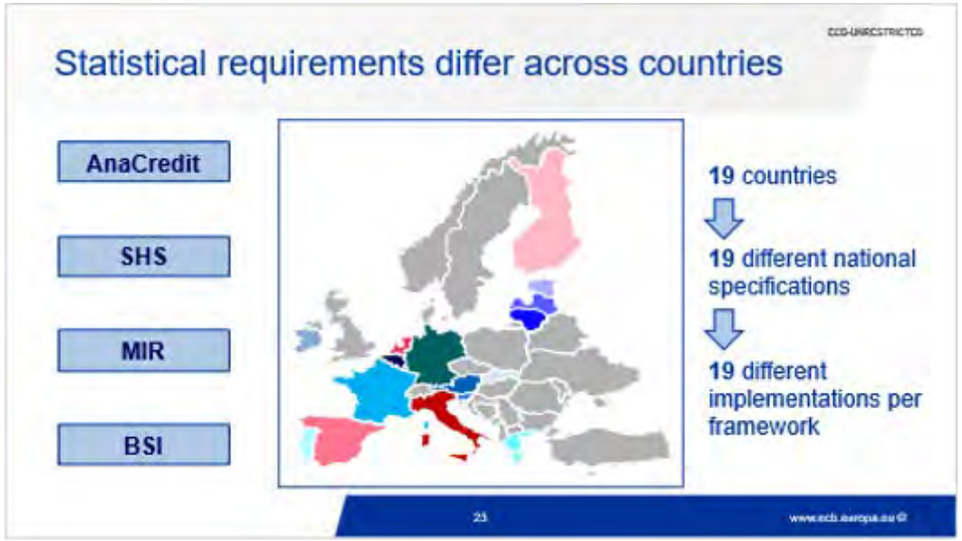
Datasets currently in scope of the IReF



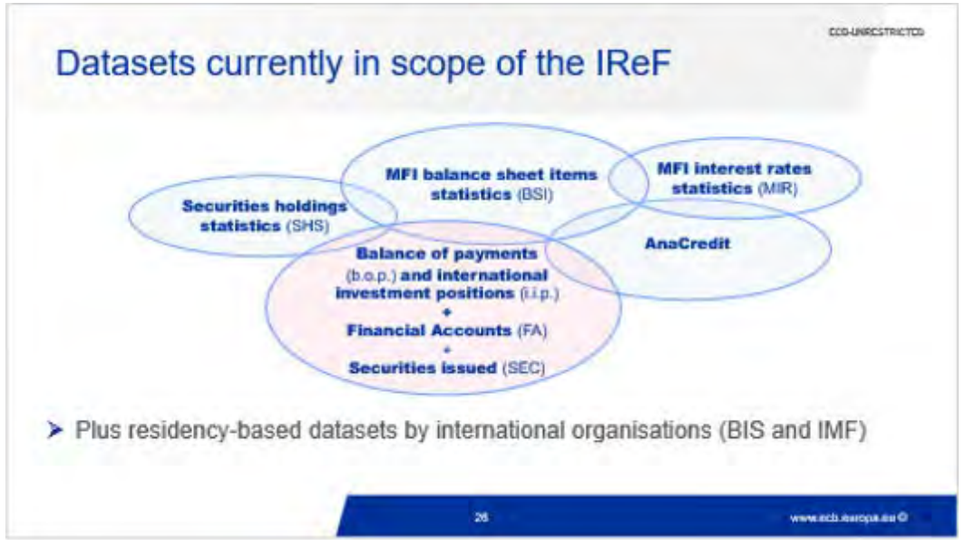
➢ *Plus supranational regulatory contents*

6. What is the IReF trying to achieve?

Cross country standardisation

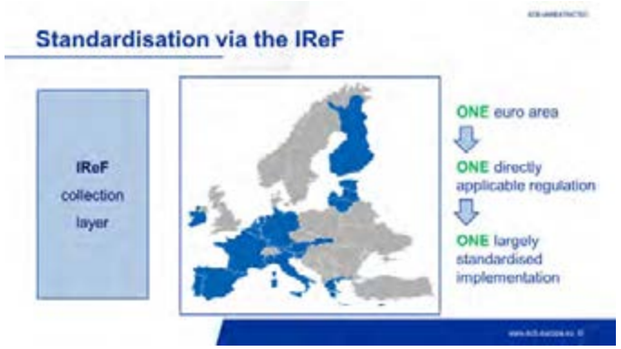


Redundancy-free reporting



6. What is the IReF trying to achieve?

Cross country standardisation



Redundancy-free reporting



www.ecb.europa.eu ©

Annex

Disclaimer: in the interest of available presentation time some of the issues in this presentation needed to be simplified. In case of further questions, please do not hesitate to contact me.

Frank Mayerlen
Head of Division
Analytical Credit & Master Data

Directorate General Statistics
EUROPEAN CENTRAL BANK
Mobile: +49 152 2261 6478
frank.mayerlen@ecb.europa.eu

www.ecb.europa.eu ©

Granular product – integration backbone

Full granularity data consisting of two tables

 Entity reference table

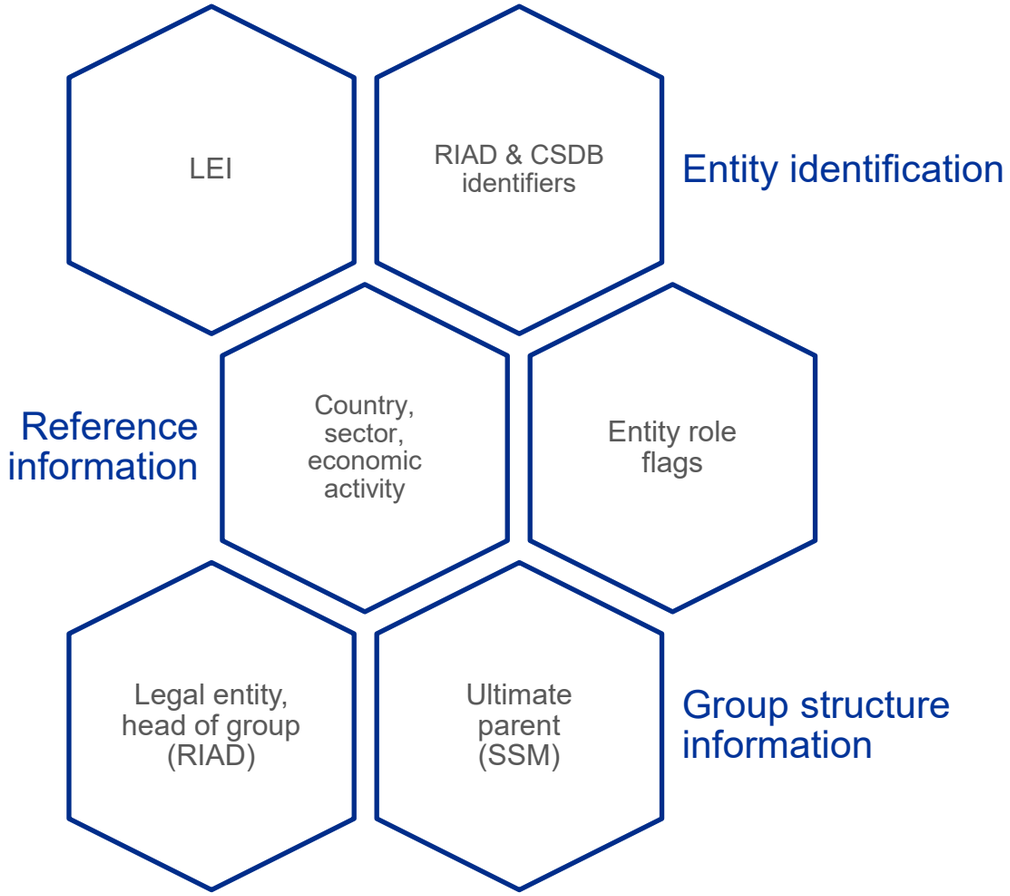
 Amount table (by entity)

Providing the ability to:

- uniquely identify entities across time and systems
- quantify the amount of securities/loans between entities
- establish a group structure to build upon

www.ecb.europa.eu ©

Entity reference table



All reference periods as of September 2018

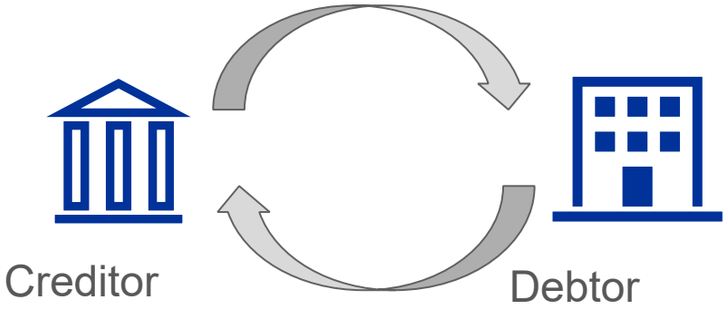
AnaCredit, CSDB, RIAD, SHSG, SSM information

Group structure based on SSM and RIAD

www.ecb.europa.eu ©

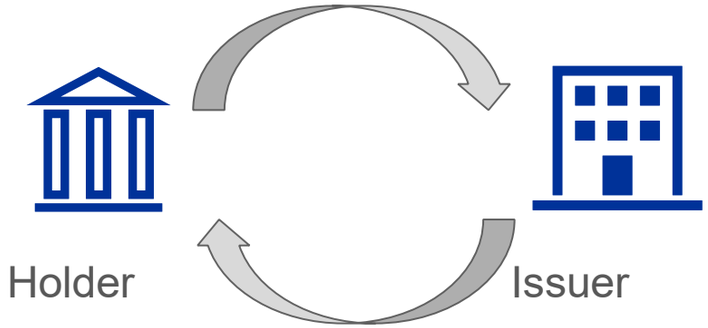
Amount table

Amount figures for the following 4 types of entity relationships:



measured by

- Carrying amount
- Outstanding nominal amount



measured by

- Holding amount (market)
- Holding amount (nominal)
- Amount outstanding
- Market capitalisation

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

The use of payment transaction data for economic forecasts[1]

Guerino Ardizzi, Giuseppe Bruno and Juri Marcucci,
Bank of Italy;

Roberto Iannaccone, Filippo Moauro,
Alessandra Righi and Davide Zurlo;
Italian National Institute of Statistics (ISTAT)

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

The Use of payment transaction data for economic forecasts

Guerino Ardizzi*, Giuseppe Bruno†, Roberto Iannaccone††, Juri Marcucci †, Filippo Moauro††, Alessandra Righi†† and Davide Zurlo††

* Bank of Italy, Currency Circulation and Retail Payments; † Bank of Italy, Economics and Statistics; †† ISTAT.

Abstract

Computation of the main components for Gross Domestic Product (GDP) and other relevant macroeconomic figures have been so far based on surveys among household and firms. These surveys are necessarily limited in sample size for obvious economic reasons. Aside from these reasons, natural phenomena like the recent Covid-19 pandemics, could completely wipe out the possibility to run this sort of surveys. On the other hand a huge amount of transaction data are regularly generated and digitally recorded from the different payment systems. These transaction data are very seldom used for building official statistics. In this work we consider billions of transactions from debit/credit card and data from the TARGET2 Gross settlement payment system as an alternative source of information for measuring variables such as consumption and tourism flows. Harnessing data arising from payment systems require special pre-processing and reconciliation operations, but once these steps are overcome they allow good improvements in pinning down the phase of the economic cycle, in squeezing the model forecasting variance and substantially shrinking the reporting times.

Even if card spending growth shows a substantial volatility compared with non-durable consumption growth, we find that normalized spending correlates quite closely with official consumption measures published by the National Statistical Office.

JEL classification: C32, C55, C81.
Keywords: Payment systems, technological platform, transaction data.

1. Bank of Italy, Research Area.
* giuseppe.bruno@bancaditalia.it
This version March 2023
The views expressed are the authors' only and do not imply those of the Bank of Italy.

1 Introduction and motivation

Doctrina bona dabit gratiam; in itinere contemptorum vorago.

Economic activity is based on monetary exchanges. Payment systems and infrastructures, which track the main commercial transactions in a timely and reliable manner, can therefore represent an important source of information for economic analysis and forecasting. Furthermore, the growing digitization of the economy favors the production of ever greater volumes of information based on digital technologies, capable of increasing analytical skills even in conditions of high macro-economic uncertainty. Statistical authorities and central banks are therefore gradually equipping themselves with tools aimed at allowing the use of this information, to build accurate and timely indicators to support economic policy decisions, as well as statistical measurement. However, although the link between economic activity and payment flows is based on a simple, intuitive consideration (think of the equivalence of Fisher exchanges), the use of payment data as an unconventional source of information for economic analysis could not be straightforward. In fact, payment systems and technological infrastructures were not developed by economists, but by specialists dedicated (rightly) to safeguarding the integrity, timeliness, cyber security of operations and, obviously, the business, with operating standards different from those that characterize economic statistics. The dynamics of monetary flows generally follows a cash criterion, not an accrual criterion; trends are usually erratic or highly seasonal, especially with reference to high-frequency flows, which do not always coincide, therefore, with those of the economic phenomena to be analysed.

The complexity in the use of payment data for economic-statistical analysis purposes is also linked to a series of reconciliation problems that need to be addressed ex ante. For example, if you want to analyze the purchases of retail goods and services, i.e. the transactions that payment experts define as person-to-business, the best thing would be to select transactions with payment cards; while, in the case of commercial exchanges between companies, the flows from bank transfer operations will be more representative. Nor is the point of observation irrelevant, which can be on the side of the originator of the payment (e.g. the consumer) or on the side of the recipient of the collection (e.g. the seller company), to be identified on the basis of the macroeconomic aggregate to be measured (e.g. household consumption or business added value).

The selection of the economic activity sector and the geolocation (territorial, national, to or from abroad) of the payment flow can also play an important role in directing the analysis. Granular information must also be extrapolated from complex payment messages processed by exchange and processing system managers or intermediaries. Often this information is not available from a single reporting agent, because it is dispersed among different players participating in the payment process (banks, processing companies, clearing and settlement systems) and it can be onerous as well as complicated to obtain and integrate the data in possession of each actor; in these cases it is therefore necessary to resort to proxies or assumptions based on the information available. For example, although you don't have specific data on wire orders ordered for retail purchases (P2B), the aggregate data on the flow of credit transfers exchanged by the interbank systems can in any case be correlated with the trend in consumption and provide useful signals for macroeconomic forecasts.

The paper shows some results in harnessing electronic payment transactions data for official statistical production purposes and in macroeconomic forecasts. The work is arranged in the following way. After this introduction section 2 illustrates the payment systems from which the relevant information is extracted for the construction of the payment indicators. In section 3 several case studies are presented in which it is shown that the introduction of transaction data improves the performance of the models of the Bank of Italy for short-term forecasts or for quick estimates, as well as Official statistics nowcasting and forecasting activities.

Finally section 4 provides some concluding remarks and suggestions for future research.

2 Payment systems and information bases for economic analysis

Every year the economic transactions are carried out with retail payment instruments for a value of more than 3 times the Italian GDP. Retail payments are those made by end customers (households, businesses, public administration) for the purchase of goods and services (P2B), for the payment of salaries and pension payments (B2P), in commercial exchanges between businesses (B2B) and in transfers between private individuals (so-called P2P).

Customer payment instruments (or services) concern all the methods and monetary means used for the settlement of "retail" economic transactions. They therefore include both legal (circulating) money and *alternative* instruments issued by banking and non-banking intermediaries (for example, credit card issuers) and by Poste Italiane and traditionally classified into four categories:

1. cheques;
2. wire transfers;
3. collection instructions;
4. payment cards.

The compositional relationship between fiat money and other instruments and, within the latter, between the various means of exchange (for example, cards, checks, etc.) is affected by cultural and institutional factors, linked to the degree of financial sophistication of a country. Furthermore, the evolution of the digital ecosystem and the application of information and communication technologies to the payment system determine continuous innovations of services and payment instruments that direct customer preferences regarding the use of the different instruments. These innovations concerned both the exchange processes and the interbank procedures for the electronic transmission of information (for example, electronic wire transfers, check truncation, use of cards via the Internet, etc.), and the physical media (for example, mobile applications , contact-less card, microcircuit, etc.). They have favored over time the progressive replacement of traditional paper-based payment instruments (and procedures) (cash, checks, etc.) with electronic ones (for example, payment cards, wire transfers, pre-authorized debits).

2.1 Payment transaction data source

Few years ago Bank of Italy set up a joint group with the National Institute for Statistics aiming to produce payment transaction data indicators and test their adoption in nowcasting or forecasting macroeconomic aggregates. The outcome of this activity has been the release of monthly and daily time series extracted from two clearance and settlement systems BI-COMP and TARGET2 retail. Albeit their timeliness and high frequency, these time series do not allow any commercial breakdown because they come at a high level of aggregation. Therefore using payment transaction data microdata coming from both BI-COMP (card data) and data from a major private card operator, after a phase of data munging and cleaning, derived a more granular set of monthly and daily time series. New series refer to the shares of the transaction by 10 Merchant category code groups (e.g., clothing, hotels and restaurants, home, work, retail, services, telecom, web and dm, travels and transport, cash advance, not defined) and the e-commerce transaction (from the acquirer side, namely, the expenditure of Italians on Italian firms).

Therefore, in our analysis we could apply the following four types of payment transaction data time series:

1. monthly time series of BI-COMP/T2 total transaction and by payment instruments (e.g., credit transfer, direct debit, payments cards - both Atm and Pos -, cheques, other); series are available since the

year 2000 (both in volume and amount);

2. monthly time series of BI-COMP e-commerce transaction in terms of indices (2014=100), available since 2014;
3. daily time series of BI-COMP/T2 total transaction and by payment instruments (POS, ATM), available since the year 2006 (both in volume and amount);
4. daily time series of BI-COMP domestic debit card transaction (considering only the proximity channel) by 10 MCC groups in terms of indices (2014=100), available since 2014.

3 Empirical application

3.1 Short-term GDP forecasts during the pandemic

Improving the accuracy of short-term forecasting models with new variables is a complex task given the scarcity of timely information available. In recent years there has been a strong push in institutions to use innovative data from unconventional sources (for example, data from payment systems, data collected on the Internet, data from social networks such as Twitter, and other textual data) to obtain timely and accurate signals on economic activity: see for example [7, 13, 1, 2, 3]. New challenges for modeling short-term forecasts have emerged following the outbreak of the Covid-19 pandemic, which caused a sharp recession and sometimes compromised the data production process, making forecasting extremely difficult (see [14] for example). Payment systems data are characterized by timeliness, high frequency and absence of relevant revisions. These characteristics make them ideal candidates for expanding the information available to forecasters. Furthermore, as electronic means of payment become more widespread, these data streams hold ever more promise for providing reliable signals of economic activity. In the last decade several studies have shown the importance of these data in anticipating the dynamics of macroeconomic aggregates, characterized by longer release times (see [3] and [4] for a review).

A broad debate has ignited in the literature on the most effective econometric strategies for predicting the strong fluctuations in macroeconomic variables observed during the pandemic in a robust manner and in real time (for a review, see Aprigliano, et al. 2021). An initial solution was to resort to models based on high-frequency data, including data from payment systems, which proved to be of fundamental importance in the construction of synthetic and timely indicators of the economic cycle (for the United States see Aruoba et al. 2009 [5] and Lewis et al., 2021; for Germany, see Eraslan and Gotz, 2021 [10]). For Italy, Delle Monache et al. 2021 [8] have developed a weekly indicator to measure economic activity (Italian Weekly Economic Index, ITWEI), estimated starting from a dataset including both weekly (the amount of debit card transactions at POS and withdrawals from ATMs, an index of expenditure made using credit cards, the total consumption of electricity, that of gas used, for industrial purposes, a search index on Google Trends of the term 'CIG' or 'Cassa Integrazione Guadagni', net job openings) and monthly (freight traffic flows, Purchasing Managers' Index - PMI of manufacturing and services, indicators from Confcommercio sources on expenditure for consumption of goods and services, cash flows generated by the collection of value added tax on imports, the effective value of transfers for the CIG). Payment data contributes substantially to the ITWEI estimate, assuming a very significant weight in the estimate of the main component on which the index is based [1]. ITWEI shows a clear cyclical pattern and a strong correlation with real GDP in the estimation sample from July 2010 to September 2021. During the pandemic period it was able to timely anticipate the collapse of economic activity in the first two quarters of the 2020, following the lockdown implemented in Italy to counter Covid-19, and the strong rebound in the third when many of the restrictions were lifted (see 3.1). 3.2 shows how the indicator provided highly accurate out-of-sample forecasts of Italian economic activity both in 2020 and 2021.

[1] ITWEI is estimated has the first principal component in the balanced del dataset with the Expectation-Maximization (EM) method proposed by Dempster et al. 1997 [9]

Figure 3.1: Economic activity weekly indicator (ITWEI) and GDP short term growth rate in the period January 2020 to September 2021.

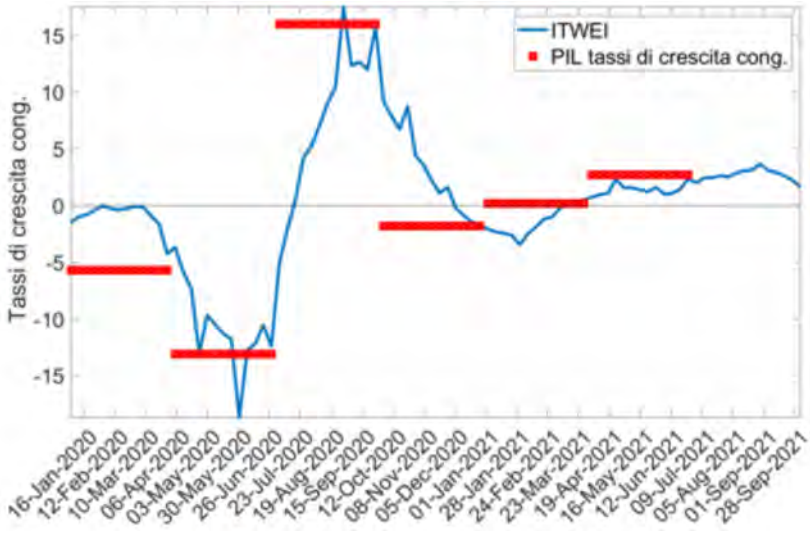
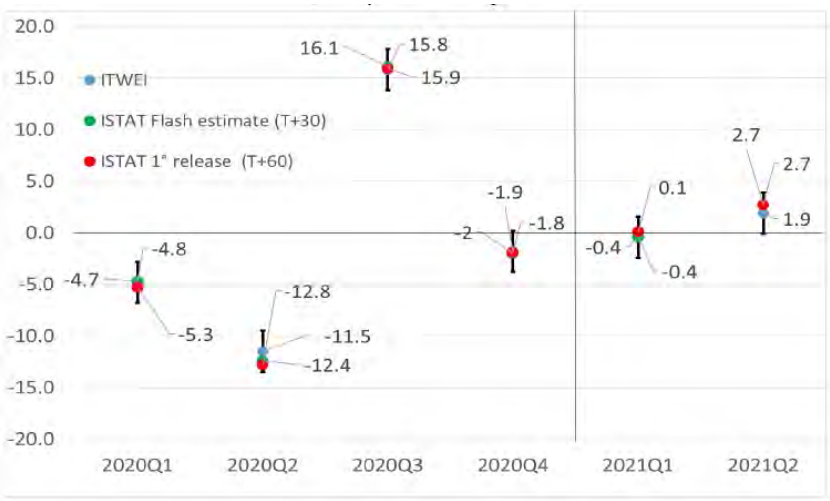


Figure 3.2: ITWEI out of sample performance during COVID-19 pandemics.



A second approach to generate robust short-term forecasts of economic activity during the pandemic resulted from the influential contribution of Ng (2021), applied by Aprigliano, Borin, Conteduca, Emiliozzi, Flaccadoro, Marchetti and Villa (2021) to the case of Italian GDP. The methodology is based on a mixed-frequency large factor model that combines macroeconomic variables, data from payment systems, epidemiological indicators and the degree of restriction of containment policies (see for example [6] and [11]). Also in this approach, the variables taken from payment systems provide an important contribution in tracing the short-term dynamics of the Italian economy. The econometric strategy used makes it possible to ”purify” the purely economic factors of the anomalous trends induced by the pandemic, making the forecasts of the model robust and more stable in the current and following quarters. The work also exploits the synergies between the econometric forecasting framework of the factor model and a sophisticated epidemic Susceptible-Infectious-Recovered (SIR) model with endogenous policy responses (see for example [15]). Despite the uncertainty induced by the numerous factors that may influence the future evolution of the pandemic, the epidemiological model is able to take into account many of the determinants of the epidemiological trends and the consequent responses in terms of containment policies.

3.2 Monthly Retail trade index and e-commerce sales index

In the context of quarterly National accounts flash estimation, retail trade indicators represent a crucial variable for estimating trade margins. The retail trade index is available monthly and it is released with a delay of around 43-50 days from the end of the month. Therefore, its information is partial for the GDP flash, which is released by ISTAT at $t + 30$ days and it requires to be supplemented by some prediction for last month. Retail trade is tracked by ISTAT through a monthly 2015=100 index detailed by type of product, employment size, and type of sales. The e-commerce index is a sub-component of total retail trade. Total retail trade is available over the period January 2000-March 2020, whereas e-commerce over a shorter period (January 2015-March 2020). We aim at verifying the size of one-step-ahead forecast errors resulting from modeling monthly retail trade index by two sets of predictors, the former provided by traditional Business surveys for firms operating in trade and the latter, more innovative, relative to payment transaction data series. Business surveys make available several statistics (monthly climate indexes, frequencies, and balances). However, after a preliminary analysis, the focus was limited to climates only. The set of electronic payment transaction data series concerns the BI-COMP/TARGET2 set of monthly series by main payment instrument (in terms of volume of transaction and in terms of value, in euro); and the BI-COMP/T2 retail components consisting of the monthly POS (BI-COMP) series in euro by nine MCC groups. Concerning the pre-treatment of BI-COMP/TARGET2 data by type of payment instrument, we removed all the outliers from the original data. Figure 1 shows these sets before (red line) and after (blue line) the outlier removal. Payment transaction data series by retail components have been backcasting to 2000 to make them homogeneous to the rest of the predictors. Finally, we computed the other two couples of series resulting from the extraction of the first two factors by principal component analysis applied respectively to the set of BI-COMP/TARGET2 by type of payment instrument and by retail component.

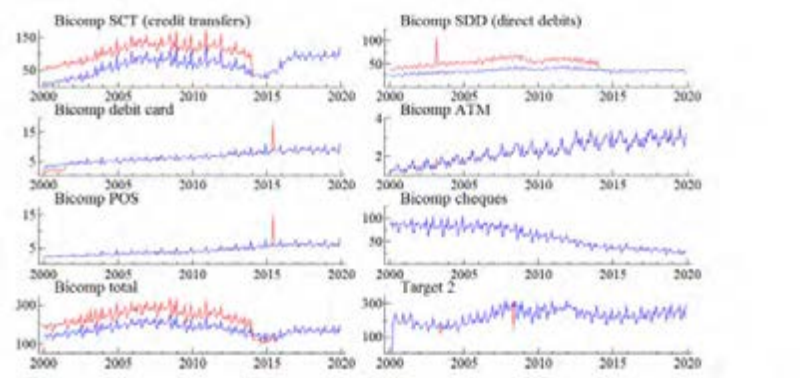


Figure 3.3: BI-COMP and TARGET2 series before (red line) and after (blue line) outlier correction.

Seasonal unadjusted data of the two retail trade indicators are shown by red lines in the two left-hand-side panels of Figure 3.4, the total on the top and the e-commerce component on the bottom. In the two right-hand-side panels, always in red, are shown the corresponding data in terms of annual growth rates. In each panel, retail trade data are accompanied by blue lines of the synthetic index of payments by retail component relative to the first-factor component, in levels for panels on the left and annual growth rates for those on the right.

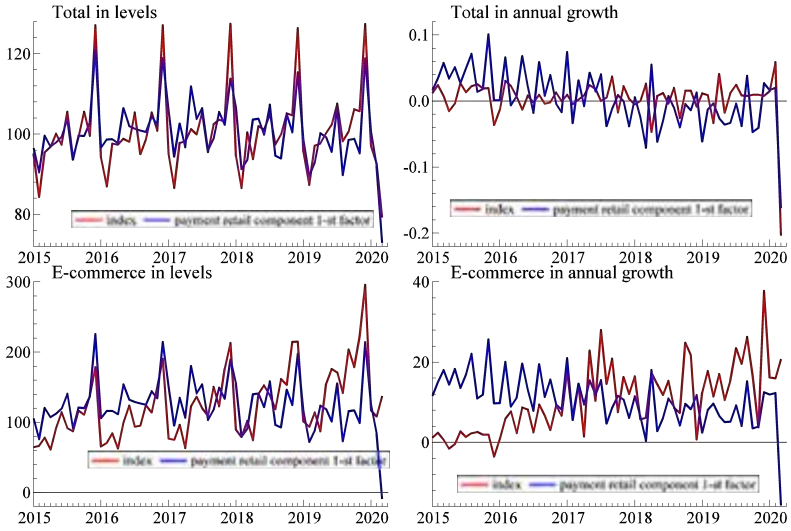


Figure 3.4: Retail trade data and synthetic index of payments by retail component

The analysis of Figure 3.4 shows that:
the first factor of payments by retail component presents higher volatility than retail trade indices, which emerge in particular over data in differences. The same evidence is even emphasized for single components of payments data (not shown here for brevity);

1. the first factor of payments by retail component fits generally better to total retail trade than the e-commerce component. This could be due to differences in the e-commerce sale boundary definition. For the construction of the Istat e-commerce index, only pure e-commerce activities are considered
2. and we only referred to the sales of goods and not services; moreover, Italian firms can sell not only to Italians but also to foreigners (even in a small share). In our payment transaction data series, we consider all the Italian firms selling only to the Italians both goods and services through a not pure e-commerce channel;
3. For both retail trade variables, the fit of same payment factor worsens moving from the comparison of data in levels (left-hand-side panels) to annual growth, with particular evidence for the e-commerce index. This could be due to both the dominance of seasonal factors in both series under comparison which is deleted out from the transformation and a higher order of the trend in the e-commerce component.

Concerning model strategy, after difficulties encountered within the multivariate Seemingly Unrelated Time Series Equation (SUTSE) system, we went for the univariate class of autoregressive distributed lag (ADL) models. This strategy has been followed both for variable selection and the setting of the final equation from which computing the forecasts. We adopted the general-to-specific (GETS) algorithm, developed into Autometrics, which is an algorithm for automatic model selection within the general-to-specific framework developed by D. Hendry (see [12]). A no-seasonal adjustment has been applied to the entire dataset, with the idea to mimic as much as possible the strategy applied in the statistical production process for which seasonal adjustment operates in a second step. The estimation sample has been set to January 2000-December 2017, whereas one-step-ahead forecasts have been computed over 27 months, from January 2018 to March 2020.

The strategy followed for the definition of the (general) unrestricted ADL model has been to start from models where the 3 sets of predictors (e.g., Business survey data; payment transaction data by type of payment instrument and by MCC groups) were all represented. In this way, we can measure the relative contribution of each set of data to the forecasting performance. For the payment transaction data series we conducted several experiments to understand if it is more favourable either the use of single components (alone, in couple, sets of three, altogether) or of the extracted factors. Other settings of the unrestricted ADL model concerned the choice of lag order of predictors (limited to 0 and 1 only), the inclusion of trading

days effects (one regressor for the number of working days, Easter and leap year effect), seasonal dummies, constant and linear trend. The presence of the lagged dependent variable among predictors has been also limited to 1. Furthermore, both models of data in levels and annual growth are tested.

The estimations are carried out by ordinary least squares (OLS).

Table 1 presents the results of estimation for the model identified by the GETS Autometrics search for total retail trade and Figure 3(a) the fitted values versus the true ones. The specification is for data in annual growth rates and turns out to be particularly parsimonious since any regressor of the payment by trade component set resulted significant, neither taken individually nor through their synthetic indices. Both lagged dependent and predictors are strongly significant. Even if trading day effects are not statistically significant we did not remove them from the final specification because they help in increasing the forecasting performance.

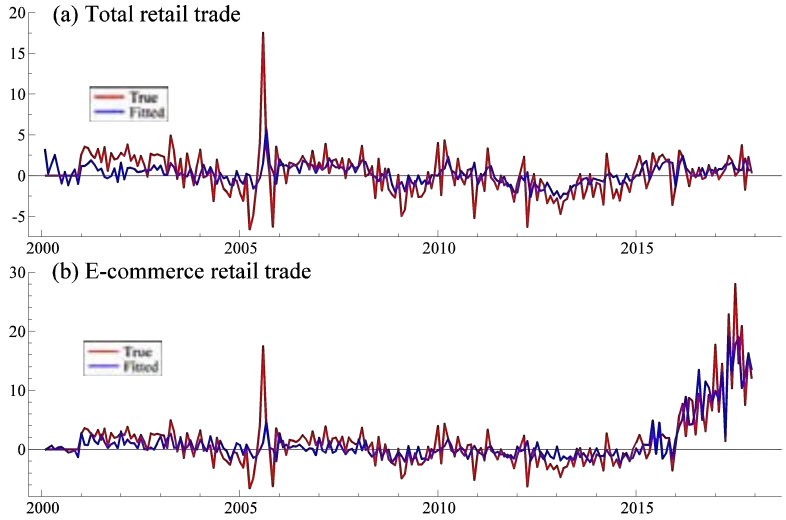


Figure 3.5: Fitted versus true values for specified ADL models

Even if the fit of the model to retail trade data is particularly effective, three negative elements emerge from Table 1: 1) the lack of payment transaction data by MCC groups within predictors, 2) the small impact of predictors measured by low values of the regression coefficient, 3) the absence of the common factor regressor for payment transaction data by type of payment instrument at lag 0 and of a negative sign at lag 1, the most promising at the beginning of the exercise. A possible explanation of this drawback is the high volatility of payments data which distorts the relationship with the dependent variable.

Table 1: ADL model for total retail trade in annual growth rates. OLS estimation results.

Predictor	Lag	Coefficient	t-students
Total retail trade	1	0.2618	3.93
BS climate index	0	0.0487	3.92
Payments by type of instrument 1-st factor	1	-0.046	-3.84
Easter effect	0	1.0157	1.74
Leap year	0	1.4961	1.27
Trading day	0	0.0984	1.78

Note: Estimation period February 2001 - December 2017.

Forecast errors in mean absolute (MAE) and mean squared (RMSE) terms are shown in Table 2, where they are computed over the annual growth rates of original data. Their values, a MAE around one for months of 2018-19, are not distant from values relative to analogous experiments conducted recently over other monthly Italian indicators (industrial production, labour force, etc.). The gains in terms of reduction of the

relative statistics once payments data are removed from the specification of Table 1 are shown in parenthesis. The gain is 12.1% in terms of MAE and 3% in terms of RMSE.

Table 2: Onestep-ahead forecasting performance of ADL model for total retail trade over the period January 2018-March 2020.

Test period	**MAE**	**RMSE**
2018:2020(1)	1.1813 (12.1%)	1.6529 (3.0%)
2018:2020(2)	1.3031 (11.7%)	1.8315 (4.5%)
2018:2020(3)	1.9582 (7.9%)	4.0727 (1.2%)

Note: In parentheses gain from the specification without payment data

As foreseeable, the forecasting performance of ADL model of Table 1 worsens when the test period extends to February and March 2020, the beginning of the Covid-19 crisis period in Italy, as provided in Table 2. Notice, in particular, the explosion of RMSE to 4.0727 when the test period spans 27 months until March 2020, the month of the lockdown of the economic activities. Results of the ADL model fitted over the monthly e-commerce sales index are in Table 3, whereas fitted versus true values are shown in Figure 3(b). The fit concerns annual growth rates over the estimation sample on the period February 2001-December 2017. As for total retail trade, Table 3 presents the best performing model obtained by the GETS Autometrics procedure.

Table 3: ADL model for the e-commerce sales index in annual growth rates. OLS estimation results.

Predictor	Lag	Coefficient	t-students
e-commerce sales index	1	0.2471	4.48
Hotels and restaurants	0	-0.16	-2.3
Hotels and restaurants	1	0.6636	8.38
Home	1	1.0157	-5.47
Service	0	1.4961	5.57
Telecommunications	0	-0.307	-4.41
Telecommunications	1	-0.1128	-2.41
Transports	1	-0.1178	-2.98
Trend	0	-0.0074	-2.69
Seasonal dummy	0	1.6089	2.23
Seasonal dummy	6	1.2383	1.77
Easter	0	1.8724	2.49
Trading day	0	0.1439	2.03

Note: Estimation period February 2001 - December 2017.

The main pieces of evidence are as follows:

1. Business survey data are not significant and therefore disappear from the final specification;
2. same evidence for payment transaction data series by type of payment instrument;
3. payment transaction data by MCC groups are here largely represented and they appear by single components instead of their synthetic factor indexes;
4. the sign of payment indicators is counter-intuitive for payment data relative to hotels and restaurant and telecommunication (lag 0). This is because of the strong volatility encountered in this set of data;
5. the presence of significant deterministic effects among selected regressors, including a linear trend, which explains the accelerating behavior of e-commerce with respect to payments. Notice also the presence of some significant seasonal components, able to explain residual seasonality and trading effects.

Moving now to the one-step-ahead forecast performance in Table 4, we see that for all sample periods considered in the experiment, both MAE and RMSE are higher than the case of total retail trade shown in Table 2 . This occurs because the experiment runs over a component of detail for the total considered before which, furthermore, features an explosive evolution in recent periods. By contrast, the gain due to the inclusion of payment transaction data by MCC groups is remarkable in this case, larger than 50% in almost all cases.

Table 4: Onestep-ahead forecasting performance of ADL model for the e-commerce sales index over the period January 2018- March 2020.

Test period	MAE	RMSE
2018:2020(1)	6.8772 (66.9%)	8.8650 (57.2%)
2018:2020(2)	6.6367 (73.6%)	8.6937 (59.8%)
2018:2020(3)	7.6178 (54.2%)	10.6501 (31.9%)

Note: In parentheses gain from the specification without payment data

3.3 Final Consumption expenditure

In this third trial, the considered target variable is the National Account (NA) aggregate referring to the quarterly final consumption expenditure (chain-linked 2015=100 and working-day adjusted). As a first exercise, an Unrestricted MIxed DAta Sampling (U-MIDAS) regression has been estimated for the time span from the 3rd quarter of 2014 to the 4th of 2019, namely before the Covid-19 Crisis. payment transaction data series by nine Merchant category code groups (for the period starting from July 2014 to December 2019) have been considered as an independent variable in the regression to evaluate the forecast performance. Because the payment transaction data variables are calculated as index number (May 2014 = 100), in the pre-treatment phase, the final consumption expenditure has been calculated as well as an index number with base 2nd quarter 2014=100. As the second step, payment transaction data and final consumption have been adjusted estimating and removing a linear or quadratic deterministic trend without removing the seasonal component.

The graphical analysis of the combination of the quarterly series of final consumption and the monthly payment transaction data series exhibited a clear correlation between them. As an example, in Figure 3.6 a plot of final consumption and the payment transaction data series referring to clothing sales and hotels (two of the best performer) is presented: the series in dots indicate the quarterly NA final consumption growth rates and the line are the growth rates for the monthly payment transaction data series for clothing sales.

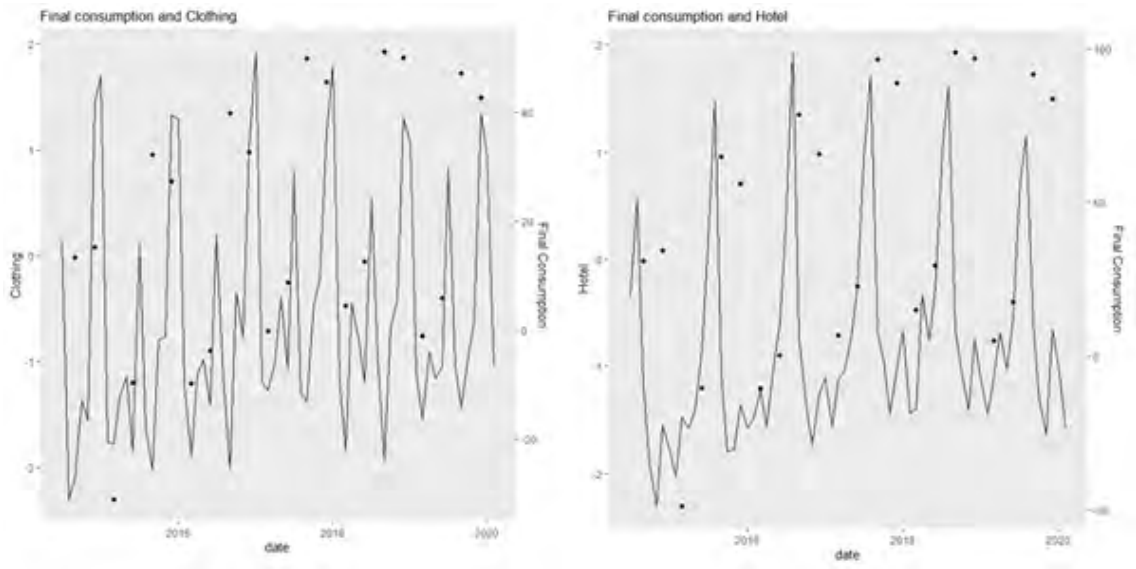


Figure 3.6: Quarterly series of final consumption and monthly payment transaction data series.

An unrestricted MIDAS model, as the one shown in formula 1, has been estimated for the period 3rd quarter 2014 to 4th 2018 :

$$y_t = \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot y_{t-2} + \beta_3 \cdot y_{t-3} + \theta_1 \cdot x_{1t} + \theta_2 \cdot x_{2t} + \theta_3 \cdot x_{3t} + \varepsilon_t \tag{1}$$

The forecasts have been evaluated out-of-sample on the period 1st quarter 2019 to 4th 2019 using a recursive and rolling window and the RMSE and the MAPE as to accuracy measures. In Table 5 the accuracy measures of our model are compared to the ones obtained estimating an auto-regressive model of order 3 as the benchmark. For values lower than one in Table 5 the predictor variable shows a better forecast performance than an AR(3) model without any predictor. Thus, results show that hotel, clothing and work series gives the best increases in forecasting performance.

	Rolling windows		**Recursive regression**	
	RMSE	**MAPE**	**RMSE**	**MAPE**
Clothing	0.99	0.87	0.91	0.84
Hotel	0.85	1.02	0.79	1.00
Food	1.57	1.06	1.77	1.15
Home	1.22	1.15	1.10	0.94
Work	0.87	0.78	0.82	0.78
Retail	1.02	0.91	1.04	0.94
Services	1.19	1.00	1.29	0.97
Communication	1.78	1.30	1.72	1.11
Transport	1.00	1.02	0.84	0.89

Table 5: Accuracy measures resulting from the MIDAS model used to forecast the Quarterly final Consumption expenditure model in the period 2019 (Q1)2019 (Q4).

The exercise has been repeated including the Covid-19 Crisis phase in the forecasting period; thus, we used the 1st quarter of 2020 for consumption and monthly data from January to March 2020 for payment transaction data series (Table 6). The same approach has been used to evaluate the forecast performance considering a recursive and rolling window and the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) as to accuracy measures. The hotel and clothing series show a better forecast performance compared to the previous exercise, particularly the improvement of 47% of the hotel series. The reason for this could be partly the autoregressive model of the benchmark, which is less able to predict the strong negative growth rates in nth1 quarter of 2020. Overall, the payment transaction data variables show an improvement thanks to their timeliness and monthly frequency that is more able to capture the strong decrease following the Covid-19 lockdown and the first signals of the consumption crisis due to the unpredictable exogenous factors.

	Rolling windows		**Recursive regression**	
	RMSE	**MAPE**	**RMSE**	**MAPE**
Clothing	0.85	0.89	0.81	0.86
Hotel	0.53	0.94	0.43	0.91
Food	1.04	1.05	1.13	1.12
Home	1.01	1.11	0.95	0.95
Work	1.01	0.85	1.00	0.85
Retail	0.95	0.93	0.92	0.95
Services	1.18	1.03	1.06	0.99
Communication	1.00	1.22	0.97	1.07
Transport	0.97	1.01	0.96	0.92

Table 6: Accuracy measures resulting from the MIDAS model used to forecast the Quarterly final Consumption expenditure model in the period 2019 (Q1) 2020 (Q1).

As a final exercise, we calculated a forecast combination of the nine payment transaction data series using as weights the in-sample root mean square error (RMSE) for the period from the 3rd quarter of 2014 to the 4th quarter of 2018. The combination led to a reduction in terms of RMSE of the forecast for the period 1st quarter 2019 to the 1st 2020 equal to 20% compared to the benchmark model (Autoregressive model of order 3).

4 Concluding Remarks

The empirical analyses carried out in this work show the relevance of payment transaction data in short-term macroeconomic forecasts and the production of official statistics. The series of electronic transactions (for business reasons, travel and transport, hotels and restaurants) recorded during the pandemic period have reduced the forecast error of expenditure for quarterly final consumption by more than 50% in some product sectors and have decreased the nowcasting forecast error of the total retail sales index by about 6% and that of the e-commerce index by about 10 percentage points.

The main advantages of using such sources of information certainly lie in their timeliness, coverage and level of detail they provide. However, some key challenges have to be overcome in order to achieve a systematic use of this information in official statistics. These challenges are essentially linked to the difficulty of accessing data, both at individual intermediaries and at private infrastructure managers, as well as the complexity and lack of standards in the IT protocols employed by the different technological platforms.

Furthermore, the preliminary phases of selection and treatment of the variables are also crucial, which have been accounted for in some contributions of this work. In our experiments, we could use only some aggregated payment transaction data series to assess how much payments-system data can contribute to the production of more accurate nowcasts/forecast of macroeconomic aggregates in Official statistics. Preliminary results look promising - although the time series are quite short for an accurate forecast -, we keep

on investigating and improving the payment transaction data time series to employ this information into Istat forecasting models. In our first trial, the automatic General To Specific Selection Procedure for the selection of the best performer series in the ADL model makes new payment transaction data series work in both models and we observed a good fit especially for the period before the Covid-19 Crisis. The gains in terms of reduction of the relative statistics are 12.1% for the MAE in the case of the retail trade index, and 67-73% for the e-commerce. Considering also the Covid-19 period of the lockdown of economic activities (March 2020), the payment transaction data series still guarantees gains in the forecasting performance of the model that is still larger than 50% in terms of MAE in the case of e-commerce sales. In our second trial on the forecasting of the quarterly final consumption expenditure, an unrestricted MIDAS model has been estimated from the 3rd quarter 2014 to 4th 2018 and evaluated both for the period 1st quarter 2019 to 4th 2019 and the period 1st quarter 2019 to 1st 2020. We compared the accuracy measures of our model to the ones obtained estimating an auto-regressive model of order 3 as the benchmark. In the case of the forecasting out of sample before the Covid-19 period, results show that hotel and work reduce the forecast RMSE of the AR3 respectively 15% and 13%. When including the Covid-19 Crisis phase in the forecasting period (1st quarter 2020), the hotel and clothing series show a further improvement in the forecasting performance (47% of reduction in term of RMSE), being the timeliness of payment transaction data series able to capture the determined strong decrease in consumption. The retail and transport series also produced an improvement, although less relevant (less than 5%). Even using a weighted combination of the nine payment transaction data series in the model, the reduction in terms of RMSE is equal to 20% compared to the benchmark model. Considering these positive results, we will continue to investigate the relevance of payments-system data and the way to employ this information into nowcasting and forecasting models. These are still largely experimental activities, which will require further analytical investigations and cooperation with other experts from the institutions and the academy. Finally it is important to stress the need for an extended and well designed experimentation before we will be able to provide a full fledged contribution to the production of official statistics.

References

[1] Cristina Angelico, Juri Marcucci, Marcello Miccoli, and Filippo Quarta. Can We Measure Inflation Expectations using Twitter? *Bank of Italy Mimeo*, 2018.

[2] Valentina Aprigliano, Guerino Ardizzi, and Libero Monteforte. Using Payment System Data to Forecast Economic Activity. *International Journal of Forecasting*, 15(4):55–80, 2019.

[3] Valentina Aprigliano, Guerino Ardizzi, Alessia Cassetta, Alessandro Cavallero, Simone Emiliozzi, Alessandro Gambini, Nazzareno Renzi, and Roberta Zizza. Exploiting payments to track italian economic activity: the experience at Banca d'Italia. *Bank of Italy Occasional Papers*, 2021.

[4] Guerino Ardizzi, Andrea Nobili, and Giorgia Rocco. A game changer in payment habits: evidence from daily data during a pandemic. *Bank of Italy Occasional Papers*, 591, 2020.

[5] S.B. Aruoba, F.X. Diebold, and Scotti C. Real-time measurement of business conditions. *Journal of Business and Economics Statistics*, 27(4):417–427, 2009.

[6] Francesco Paolo Conteduca. Measuring Covid-19 Restrictions in Italy during the Second-Wave. *Bank of Italy Covid-19 Note*, 2021.

[7] Francesco D'Amuri and Juri Marcucci. The Predictive Power of Google Searches in Forecasting US unemployment. *International Journal of Forecasting*, 33:801–816, 2017.

[8] Davide Delle Monache, Simone Emiliozzi, and Andrea Nobili. Tracking economic growth during the covid-19: a weekly indicator for italy. *Covid-19 Note*, 2021.

[9] A.P. Dempster, N.M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm,. *Journal of Royal Statistical Society*, 39:1–22, 1977.

[10] S Eraslan and Götz T. An unconventional weekly economic activity index for germany. *Economic Letters*, 204, 2021.

[11] Thomas Hale, Rafael Goldszmit, Beatrix Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, and Helen Tatlow. A global panel database of pandemic policies. *Nature Human Behaviour*, 5(4):529–538, 2021.

[12] David Forbes Hendry. Predictive Failure and Econometric Modelling, Macroeconomics: The Transactions Demand for Money. *Economic Modelling, Heinemann*, pages 217–242, 1979.

[13] Michele Loberto, Andrea Luciani, and Marco Pangallo. The Potential of Big Housing Data: an Application to the Italian real-estate Market. *Bank of Italy working papers*, 1171, 2018.

[14] Alberto Locarno and Roberta Zizza. Previsioni ai tempi del Coronavirus. *Bank of Italy Covid-19 Note*, 2020.

[15] Sabina Marchetti, Alessandro Borin, Francesco Paolo Conteduca, Giuseppe Ilardi, Giorgio Guzzetta, Paolo Poletti, Patrizio Pezzotti, Antonino Bella, Paola Stefanelli, Flavia Riccardo, Stefano Merler, Andrea Brandolini, and Silvio Brusaferro. An epidemic model for SARS-cov-2 with self-adaptive containement measures. *Bank of Italy Occasional Papers*, 681, 2022.

Payment Microdata for Macroeconomic Forecasts

Guerino Ardizzi[1], Giuseppe Bruno[2], Roberto Iannaccone[3],
Juri Marcucci[2], Filippo Moauro[3], Alessandra Righi[3], Davide Zurlo[3]

[1]Money circulation Directorate; [2]Economics and statistics Directorate,
Bank of Italy;
[3]National Statistical Office, ISTAT.

Granular data: new horizons and challenges for central banks.
BoC-IFC Satellite seminar.
Bank of Canada, Ottawa, July 15th

The views expressed here are the authors' only and do not imply those of the Bank of Italy.

Outline

1. Motivation
2. Payment systems and Economic analysis
   - Short-term GDP forecasts during the pandemic
   - Monthly Retail trade index and e-commerce sales index
   - Final Consumption expenditure
3. Concluding Remarks

Statistical value of payment data

Computing the main components for Gross Domestic Product (GDP) and other relevant macroeconomic figures is traditionally based on surveys among household and firms.

- surveys are necessarily limited in sample size;
- they are usually very expensive;
- natural phenomena like the Covid-19 pandemics could wipe out the possibility to run these surveys;

Statistical value of payment data

Albeit the close relationship between payment flows and economic activity, the employment of payment data for the macroeconomic analysis is not straightforward.

Payment systems and technological infrastructures have been developed with the goal of protecting the integrity, the speed and the security of the involved computing platforms.

Statistical value of payment data

Albeit the close relationship between payment flows and economic activity, the employment of payment data for the macroeconomic analysis is not straightforward.
Payment systems and technological infrastructures have been developed with the goal of protecting the integrity, the speed and the security of the involved computing platforms.

The payment instruments in Italy

The Bank of Italy promotes the reliability and efficiency of payment instruments in order to preserve confidence in money and non-cash alternatives.

Every year the economic transactions are carried out with retail payment instruments for a value of more than 3 times the Italian GDP.

1. cheques;
2. wire transfers;
3. collection instructions;
4. debit/credit and prepaid cards.

The payment instruments in Italy

The Bank of Italy promotes the reliability and efficiency of payment instruments in order to preserve confidence in money and non-cash alternatives.
Every year the economic transactions are carried out with retail payment instruments for a value of more than 3 times the Italian GDP.

1. cheques;
2. wire transfers;
3. collection instructions;
4. debit/credit and prepaid cards.

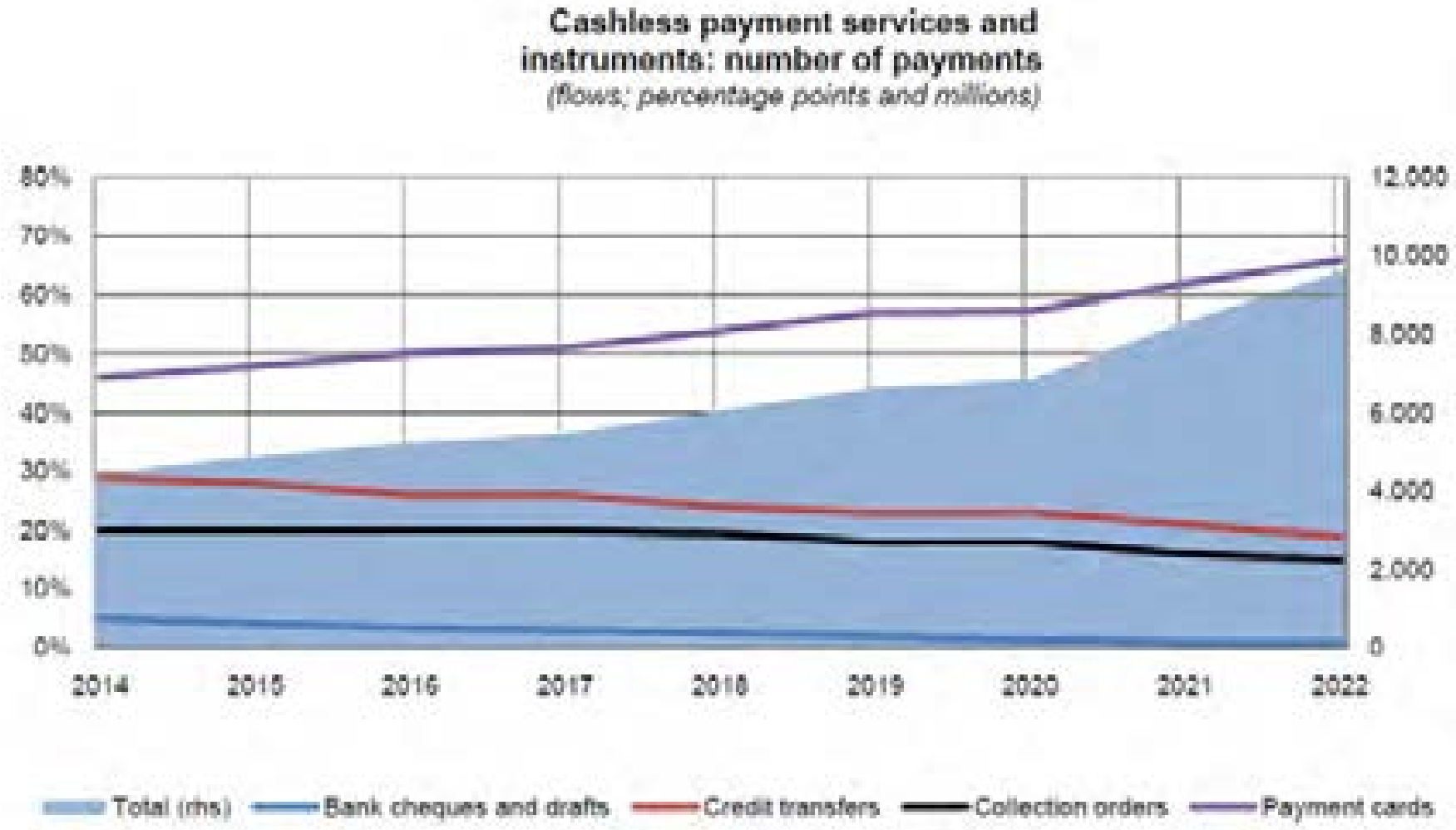
The payment instruments in Italy

The Bank of Italy promotes the reliability and efficiency of payment instruments in order to preserve confidence in money and non-cash alternatives.
Every year the economic transactions are carried out with retail payment instruments for a value of more than 3 times the Italian GDP.

1. cheques;
2. wire transfers;
3. collection instructions;
4. debit/credit and prepaid cards.

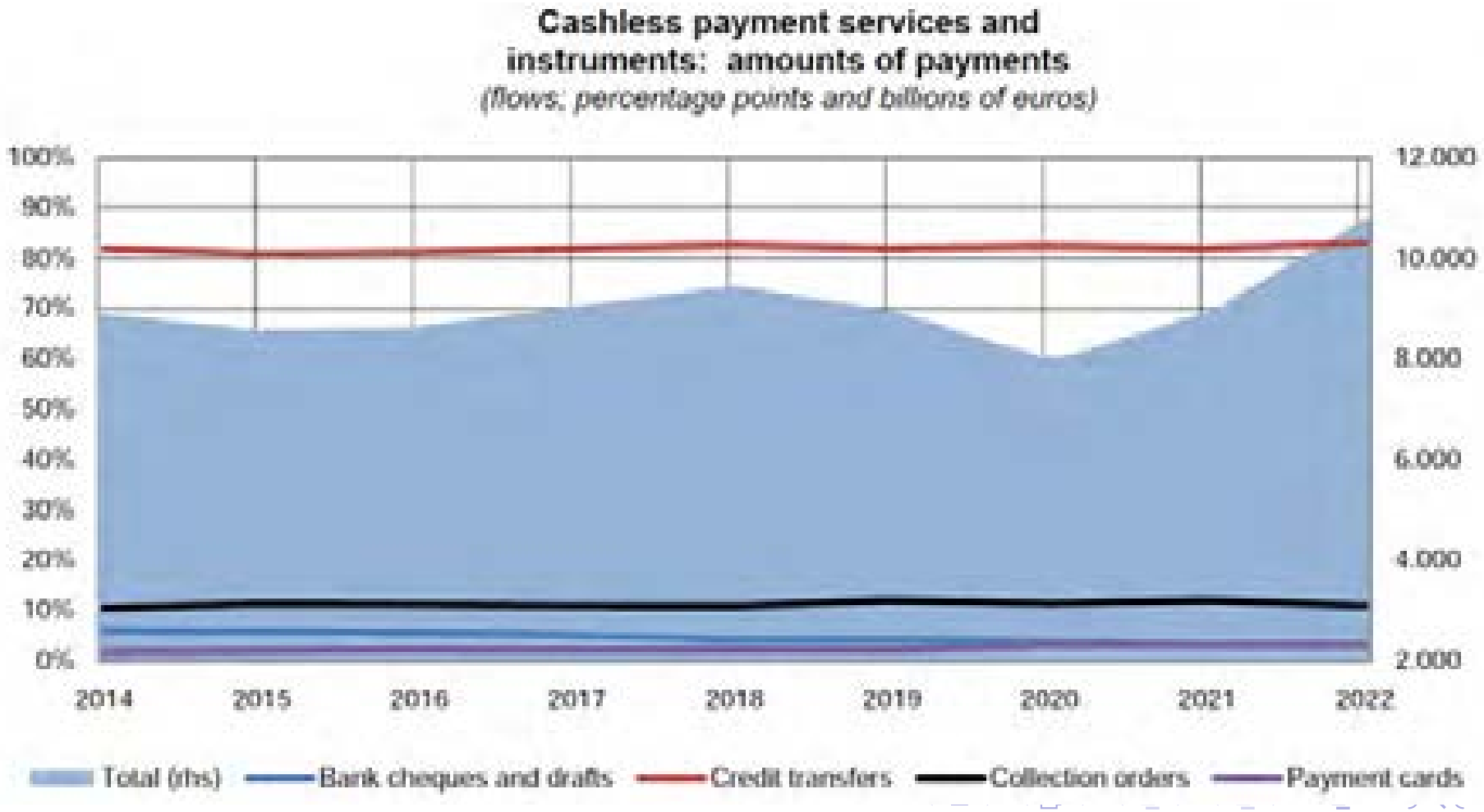
The distribution among the instruments: Volumes

The evolution of the volumes of payment instruments in Italy



The distribution among the instruments:Values

The evolution of the values of payment instruments in Italy



Banks vs digital platforms

Bank of Italy manages and oversees the national retail payment system BI-COMP and cooperates along with Bundesbank and Banque de France in the management and oversight of TARGET2, the Eurosystem Large-Value Payment Systems (LVPS).

The information collected from these two systems has played an essential role in building the databases employed for our macroeconomic analysis.

These data are available at high frequency and very timely, but . . . .

They are not granular in the commodity dimension.

Banks vs digital platforms

Bank of Italy manages and oversees the national retail payment system BI-COMP and cooperates along with Bundesbank and Banque de France in the management and oversight of TARGET2, the Eurosystem Large-Value Payment Systems (LVPS).
The information collected from these two systems has played an essential role in building the databases employed for our macroeconomic analysis.
These data are available at high frequency and very timely, but . . . .

They are not granular in the commodity dimension.

Banks vs digital platforms

We do have two possible data sources:

- agreements with the banks;
- agreements with payment technological platform.

We followed the easiest avenue: the technological platform



Nexi is present in Europe with around 170 m cards and 2.2 m merchants.
In Italy Nexi has around 41 m cards and 890 K merchants.

Banks vs digital platforms

We do have two possible data sources:

- agreements with the banks;
- agreements with payment technological platform.

We followed the easiest avenue: the technological platform

Nexi is present in Europe with around 170 m cards and 2.2 m merchants.
In Italy Nexi has around 41 m cards and 890 K merchants.

data exchange agreements

We have made an informal agreement with Nexi

- agreements on avoiding release of any confidential information;
- data granularity in order to prevent any privacy breach;
- data frequency;
- updating policy;

There is the need of internal agreements for making data available to different department

- a light dbms in postgreSQL;
- data authorization in a centralised fashion;
- user driven data analytics with open source back-ends;
- Big Data paradigms for data storing (**parquet**) and data processing (**Spark**).

data exchange agreements

We have made an informal agreement with Nexi

- agreements on avoiding release of any confidential information;
- data granularity in order to prevent any privacy breach;
- data frequency;
- updating policy;

There is the need of internal agreements for making data available to different department

- a light dbms in postgreSQL;
- data authorization in a centralised fashion;
- user driven data analytics with open source back-ends;
- Big Data paradigms for data storing (**parquet**) and data processing (**Spark**).

data exchange agreements

We have made an informal agreement with Nexi

- agreements on avoiding release of any confidential information;
- data granularity in order to prevent any privacy breach;
- data frequency;
- updating policy;

There is the need of internal agreements for making data available to different department

- a light dbms in postgreSQL;
- data authorization in a centralised fashion;
- user driven data analytics with open source back-ends;
- Big Data paradigms for data storing (**parquet**) and data processing (**Spark**).

Outline

1 Motivation

2 Payment systems and Economic analysis
- Short-term GDP forecasts during the pandemic
- Monthly Retail trade index and e-commerce sales index
- Final Consumption expenditure

3 Concluding Remarks

Payment data for GDP forecast

Improving the accuracy of GDP short-term forecasting models is of the utmost relevance in high volatility times. Introducing new variables is a complex task given the scarcity of timely information available.

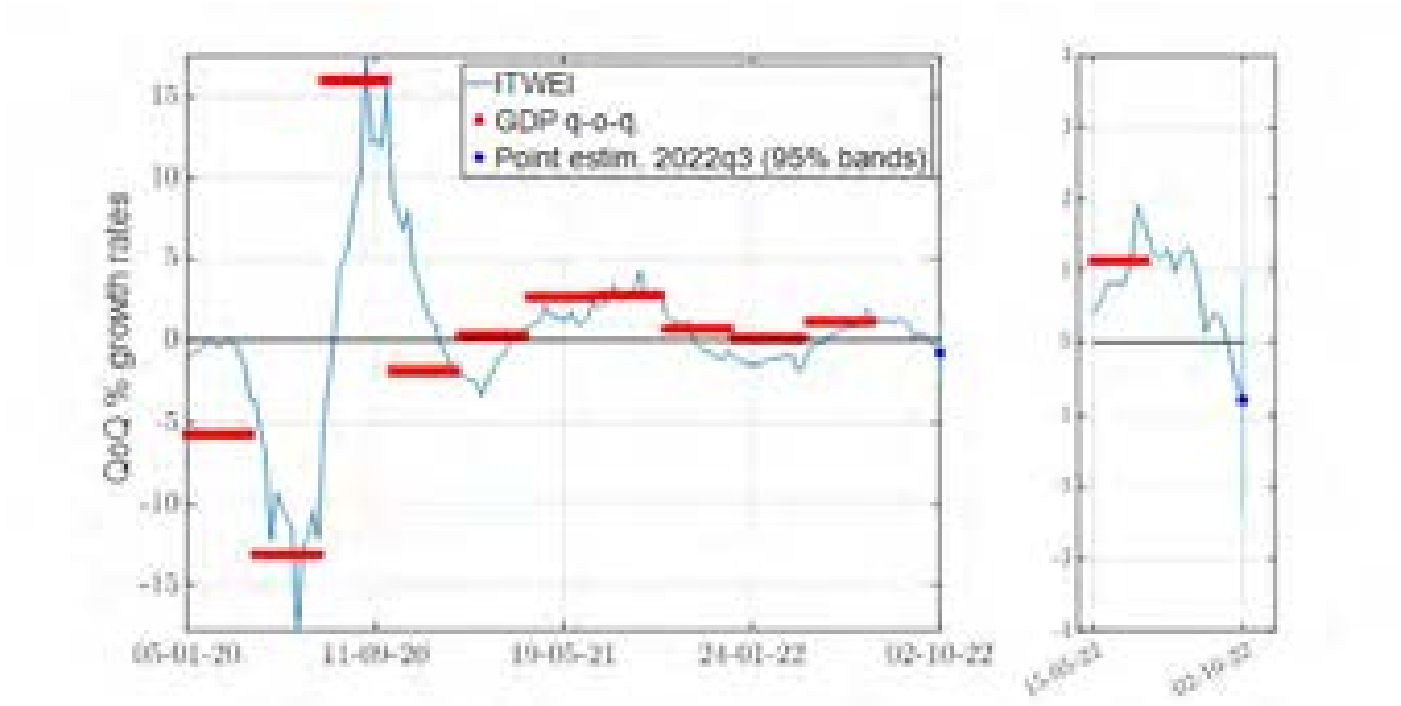
In recent years there has been a strong push in institutions to use innovative data from unconventional sources (for example, data from payment systems, data collected on the Internet, data from social networks such as Twitter, and other textual data)

Payment data for GDP forecast

Improving the accuracy of GDP short-term forecasting models is of the utmost relevance in high volatility times. Introducing new variables is a complex task given the scarcity of timely information available.
In recent years there has been a strong push in institutions to use innovative data from unconventional sources (for example, data from payment systems, data collected on the Internet, data from social networks such as Twitter, and other textual data)

Italian GDP forecast

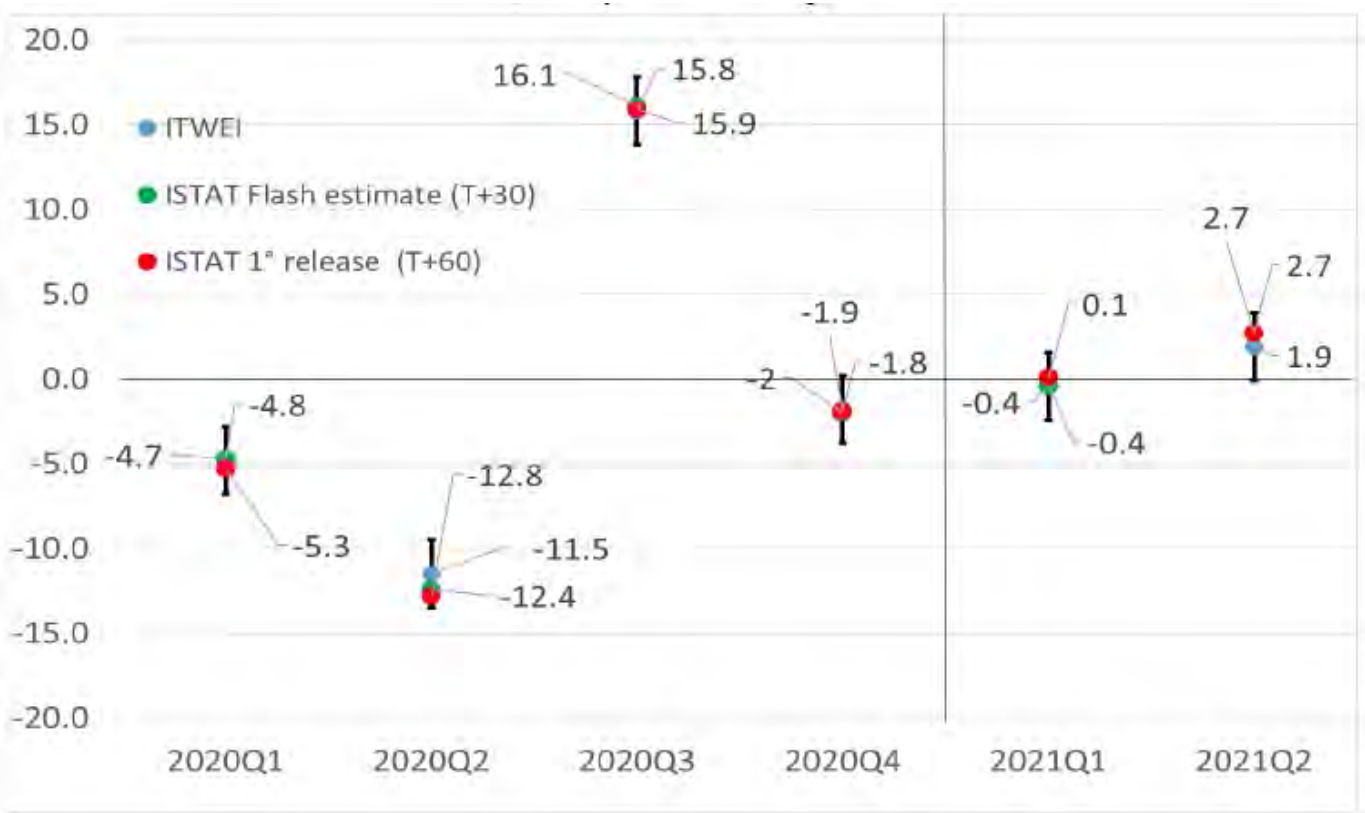
The evolution of the payment instruments in the last years in Italy



Italian weekly economic activity indicator and GDP growth rate from January 2020 to October 2022

GDP forecast

During the pandemics, ITWEI allowed us to anticipate the collapse of economic activity in 2020 Q1 and Q2, following the lockdown implemented in Italy to counter Covid-19, and the strong rebound in Q3 when many of the restrictions were lifted.



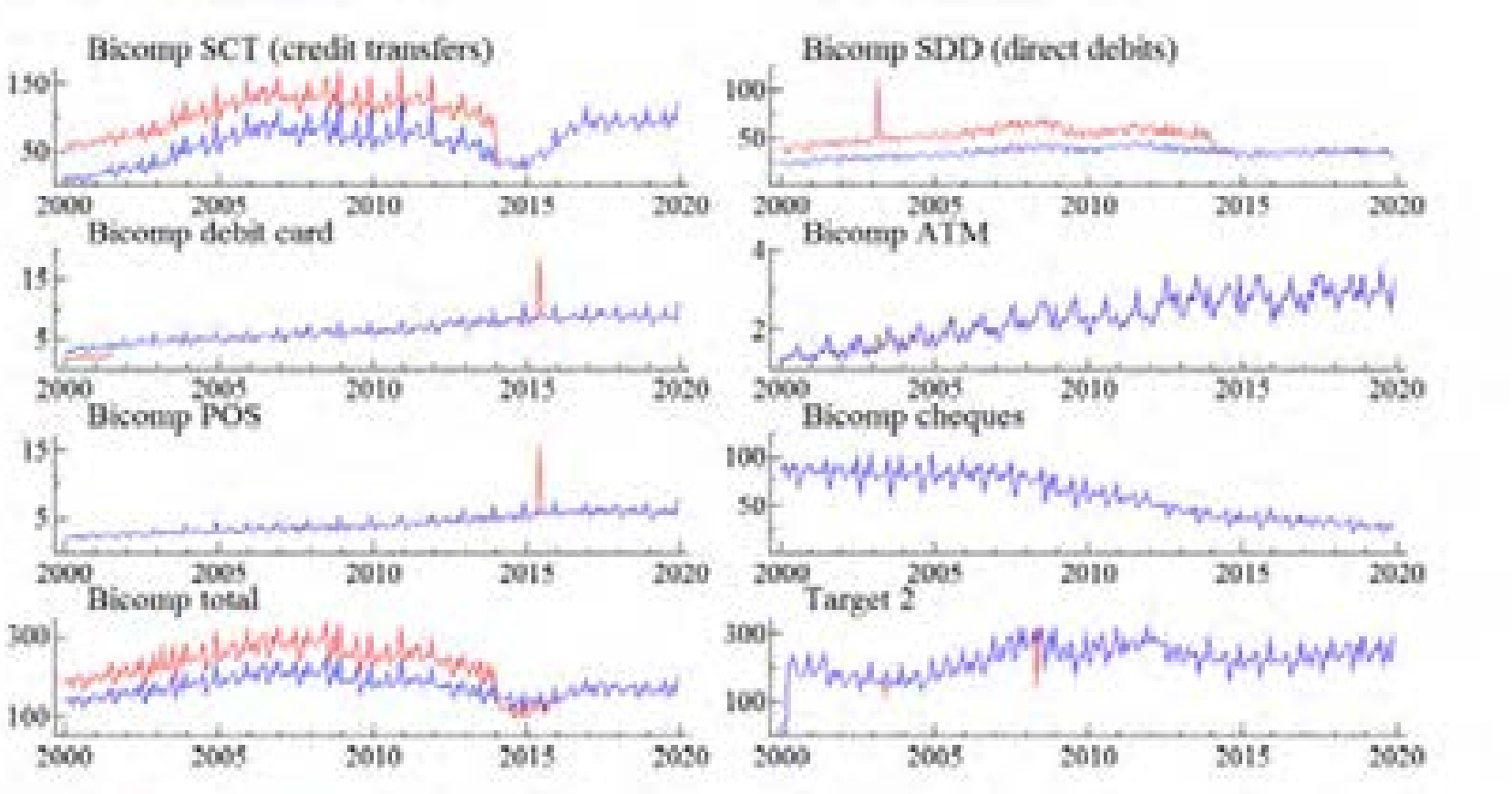
ITWEI out of sample performance in GDP forecast during COVID-19 pandemics

Outline

1. Motivation
2. Payment systems and Economic analysis
   - Short-term GDP forecasts during the pandemic
   - Monthly Retail trade index and e-commerce sales index
   - Final Consumption expenditure
3. Concluding Remarks

Monthly retail trade index

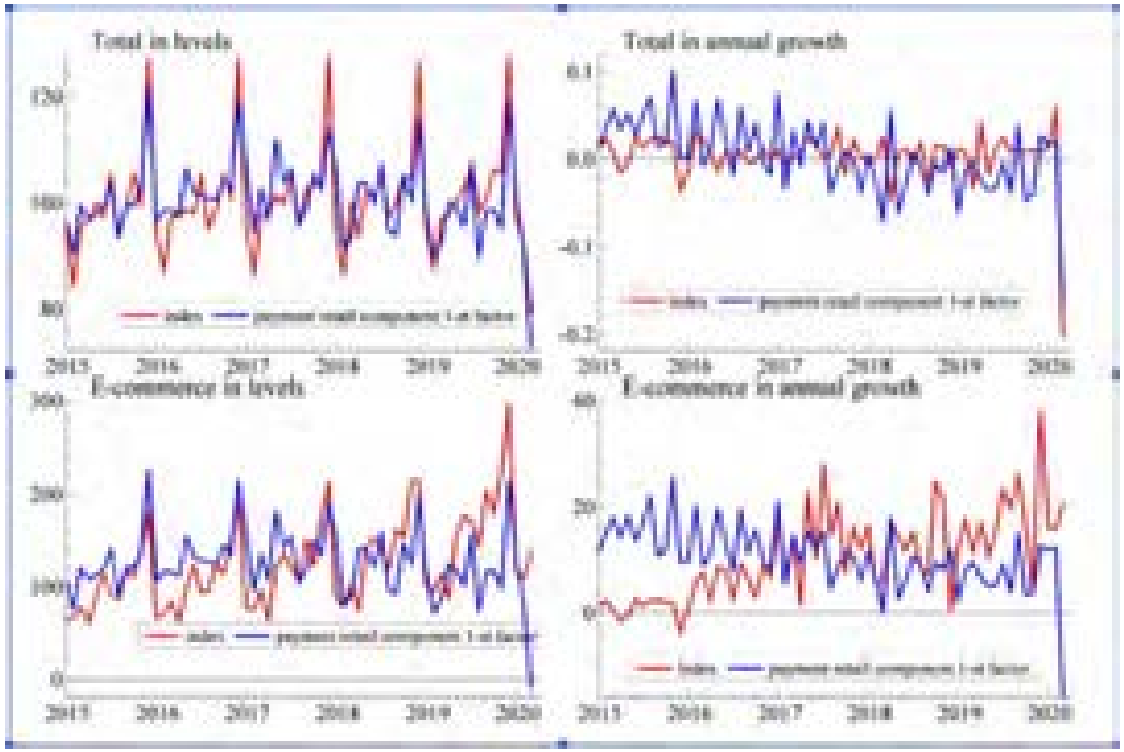
In the context of quarterly National accounts flash estimation, retail trade indicators represent a crucial variable for estimating trade margins.



BI-COMP and TARGET2 series before (red line) and after (blue line) outlier correction

Monthly retail trade index

Comparison between retail trade index series total and e-commerce and POS payments, 2014-2021 (levels and yearly variational trend).



Retail trade data total and e-commerce

Outline

1. Motivation
2. Payment systems and Economic analysis
   - Short-term GDP forecasts during the pandemic
   - Monthly Retail trade index and e-commerce sales index
   - Final Consumption expenditure
3. Concluding Remarks

The distributed lag model

Here the considered target variable is the National Account aggregate of the quarterly final consumption expenditure.

We estimated an Unrestricted MIxed DAta Sampling (U–MIDAS) regression from 2014 the 3rd to the 4th quarter of 2018:

$$y_0 = \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot y_{t-2} + \beta_3 \cdot y_{t-3} + \theta_1 \cdot x_{1t} + \theta_2 \cdot x_{2t} + \theta_3 \cdot x_{2t} + \varepsilon_t$$

The forecasts have been evaluated on the period 1st to 4th of 2019 using both a recursive and rolling window and the RMSE and the MAPE as to accuracy measures

The distributed lag model

Here the considered target variable is the National Account aggregate of the quarterly final consumption expenditure.

We estimated an Unrestricted MIxed DAta Sampling (U–MIDAS) regression from 2014 the 3rd to the 4th quarter of 2018:

$$y_0 = \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot y_{t-2} + \beta_3 \cdot y_{t-3} + \theta_1 \cdot x_{1t} + \theta_2 \cdot x_{2t} + \theta_3 \cdot x_{2t} + \varepsilon_t$$

The forecasts have been evaluated on the period 1st to 4th of 2019 using both a recursive and rolling window and the RMSE and the MAPE as to accuracy measures

The distributed lag model

Here the considered target variable is the National Account aggregate of the quarterly final consumption expenditure.

We estimated an Unrestricted MIxed DAta Sampling (U–MIDAS) regression from 2014 the 3rd to the 4th quarter of 2018:

$$y_0 = \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot y_{t-2} + \beta_3 \cdot y_{t-3} + \theta_1 \cdot x_{1t} + \theta_2 \cdot x_{2t} + \theta_3 \cdot x_{2t} + \varepsilon_t$$

The forecasts have been evaluated on the period 1st to 4th of 2019 using both a recursive and rolling window and the RMSE and the MAPE as to accuracy measures

The statistical value of the expenditures

We compare the total RMSE and MAPE of the previous model with a simple 3–period distributed lag of the final consumption

	Rolling windows		**Recursive regression**	
	RMSE	**MAPE**	**RMSE**	**MAPE**
Clothing	0.99	0.87	0.91	0.84
Hotel	0.85	1.02	0.79	1.00
Food	1.57	1.06	1.77	1.15
Home	1.22	1.15	1.10	0.94
Work	0.87	0.78	0.82	0.78
Retail	1.02	0.91	1.04	0.94
Services	1.19	1.00	1.29	0.97
Communication	1.78	1.30	1.72	1.11
Transport	1.00	1.02	0.84	0.89

Accuracy measures resulting from the MIDAS model used to forecast the Quarterly final Consumption expenditure model in the period 2019 (Q1) 2019 (Q4).

The forecast error for total retail index

We compare the total RMSE and MAPE of the one step-ahead forecast error

Testing range	Mean Absolute Error (MAE)	Root Mean Square Error
2020(1) - 2020(6)	10.1414 (-5.5%)	14.0314 (-5.2%)
2020(1) - 2020(12)	6.7623 (-5.9%)	10.3337 (-5.6%)
2020(1) - 2021(6)	9.1654 (-1.7%)	12.9199 (-0.1%)
2020(1) - 2021(9)	8.4064 (-1.0%)	12.0768 (0.0%)

One step ahead forecast error for the total index of retail sale. In parentheses the percentage loss without payment data.

The forecast error for e-commerce index

We compare the total RMSE and MAPE of the one step-ahead forecast error

Testing range	Mean Absolute Error (MAE)	Root Mean Square Error
2020(1) - 2020(6)	16.7898 (-6.6%)	21.1267 (-9.4%)
2020(1) - 2020(12)	20.6482 (-10.2%)	24.9794 (-9.8%)
2020(1) - 2021(6)	20.9022 (-8.3%)	24.3784 (-9.0%)
2020(1) - 2021(9)	18.8256 (-9.2%)	22.7444 (-9.3%)

One step ahead forecast error for the e-commerce index. In parentheses the percentage loss without payment data.

Concluding remarks, part 1

The empirical analyses carried out in this work show the relevance of payment transaction data in short-term macroeconomic forecasts and the production of official statistics. Main advantages in employing payment data:

- timely availability;
- wide geographic coverage;
- highly detailed information on merchant category code

There are disadvantages to deal with as well:

- selecting a good set of metadata to prevent privacy breach;
- presence of different seasonal components;
- susceptibility to the market share variability of the payment platform.

Concluding remarks, part 1

The empirical analyses carried out in this work show the relevance of payment transaction data in short-term macroeconomic forecasts and the production of official statistics. Main advantages in employing payment data:

- timely availability;
- wide geographic coverage;
- highly detailed information on merchant category code

There are disadvantages to deal with as well:

- selecting a good set of metadata to prevent privacy breach;
- presence of different seasonal components;
- susceptibility to the market share variability of the payment platform.

Concluding remarks, part 2

- The series of electronic transactions (for business reasons, travel and transport, hotels and restaurants) recorded during the pandemic period have reduced the forecast error of expenditure for quarterly final consumption by more than 50% in some product sectors;
- in the *nowcasts* the forecast error of the total retail sales index improves by about 6% and that of the e–commerce index by about 10 percentage points.

Concluding remarks, part 2

- The series of electronic transactions (for business reasons, travel and transport, hotels and restaurants) recorded during the pandemic period have reduced the forecast error of expenditure for quarterly final consumption by more than 50% in some product sectors;
- in the *nowcasts* the forecast error of the total retail sales index improves by about 6% and that of the e–commerce index by about 10 percentage points.

Discussion

Thank you very much for your Questions

Merci beaucoup pour vos Questions



IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Retail credit estimates in Federal Reserve Board's G.19 "Consumer Credit" Statistical Release[1]

Michael Chernousov,
Federal Reserve Board of Governors

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

The views in this presentation are those of the author and do not necessarily reflect those of the Board of Governors of the Federal Reserve System or its staff.

Retail Credit Estimates in Federal Reserve Board's G.19 "Consumer Credit" Statistical Release

Michael M. Chernousov

Federal Reserve Board

Bank of Canada, July 15, 2023

Contents

- **The G.19 "Consumer Credit" Statistical Release**

Contents

- **The G.19 "Consumer Credit" Statistical Release**
- **Old Method**

Contents

- **The G.19 "Consumer Credit" Statistical Release**
- **Old Method**
- **Challenge**

Contents

- **The G.19 "Consumer Credit" Statistical Release**
- **Old Method**
- **Challenge**
- **Solution**

Contents

- **The G.19 "Consumer Credit" Statistical Release**
- **Old Method**
- **Challenge**
- **Solution**
- **New Estimates**

Contents

- **The G.19 "Consumer Credit" Statistical Release**
- **Old Method**
- **Challenge**
- **Solution**
- **New Estimates**
- **Implementation**

Contents

- **The G.19 "Consumer Credit" Statistical Release**
- **Old Method**
- **Challenge**
- **Solution**
- **New Estimates**
- **Implementation**
- **Conclusion**

G.19 "Consumer Credit"

- **G.19 "Consumer Credit" statistical release**
  - Provides timely data on several aspects of consumer credit:
    - Aggregate consumer credit outstanding and change relative to previous month
    - Interest rates on loans at commercial banks
    - Terms on new car loans at auto finance companies
  - Is a Primary Federal Economic Indicator for U.S. economy
  - Data receive media attention and are used in economic policy and research

G.19 "Consumer Credit"

- **Data sources**
  - We cover receivables for six types of lenders

G.19 "Consumer Credit"

- **Data sources**
  - We cover receivables for six types of lenders
    - Depository Institutions

G.19 "Consumer Credit"

- **Data sources**
  - We cover receivables for six types of lenders
    - Depository Institutions
    - Finance Companies

G.19 "Consumer Credit"

- **Data sources**
  - We cover receivables for six types of lenders
    - Depository Institutions
    - Finance Companies
    - Credit Unions

G.19 "Consumer Credit"

- **Data sources**
  - We cover receivables for six types of lenders
    - Depository Institutions
    - Finance Companies
    - Credit Unions
    - Federal Government

G.19 "Consumer Credit"

- **Data sources**
  - We cover receivables for six types of lenders
    - Depository Institutions
    - Finance Companies
    - Credit Unions
    - Federal Government
    - Nonprofit and Educational Institutions

G.19 "Consumer Credit"

- **Data sources**
  - We cover receivables for six types of lenders
    - Depository Institutions
    - Finance Companies
    - Credit Unions
    - Federal Government
    - Nonprofit and Educational Institutions
    - **Nonfinancial Business**

G.19 "Consumer Credit"

- **Data sources**
  - We cover receivables for six types of lenders
    - Depository Institutions
    - Finance Companies
    - Credit Unions
    - Federal Government
    - Nonprofit and Educational Institutions
    - **Nonfinancial Business**
  - Nonfinancial Business:
    Institutions that hold consumer credit receivables for whom loans and leases DO NOT make up 50 percent or more of their total assets.

G.19 "Consumer Credit"

FEDERAL RESERVE statistical release

G.19 — Consumer Credit — April 2023 — For release at 3 p.m. (Eastern Time) June 7, 2023

In April, consumer credit increased at a seasonally adjusted annual rate of 6.7 percent. Revolving credit increased at an annual rate of 13.1 percent, while nonrevolving credit increased at an annual rate of 3.2 percent.

Consumer Credit Outstanding[1]
Seasonally adjusted. Billions of dollars except as noted

	2018	2019	2020	2021[r]	2022[r]	2022 Q1[r]	2022 Q2	2022 Q3	2022 Q4[r]	2023 Q1[r]	2023 Feb[r]	2023 Mar[r]	2023 Apr[p]
Total percent change (annual rate)[2]	4.5	4.6	-0.3	5.9	8.0	8.2	8.2	7.3	7.4	4.3	2.0	5.7	5.7
Revolving	3.7	3.6	-11.2	6.9	10.4	16.8	14.3	10.2	14.1	6.3	3.0	14.6	13.1
Nonrevolving[3]	4.8	5.0	3.5	5.6	6.7	5.6	6.2	5.4	5.2	2.6	1.7	2.7	3.2
Total flow (annual rate)[2,4]	172.7	185.1	-12.0	246.0	354.9	365.1	368.6	336.9	349.0	204.5	98.0	274.9	276.1
Revolving	37.3	38.1	-132.1	67.5	160.3	175.3	154.9	148.3	163.3	111.4	36.4	172.7	161.7
Nonrevolving[3]	135.3	147.0	110.1	178.3	194.6	189.8	214.5	188.6	185.6	93.1	61.6	98.4	114.4
Total outstanding	4,037.0	4,192.3	4,184.0	4,430.8	4,785.7	4,622.1	4,814.3	4,998.6	4,785.7	4,836.0	4,814.0	4,836.0	4,850.0
Revolving	1,065.8	1,092.0	974.6	1,032.4	1,208.6	1,098.1	1,194.6	1,161.1	1,208.6	1,220.1	1,214.0	1,220.1	1,243.8
Nonrevolving[3]	2,943.3	3,100.3	3,210.3	3,398.4	3,583.2	3,436.0	3,469.7	3,634.6	3,583.2	3,616.0	3,599.9	3,616.0	3,616.0

Terms of Credit
Not seasonally adjusted. Percent except as noted

Commercial bank interest rates[5]													
New car loans													
60-month	5.02	5.31	5.02	4.53	5.26	4.23	4.65	5.50	6.55	7.48	7.48	n.a.	n.a.
72-month	5.13	5.36	5.21	4.62	5.50	4.54	5.19	5.61	6.64	6.97	6.97	n.a.	n.a.
Credit card plans													
All accounts	14.22	15.05	14.71	14.50	16.26	14.56	15.13	16.27	19.07	20.09	20.09	n.a.	n.a.
Accounts assessed interest	16.04	16.96	16.39	16.49	17.91	16.17	16.69	18.43	20.40	20.92	20.92	n.a.	n.a.
Personal loans													
24-month	10.32	10.30	9.61	9.38	9.87	9.39	9.73	10.16	11.21	11.48	11.48	n.a.	n.a.
Finance companies (new car loans)[6]													
Interest rates	6.1	6.4	5.2	4.6	5.2	4.4	5.0	5.5	5.1	5.4	n.a.	5.4	n.a.
Maturity (months)	66	67	68	67	67	66	66	66	67	66	n.a.	66	n.a.
Amount financed (dollars)	30,173	31,311	34,449	36,307	38,600	37,991	38,044	40,156	39,407	39,048	n.a.	39,048	n.a.

The release is generally issued on the fifth business day of each month. See the Statistical Release for more information.
Footnotes appear on the second and third pages.

G.19 "Consumer Credit"

Consumer Credit Outstanding (Levels)
Not seasonally adjusted
Billions of dollars

	2018	2019	2020	2021	2022[r]	2022 Q1	2022 Q2[r]	2022 Q3	2022 Q4[r]	2023 Q1[r]	2023 Feb[r]	2023 Mar[r]	2023 Apr[p]
Total	4,007.0	4,192.2	4,184.9	4,400.8	4,785.7	4,462.6	4,583.4	4,681.1	4,785.7	4,777.4	4,769.5	4,777.4	4,802.9
Major holders													
Depository institutions	1,687.4	1,774.1	1,687.5	1,827.2	2,002.2	1,830.9	1,915.1	1,959.3	2,002.2	2,002.5	1,994.8	2,002.5	2,021.7
Finance companies	534.4	537.7	551.4	577.0	573.4	572.8	570.4	570.2	573.4	575.2	573.6	575.2	576.1
Credit unions	481.2	498.0	505.1	532.0	636.8	547.2	584.4	616.3	636.8	643.6	641.0	643.6	647.9
Federal government[7]	1,236.3	1,319.2	1,381.0	1,436.4	1,487.3	1,455.0	1,457.1	1,479.1	1,487.3	1,501.3	1,504.9	1,501.3	1,502.6
Nonprofit and educational institutions[8]	31.3	27.3	24.1	22.4	20.3	22.0	21.6	21.1	20.3	20.2	20.2	20.2	20.0
Nonfinancial business	36.5	35.8	35.8	35.8	35.8	34.7	34.9	35.0	35.8	34.7	35.0	34.7	34.6
Major types of credit, by holder													
Revolving	1,053.8	1,092.0	974.6	1,042.2	1,202.5	1,024.8	1,088.3	1,123.2	1,202.5	1,169.2	1,163.2	1,169.2	1,184.6
Depository institutions	947.2	983.6	875.3	944.2	1,095.7	928.2	989.3	1,021.9	1,095.7	1,064.8	1,058.2	1,064.8	1,079.1
Finance companies	23.7	21.9	17.1	13.4	12.2	12.9	12.2	11.9	12.2	11.4	11.5	11.4	11.4
Credit unions	62.4	66.5	62.3	64.7	74.7	64.8	67.7	70.4	74.7	74.2	74.3	74.2	75.3
Federal government[7]	—	—	—	—	—	—	—	—	—	—	—	—	—
Nonprofit and educational institutions[8]	—	—	—	—	—	—	—	—	—	—	—	—	—
Nonfinancial business	20.5	20.0	20.0	20.0	20.0	18.9	19.1	19.1	20.0	18.9	19.1	18.9	18.8
Nonrevolving	2,953.2	3,100.2	3,210.3	3,388.6	3,583.2	3,437.8	3,495.1	3,557.8	3,583.2	3,608.3	3,606.4	3,608.3	3,618.3
Depository institutions	740.2	790.5	812.2	883.0	906.5	902.7	925.7	937.5	906.5	937.7	936.6	937.7	942.6
Finance companies	510.7	515.9	534.3	563.6	561.2	559.9	558.2	558.3	561.2	563.8	562.0	563.8	564.6
Credit unions	418.8	431.5	442.9	467.4	562.1	482.4	516.8	545.9	562.1	569.4	566.7	569.4	572.6
Federal government[7]	1,236.3	1,319.2	1,381.0	1,436.4	1,487.3	1,455.0	1,457.1	1,479.1	1,487.3	1,501.3	1,504.9	1,501.3	1,502.6
Nonprofit and educational institutions[8]	31.3	27.3	24.1	22.4	20.3	22.0	21.6	21.1	20.3	20.2	20.2	20.2	20.0
Nonfinancial business	16.0	15.8	15.8	15.8	15.8	15.8	15.8	15.9	15.8	15.8	15.9	15.8	15.8
Memo													
Student Loans[9]	1,566.9	1,637.9	1,693.9	1,733.4	1,762.2	1,747.5	1,744.0	1,761.7	1,762.2	1,774.1	n.a.	1,774.1	n.a.
Motor Vehicle Loans[10]	1,139.6	1,184.1	1,224.4	1,314.2	1,417.5	1,332.2	1,366.4	1,396.9	1,417.5	1,431.4	n.a.	1,431.4	n.a.

Old Method

- Benchmark Data:
  U.S. Census Bureau's Annual Retail Trade Survey
  - sampled employer businesses classified in the retail trade sector
  - contained information about sales, inventories, purchases, expenses, margins, and accounts receivable for a sample of businesses in the retail trade sector
  - included accounts receivable data that measure amounts owed to retail stores by their customers for purchases made on credit

Old Method

- Benchmark Data:
  U.S. Census Bureau's Annual Retail Trade Survey
  - sampled employer businesses classified in the retail trade sector
  - contained information about sales, inventories, purchases, expenses, margins, and accounts receivable for a sample of businesses in the retail trade sector
  - included accounts receivable data that measure amounts owed to retail stores by their customers for purchases made on credit
- Indicator Data:
  The **seasonal factors** for the remaining G.19 sectors

Challenge

The discontinuation of the collection of accounts receivable data on the **U.S. Census Bureau's Annual Retail Trade Survey** caused the need to examine alternative sources of data for the retail credit estimates included in the G.19 under the Nonfinancial Business sector.

Solution

Federal Reserve Bank of New York's Consumer Credit Panel (CCP)/Equifax Data!

Solution

- **Federal Reserve Bank of New York Consumer Credit Panel (CCP)/Equifax:**
  - A database on consumers' credit use and payment performance drawn from Equifax credit bureau records.
  - Anonymized 5% random sample representative of the population of credit users in each quarter.
  - The database has two versions and contains individual- and tradeline/account-level information on virtually every debt owed by each consumer and the Equifax credit bureau score.
  - The tradeline-level data also includes specific indicators for each industry and account type (revolving/nonrevolving).

Solution

- **Federal Reserve Bank of New York Consumer Credit Panel (CCP)/Equifax:**
  - A database on consumers' credit use and payment performance drawn from Equifax credit bureau records.
  - Anonymized 5% random sample representative of the population of credit users in each quarter.
  - The database has two versions and contains individual- and tradeline/account-level information on virtually every debt owed by each consumer and the Equifax credit bureau score.
  - The tradeline-level data also includes specific indicators for each industry and account type (revolving/nonrevolving).
- To obtain the revised estimates for the Nonfinancial Business sector, we combine information from the tradeline versions of the **credit card** and the **consumer finance** datasets.

Solution (continued...)

- **Credit card tradeline data industry codes selected:**

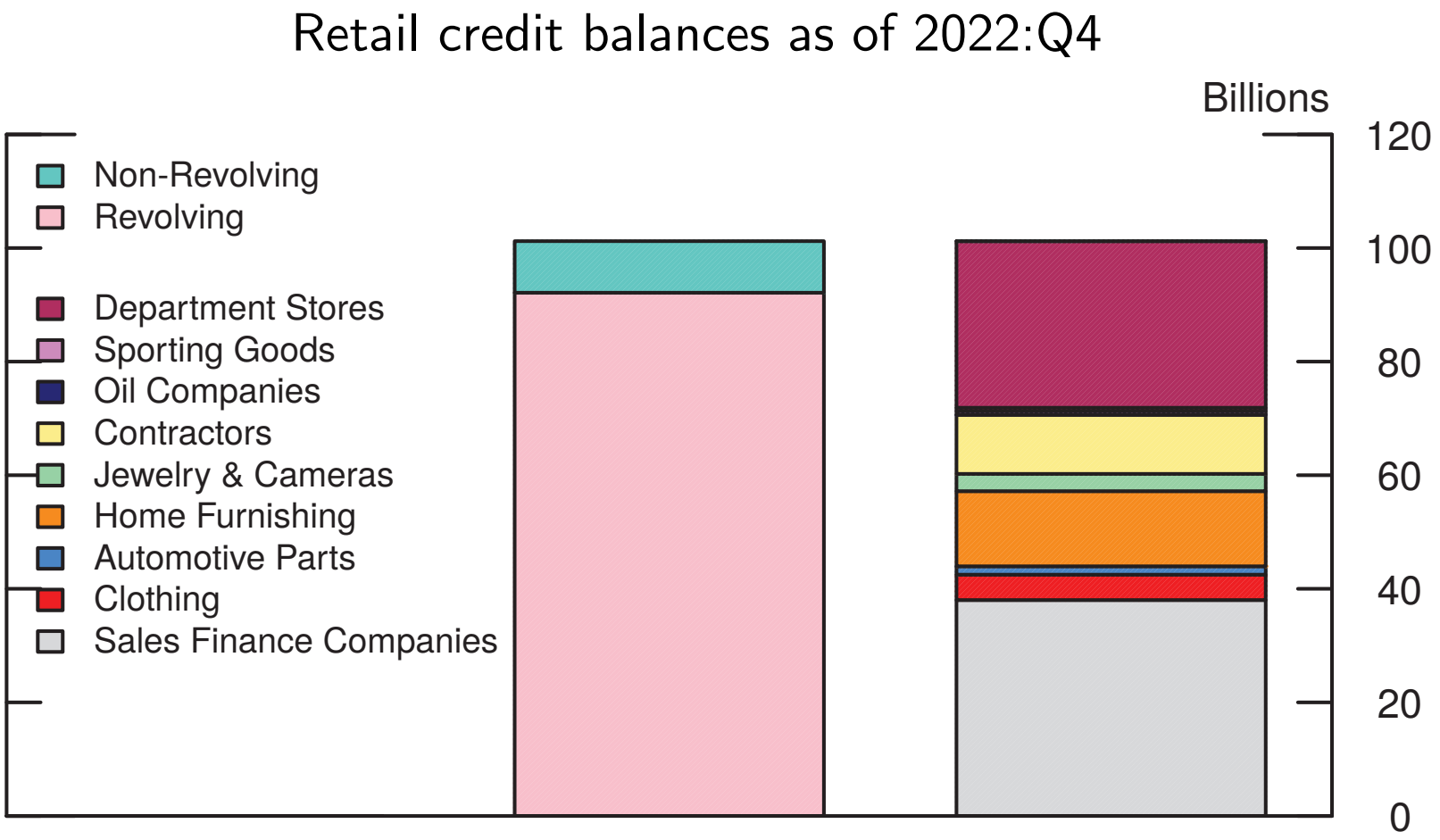
- AP - Automotive Parts
- AT - TBA Stores, Tire Dealers
- AZ - Automotive Miscellaneous
- CG - General Clothing Stores
- CS - Specialty – Shoe, Hat, Etc.
- CZ - Miscellaneous Clothing
- DC - Complete Dept. Stores
- DM - Mail Order Firms
- DV - Variety Stores
- DZ - Miscellaneous Stores
- HA - Appliances Sales & Service
- HF - Furniture, Etc.
- HM - Music & Records
- HT - Television & Radio
- HZ - Miscellaneous Home Furnishings
- JA - Jewelers
- JC - Cameras
- LA - Air Conditioning, Plumbing, Electrical Sales & Service
- LH - Hardware Stores
- LZ - Miscellaneous Hardware
- OC - Oil Companies
- SG - Sporting Goods Stores
- SZ - Miscellaneous Sporting Goods
- TN - Nursery & Landscaping Supplies & Services
- TZ - Miscellaneous Farm & Garden Supply

Solution (continued...)

- **Consumer finance tradeline data industry code selected:**
  - FF - Sales Financing

- Both sets of balances are separated by account type to obtain the revolving and nonrevolving credit shares.

Retail Credit



Source: Federal Reserve Bank of New York Consumer Credit Panel (CCP)/Equifax.

New Estimates

- For example, as of 2022:Q4:
  - The credit card tradeline data reveal over $62 billion in revolving credit balances and almost $1 billion in nonrevolving credit balances.
  - The consumer finance tradeline data show close to $30 billion in revolving credit balances and $8 billion in nonrevolving credit balances held by sales finance companies.

New Estimates

- For example, as of 2022:Q4:
  - The credit card tradeline data reveal over $62 billion in revolving credit balances and almost $1 billion in nonrevolving credit balances.
  - The consumer finance tradeline data show close to $30 billion in revolving credit balances and $8 billion in nonrevolving credit balances held by sales finance companies.
- The final estimate of $101 billion consists of two parts:
  - For revolving:
    $62 billion + $30 billion = $92 billion.
  - For nonrevolving:
    $1 billion + $8 billion = $9 billion.

New Estimates

Estimates obtained via two different methodologies

Nonfinancial Business	Old (billion $) I	New (billion $) II	Difference (billion $) III
Total	35	101	66
Revolving	19	92	73
Nonrevolving	16	9	-7

Source: G.19 Consumer Credit release and FRBNY (CCP)/Equifax .

Implementation

- One issue:
  - The estimates found in the Equifax/CCP data for retail credit are primarily financed by banks and sales finance companies.
  - Other sectors in the G.19 Consumer Credit release may contain these estimates. We want to avoid double counting.

Implementation

- One issue:
  - The estimates found in the Equifax/CCP data for retail credit are primarily financed by banks and sales finance companies.
  - Other sectors in the G.19 Consumer Credit release may contain these estimates. We want to avoid double counting.
- Moving forward:
  - We can adjust the estimates for other sectors to exclude the estimates for retail credit.
  - We can classify the retail credit estimates separately on a memo line item.

Conclusion

- Challenges of loan-level data:
  - Are we properly categorizing these loans?
  - Is there some overlap with other sectors?
  - Would we notice if a loan shifted from one sector to another?
  - Is there a way to capture smaller "mom and pop" lenders?

Conclusion

- Challenges of loan-level data:
  - Are we properly categorizing these loans?
  - Is there some overlap with other sectors?
  - Would we notice if a loan shifted from one sector to another?
  - Is there a way to capture smaller "mom and pop" lenders?
- Advantages of loan-level data:
  - Data is available more frequently and at higher levels of granularity.
  - Micro data can be useful to generate context for macro-level trends.
  - Data is available for a sector that might otherwise go uncovered.

Thank you!

Special thanks to colleagues Jessica Flagg for her work on visuals and Simona Hannon for her background and expertise.

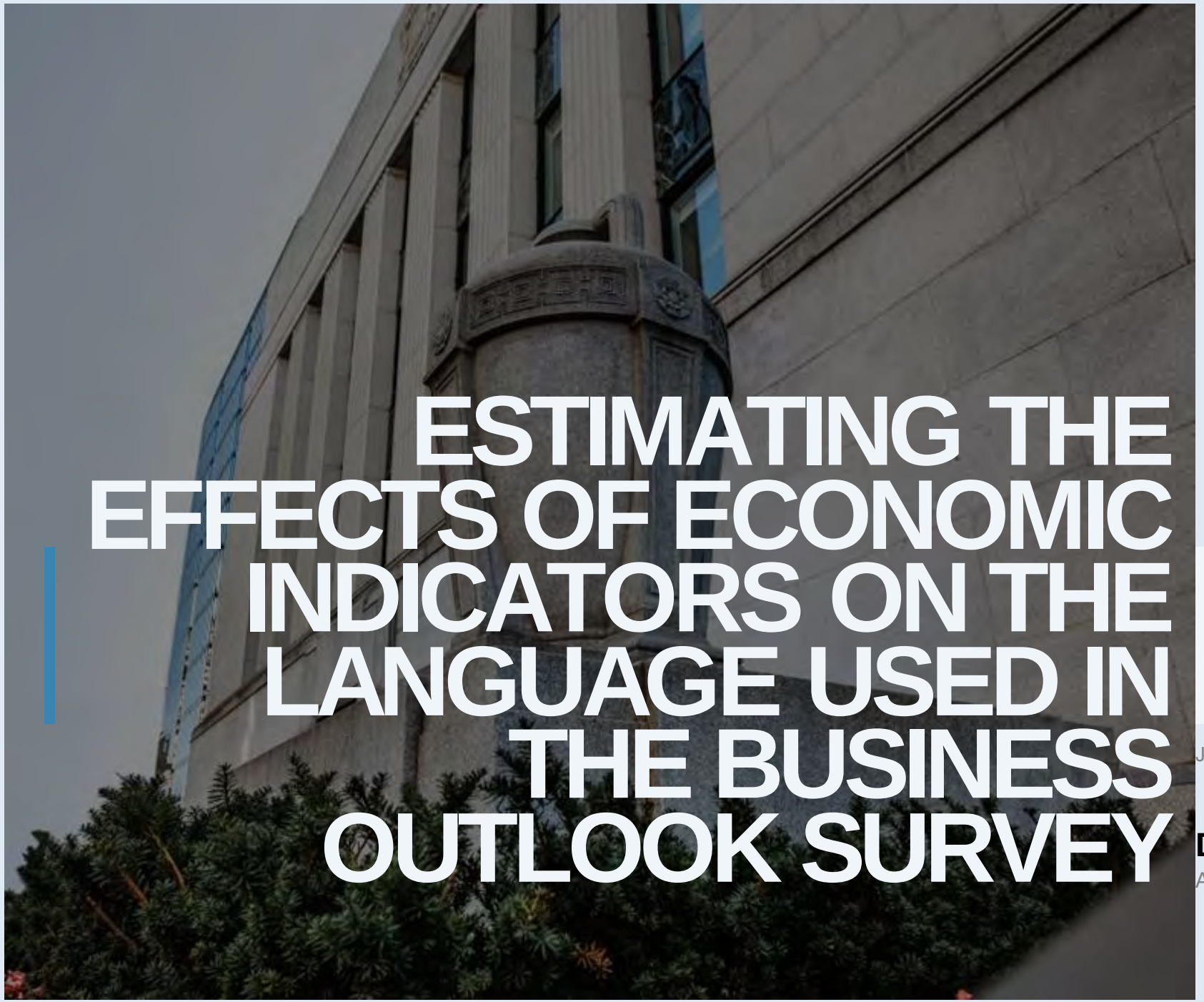
Michael M. Chernousov
Sr. Financial Analyst – Research and Statistics
Michael.M.Chernousov at frb·*gov*

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Estimating the effects of economic indicators on the language used in the business outlook survey[1]

Dave Campbell, Charita Koya and Naveen Rai,
Bank of Canada

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.



ESTIMATING THE EFFECTS OF ECONOMIC INDICATORS ON THE LANGUAGE USED IN THE BUSINESS OUTLOOK SURVEY

BANK OF CANADA
BANQUE DU CANADA

Charita Koya, Naveen Rai

July 2023

Dave Campbell
Assistant Director Data Science Applications, Bank of Canada

Business Outlook Survey

https://www.bankofcanada.ca/publications/bos/business-outlook-survey-data/

“Four times a year, our regional offices interview business leaders from about 100 firms to gather a broad range of economic perspectives …”

Stories are gathered about wages, labour, sales, investment, and more

How is the language of stories impacted by the state of the economy?

Outline

Language Processing Options

Some insights from the Business Outlook Survey

Regression on Language Models

Using the topics of discourse (X) to estimate economic variables (Y)

$$Y = f(X_{text}) + \varepsilon$$

Estimate Y and X together or condition on ‘known’ X

Algorithm results are unstable or unidentifiable to rotation

Regression of Language Models

How do Consumer Price Index (X) impact the discourse around investment in the Business Outlook Survey (Y)?

$$Y_{text} = f(X) + \varepsilon$$

Y needs to be meaningfully measured to quantify stability

Converting Text into Numbers

Language Processing = {data acquisition, pre-processing, algorithm}

Each response to a labour question from 2009 - 2022 is a document

	Docs				
Terms	15950	6515	1281	14677	11726
people	0	2	3	2	2
labour	0	0	3	1	0
staff	2	0	0	0	4
work	0	2	2	5	1
year	3	0	0	0	5
employees	2	0	0	2	0
wage	1	1	0	0	1
demand	0	1	0	1	0
workers	0	0	0	0	1
increase	2	0	0	0	1
covid	0	2	1	1	0
get	0	1	1	1	2
shortages	0	1	2	0	0
back	0	0	0	2	1
due	0	0	1	0	0
higher	2	0	0	0	0
last	1	1	0	0	5
skilled	0	1	1	0	1
market	6	0	1	0	0
months	0	2	1	1	0

- After pre-processing construct a Term x Document matrix
- #Terms = 8,677
- #Documents = 16,776
- Define the ‘word space’ of the text

Decomposing the Text Stories into Substructure

- term-document matrix N_{docs} X N_{terms}.

	Docs				
Terms	15950	6515	1281	14677	11726
people	0	2	3	2	2
labour	0	0	3	1	0
staff	2	0	0	0	4
work	0	2	2	5	1
year	3	0	0	0	5
employees	2	0	0	2	0
wage	1	1	0	0	1
demand	0	1	0	1	0
workers	0	0	0	0	1
increase	2	0	0	0	1
covid	0	2	1	1	0
get	0	1	1	1	2
shortages	0	1	2	0	0
back	0	0	0	2	1
due	0	0	1	0	0
higher	2	0	0	0	0
last	1	1	0	0	5
skilled	0	1	1	0	1
market	6	0	1	0	0
months	0	2	1	1	0

- Extract Substructures = Topics of discourse based on how words are used together

Non-Negative Matrix Factorization with Anchor Words

Algorithm basics:

Idea: Decompose the Term x Document matrix into [Term xTopic] and [Topic x Document] matrices

Model: topic weight within document = $f(X) + \varepsilon$

Inference on *f* comes from Bootstrap

Non-negative Matrix Factorization: “eigen-ish” decomposition

- term-document matrix N_{docs} X N_{terms}. Decomposed

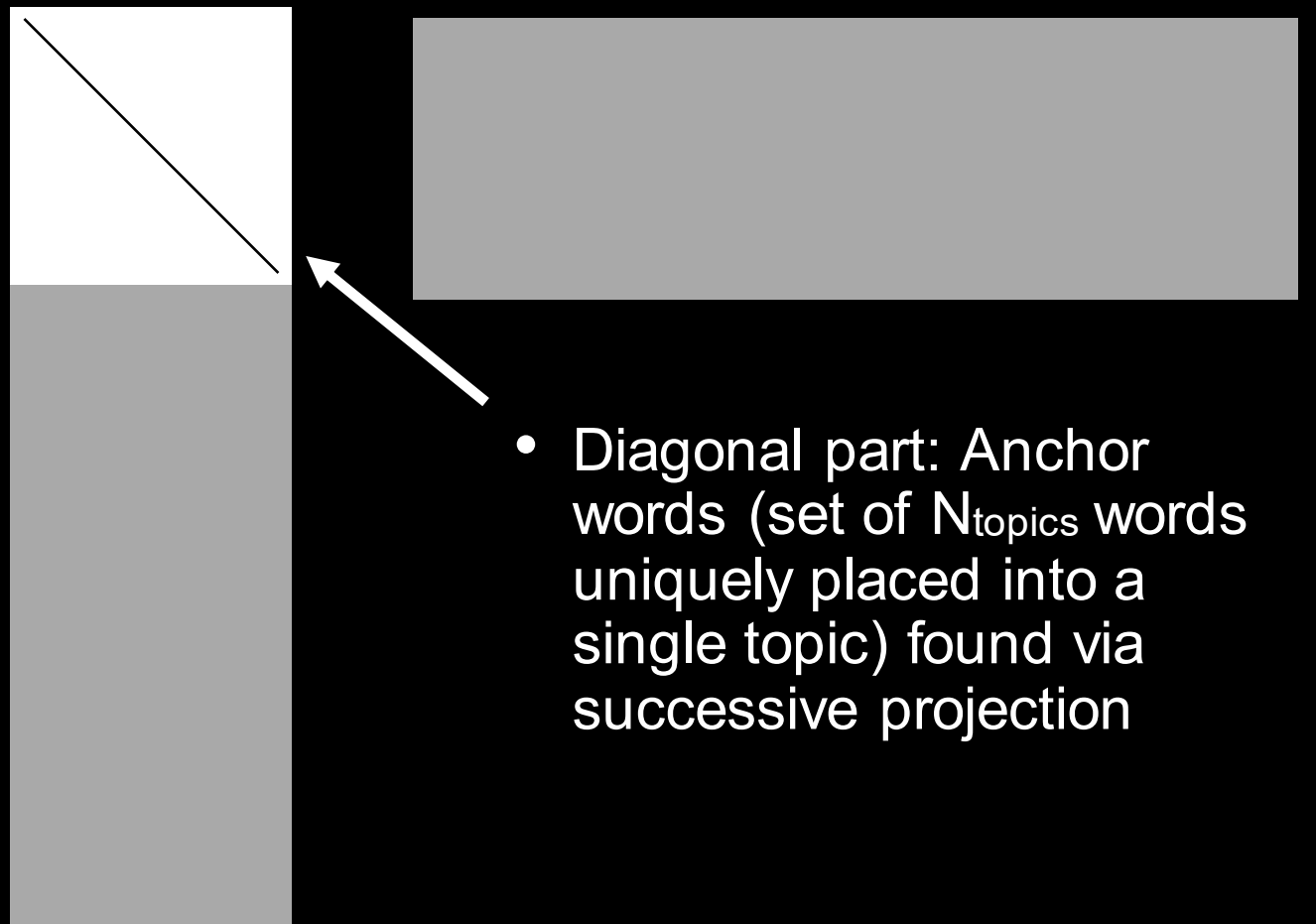
=

terms X topics (N_{terms} X N_{topics})

topics X document (N_{topics} X N_{docs})

	Docs				
Terms	15950	6515	1281	14677	11726
people	0	2	3	2	2
labour	0	0	3	1	0
staff	2	0	0	0	4
work	0	2	2	5	1
year	3	0	0	0	5
employees	2	0	0	2	0
wage	1	1	0	0	1
demand	0	1	0	1	0
workers	0	0	0	0	1
increase	2	0	0	0	1
covid	0	2	1	1	0
get	0	1	1	1	2
shortages	0	1	2	0	0
back	0	0	0	2	1
due	0	0	1	0	0
higher	2	0	0	0	0
last	1	1	0	0	5
skilled	0	1	1	0	1
market	6	0	1	0	0
months	0	2	1	1	0

=



- Diagonal part: Anchor words (set of N_{topics} words uniquely placed into a single topic) found via successive projection

Non-negative Matrix Factorization: “eigen-ish” decomposition

- term-document matrix N_{docs} X N_{terms}. Decomposed

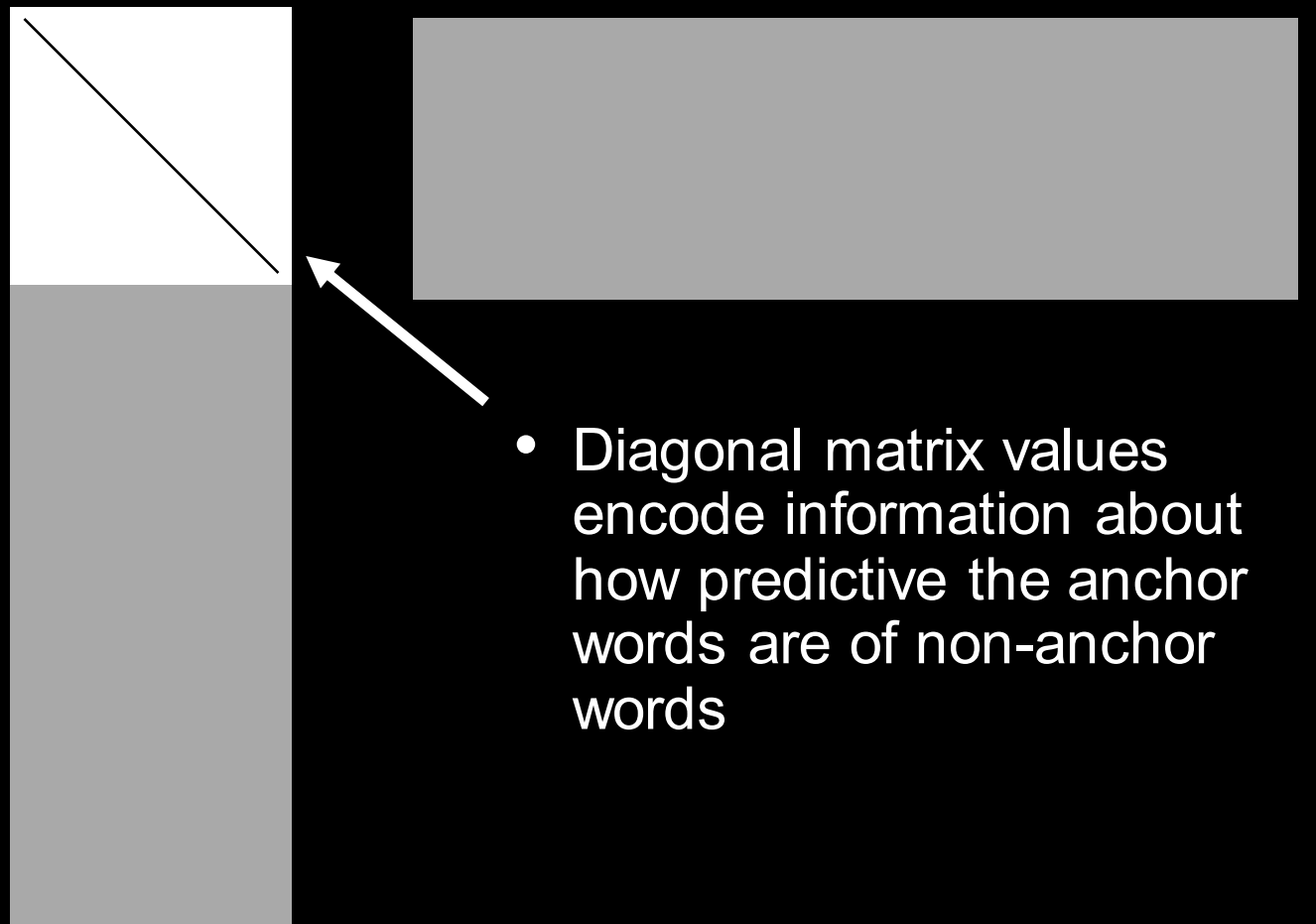
=

terms X topics (N_{terms} X N_{topics})

topics X document (N_{topics} X N_{docs})

Terms	Docs 15950	6515	1281	14677	11726
people	0	2	3	2	2
labour	0	0	3	1	0
staff	2	0	0	0	4
work	0	2	2	5	1
year	3	0	0	0	5
employees	2	0	0	2	0
wage	1	1	0	0	1
demand	0	1	0	1	0
workers	0	0	0	0	1
increase	2	0	0	0	1
covid	0	2	1	1	0
get	0	1	1	1	2
shortages	0	1	2	0	0
back	0	0	0	2	1
due	0	0	1	0	0
higher	2	0	0	0	0
last	1	1	0	0	5
skilled	0	1	1	0	1
market	6	0	1	0	0
months	0	2	1	1	0

=



- Diagonal matrix values encode information about how predictive the anchor words are of non-anchor words

Non-negative Matrix Factorization: “eigen-ish” decomposition

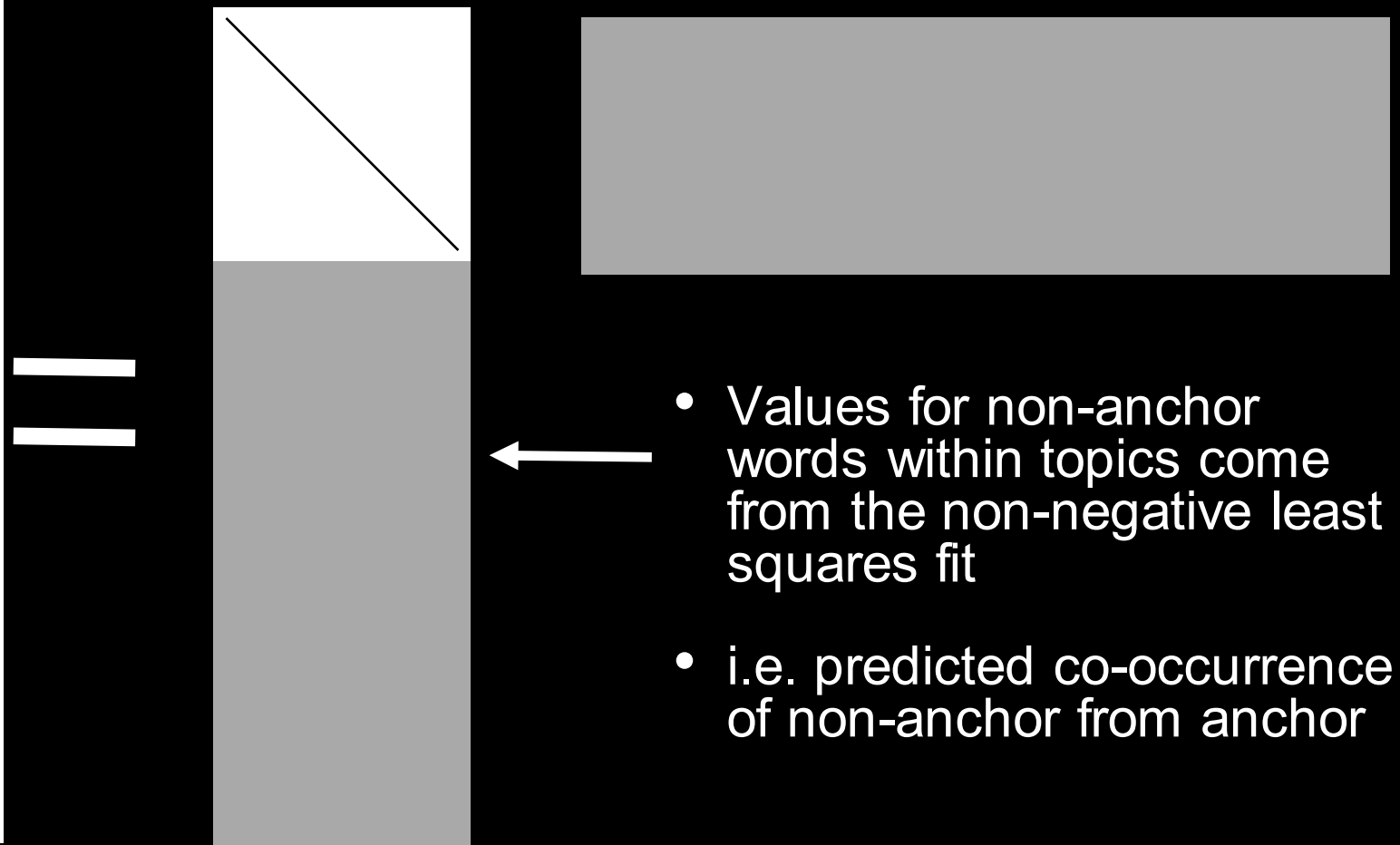
- term-document matrix N_{docs} X N_{terms}. Decomposed

=

terms X topics (N_{terms} X N_{topics})

topics X document (N_{topics} X N_{docs})

Terms	Docs 15950	6515	1281	14677	11726
people	0	2	3	2	2
labour	0	0	3	1	0
staff	2	0	0	0	4
work	0	2	2	5	1
year	3	0	0	0	5
employees	2	0	0	2	0
wage	1	1	0	0	1
demand	0	1	0	1	0
workers	0	0	0	0	1
increase	2	0	0	0	1
covid	0	2	1	1	0
get	0	1	1	1	2
shortages	0	1	2	0	0
back	0	0	0	2	1
due	0	0	1	0	0
higher	2	0	0	0	0
last	1	1	0	0	5
skilled	0	1	1	0	1
market	6	0	1	0	0
months	0	2	1	1	0



- Values for non-anchor words within topics come from the non-negative least squares fit
- i.e. predicted co-occurrence of non-anchor from anchor

Non-negative Matrix Factorization: “eigen-ish” decomposition

- term-document matrix N_{docs} X N_{terms}. Decomposed

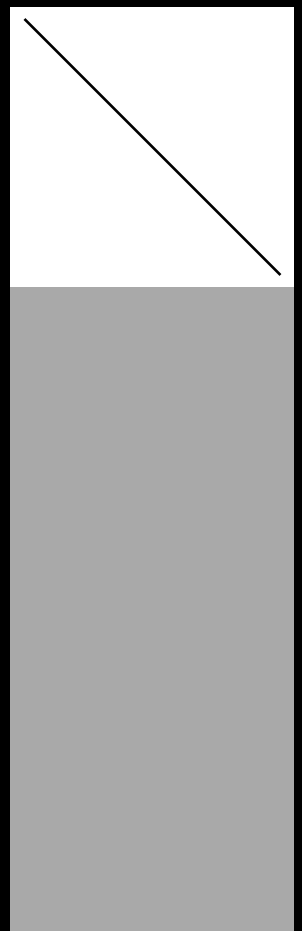
=

terms X topics (N_{terms} X N_{topics})

topics X document (N_{topics} X N_{docs})

Terms	Docs 15950	6515	1281	14677	11726
people	0	2	3	2	2
labour	0	0	3	1	0
staff	2	0	0	0	4
work	0	2	2	5	1
year	3	0	0	0	5
employees	2	0	0	2	0
wage	1	1	0	0	1
demand	0	1	0	1	0
workers	0	0	0	0	1
increase	2	0	0	0	1
covid	0	2	1	1	0
get	0	1	1	1	2
shortages	0	1	2	0	0
back	0	0	0	2	1
due	0	0	1	0	0
higher	2	0	0	0	0
last	1	1	0	0	5
skilled	0	1	1	0	1
market	6	0	1	0	0
months	0	2	1	1	0

=

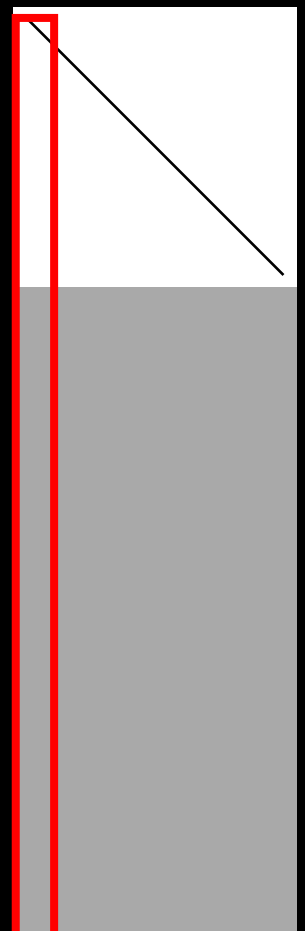


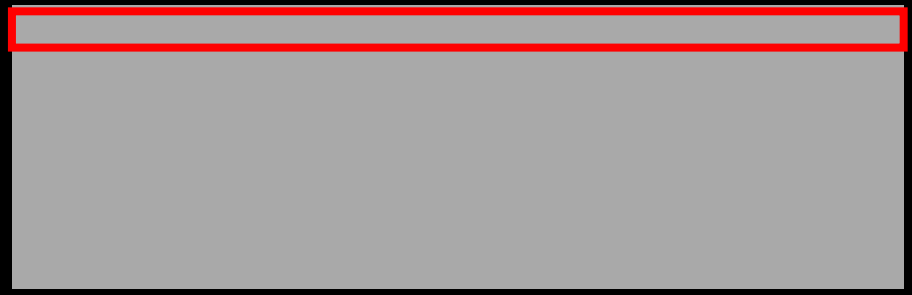


- The Topic intensity per document is found by projecting the Term-Document matrix into the anchor space

Non-negative Matrix Factorization (NMF)

- term-document matrix N_{docs} X N_{terms}. Decomposed = terms X topics (N_{terms} X N_{topics}) topics X document (N_{topics} X N_{docs})
  - Use Term x Topic to define what we are measuring



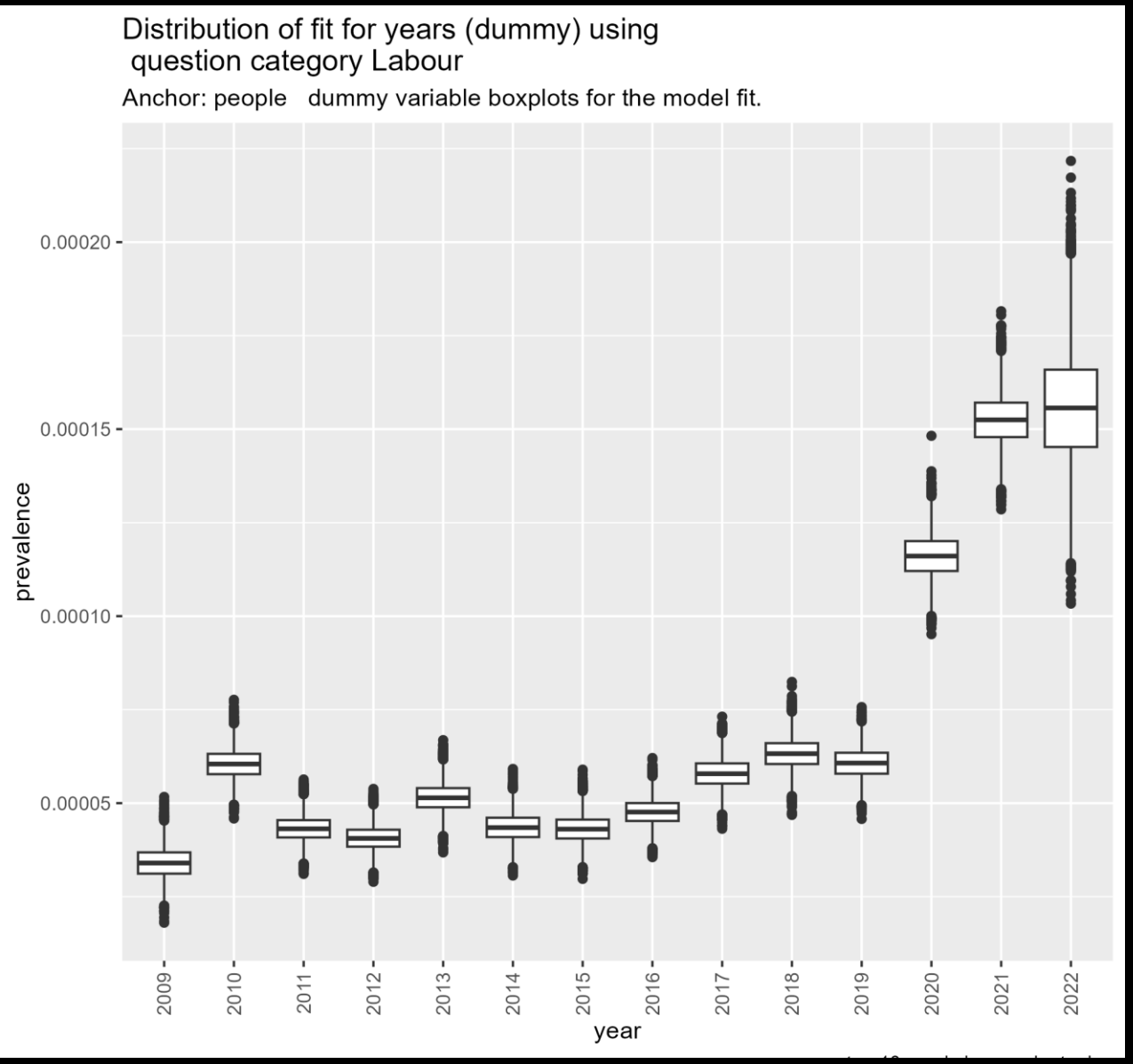


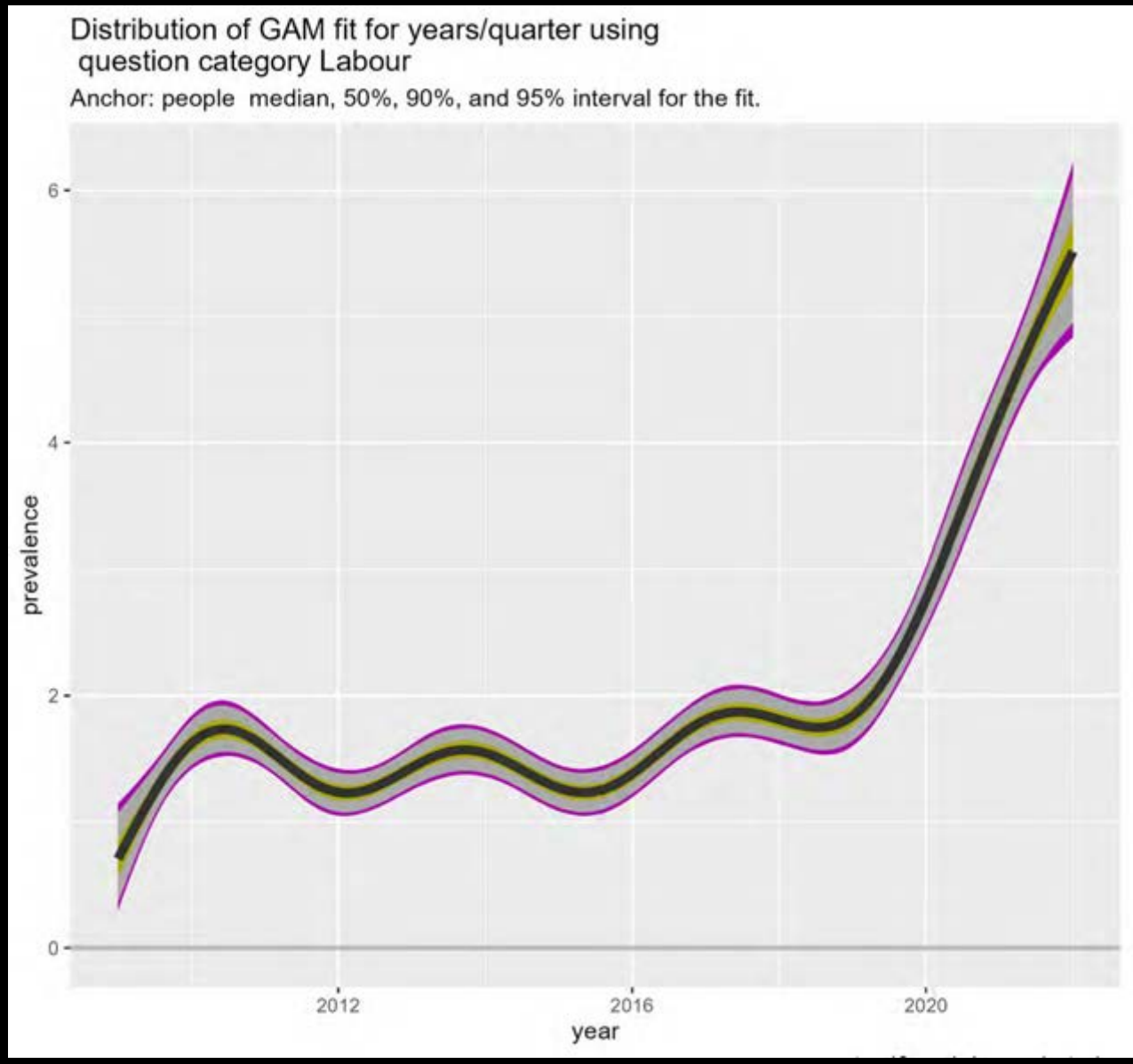
- Within the labour questions:
- Anchor words:
- "people" "staff" "year" "labour" "wage"

people topic:

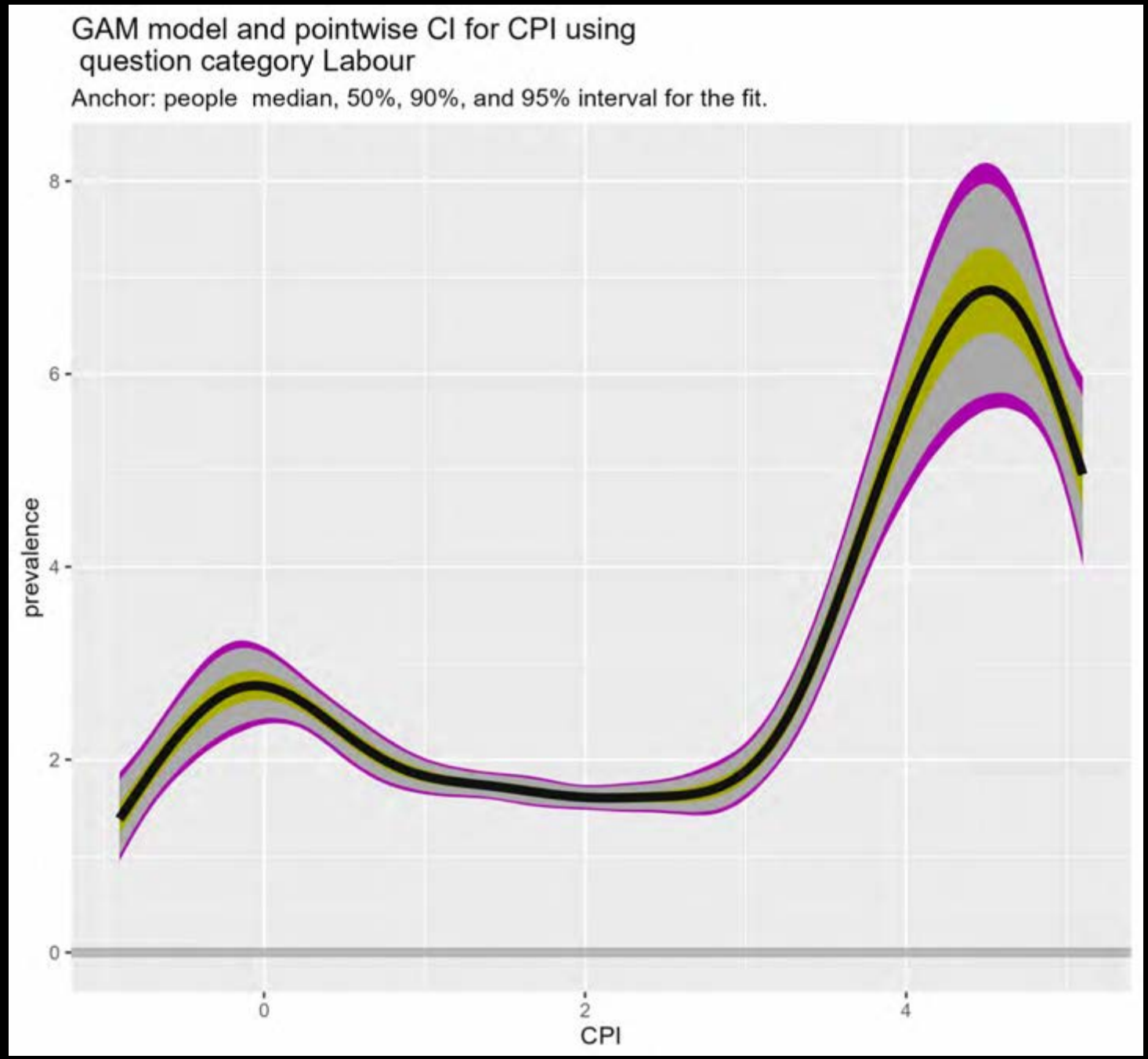
abated, workaround, hires, ghosting, newest, backgrounds, wanted

- Use corresponding row of Topic x Document as our Y
- Economic variables as our X
- Fit model $Y_{text} = f(X) + \varepsilon$
- Inference from bootstrap

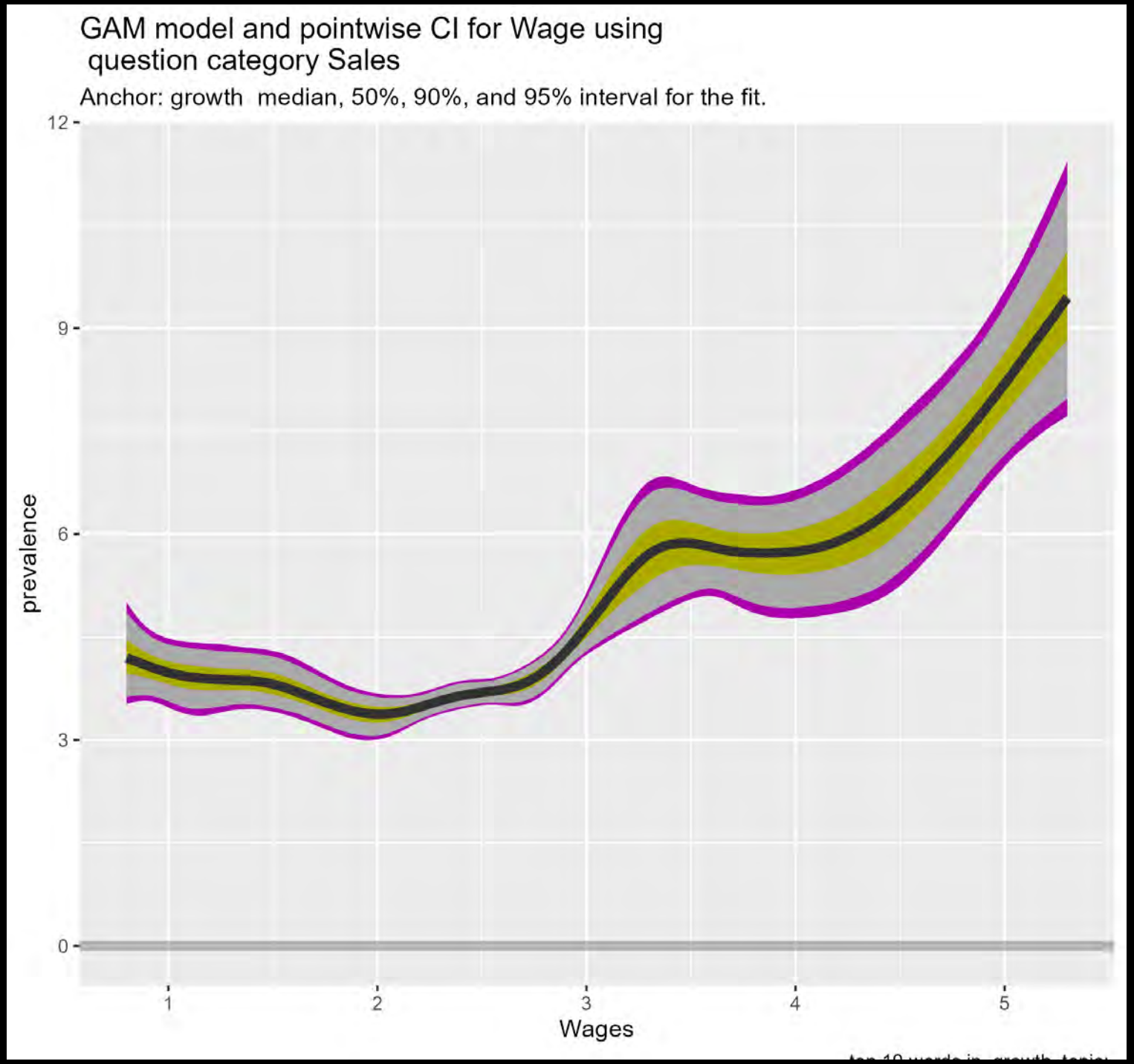

Distribution of fit for years (dummy) using
question category Labour
Anchor: people dummy variable boxplots for the model fit.
prevalence
0.00020
0.00015
0.00010
0.00005
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
year


Distribution of GAM fit for years/quarter using
question category Labour
Anchor: people median, 50%, 90%, and 95% interval for the fit.
prevalence
6
4
2
0
2012
2016
2020
year

- X = CPI


GAM model and pointwise CI for CPI using question category Labour
Anchor: people median, 50%, 90%, and 95% interval for the fit.
prevalence
8
6
4
2
0
0
2
4
CPI

- X = Wages


GAM model and pointwise CI for Wage using question category Sales
Anchor: growth median, 50%, 90%, and 95% interval for the fit.
12
9
6
3
0
prevalence
1
2
3
4
5
Wages

End Notes

Ongoing work: New model types, serial correlations in time, correlated multivariate outcomes,…

DCampbell@Bank-Banque-Canada.ca

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Unit labour cost, wages and labour productivity in Canada: A bottom-up analysis[1]

Joshua Brault, Maryam Haghighi and Masoud Nasari,
Bank of Canada,

and

Hashmat Khan,
Carleton University

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Unit Labour Cost, Wages, and Labour Productivity in Canada: A Bottom-up Analysis*

Joshua Brault† Maryam Haghighi‡
Hashmat Khan§ Masoud M Nasari¶

September 28, 2023

Abstract

We provide a granular view of Canadian real unit labour cost growth and its two components — real wages and labour productivity — from 1997 to 2019. We develop a growth accounting framework allowing us to decompose aggregates into contributions from 19 industries across time. Our approach allows us to generate insights into the behaviour of Canada's unit labour cost, wages, and labour productivity that are not possible to obtain by using aggregate data alone.

Our main findings demonstrate that there are considerable differences in the growth of real wages and labour productivity across industries over this period, but within industries real wage and labour productivity growth are strongly correlated. As a result, within industry real unit labour cost growth is small. Our framework highlights that industry effects on the aggregate depend not only on within industry changes, but also relative levels of wages/labour productivity and the input share of each industry. If an industry has a high relative level of labour productivity/wages, then shifts in input shares can impact the aggregate even in the absence of any growth within industry.

Two important themes emerge from our paper. First, analysis of macroeconomic aggregates runs into the usual challenges that such aggregates reflect complex and often non-linear interactions, making it difficult to pinpoint the exact sources of changes in aggregate variables. One way to ameliorate these challenges is to make use of more granular data, as we do. Second, a coherent economic framework is needed to reconcile and connect granular data to aggregate information, and to draw conclusions about macroeconomic aggregates from granular data. Our work proposes such a framework within the Canadian context for real unit labour cost, real wages, and labour productivity.

*The views expressed in this paper are those of the authors and do not neccessarily reflect the official views of the Bank of Canada. We thank Olivier Audet-Yang for providing excellent research assistance.

†Bank of Canada
‡Bank of Canada
§Department of Economics, Carleton University
¶Bank of Canada

1 Introduction

Unit labour cost (ULC) — a measure of the labour cost to produce a single unit of real output — is a commonly cited metric in discussions of both short- and long-run economic outcomes in Canada. In the short-run, ULC may be informative about current and future inflationary pressures, as rising labour costs can lead businesses to pass on such costs to consumers in the form of higher prices.[1] In the long-run ULC can be useful to assess Canada's competitiveness in global markets, as lower ULC makes domestic products more attractive and increase exports (Kaci and Maynard, 2002).

ULC is defined as the ratio of the wage rate and labour productivity. Increases in ULC can then occur whenever wage growth exceeds labour productivity growth. From 1997 to 2019 real ULC in Canada decreased by roughly 3%, explained by the fact that the increase in labour productivity over this period (+30%) outpaced increases in real wages (+27%). However, it is often difficult to pinpoint the exact sources of growth in these variables as they reflect the aggregation of activity and prices across a number of industries and regions. Further, analysis from a purely aggregate standpoint potentially conceals important trends occurring at a more granular level, which can be informative about the future performance of the Canadian economy.[2]

In this paper we provide an analysis of real unit labour cost and its two components, namely real wages and labour productivity, in Canada from a granular viewpoint. We develop a growth accounting framework which allows us to decompose aggregates of these variables into contributions from each industry across time. Further, our approach allows us to distinguish between changes in aggregates resulting from growth at the disaggregated level versus changes resulting from shifts in the composition of activity/prices. Our industry-level analysis focuses on 19 different industries, corresponding to the 2-digit North American Industry Classification System (NAICS) level, over the period 1997 to 2019.[3] Our approach allows us to generate insights into the behaviour of Canada's unit labour cost, wages, and labour productivity not possible using aggregate data alone, and helps us to address questions such as: What are the industries contributing to Canada's ULC, wage, and labour productivity growth? Have the primary contributors changed over time? How

[1] See Perrier and Gu (2023) for a recent example in the Canadian context. Barlevy and Hu (2023) find that ULC does not help to forecast US inflation, but that this finding is consistent with an economy where wages adjust more slowly than prices.

[2] For example, Foerster, Hornstein, Sarte and Watson (2022) show that US sectors exhibit very disparate trends in productivity over the post-war period. Some sectors exhibit positive productivity growth, while other sectors are stagnant.

[3] The NAICS classification structure can be found at https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=1181553

are the contributions distributed across industries? Is growth in these variables stagnant in some industries?

Our main findings can be summarized as follows. First, we find that across industries there are considerable differences in the growth of real wages and labour productivity over this period. Industries such as agriculture, wholesale trade, and retail trade experienced large increases in labour productivity, while others such as mining, oil, & gas, construction, and administrative & waste management services experienced low or even negative labour productivity growth over this period. In most cases real wage growth is strongly correlated with labour productivity growth at the industry-level, meaning that within industry changes in real unit labour cost are small.

Second, our decomposition highlights that within industry growth rates in isolation cannot fully determine the effect of a particular industry on aggregate outcomes. For example, the mining, oil, and gas industry averaged negative labour productivity growth and positive real wage growth over this period – implying a rise in real unit labour cost. In the aggregate however, mining, oil, and gas contributed positively to Canada's level of labour productivity and negatively to its real unit labour cost. Our framework highlights that industry effects on the aggregate depend not only on within industry changes, but also relative levels of wages/labour productivity and the input share of each industry. If an industry has a high relative level of labour productivity/wages, then shifts in input shares can impact the aggregate even in the absence of any growth within industry. In the case of the mining, oil, and gas industry, even though within industry labour productivity was falling and real unit labour cost was rising, the industry has a high relative level of labour productivity and low relative level of unit labour cost. As a result, the within industry effects are dominated by changes in the input shares, pushing aggregate labour productivity upwards and aggregate real unit labour cost downwards.

Lastly, we find that the industries responsible for short-run changes in unit labour cost, wages, and labour productivity are quite different. Year-over-year changes in real unit labour cost and labour productivity growth primarily reflect developments in goods-producing sectors, while real wages reflect developments in service-producing and non-business industries.

Two important themes emerge from our paper. First, analysis of macroeconomic aggregates runs into the usual challenges that such aggregates reflect complex and often non-linear interactions, making it difficult to pinpoint the exact sources of changes in aggregate variables. One way to ameliorate these challenges is to make use of more granular data, as we do. Second, to draw conclusions about macroeconomic aggregates from granular data

requires a coherent economic framework. In some cases, as we show, the effect of within industry changes on aggregate outcomes can be counterintuitive.

The rest of the paper is organized as follows. Section 2 derives and outlines our growth accounting framework. Section 3 discusses our data and our findings. Section 4 concludes.

2 A Growth Accounting Framework

Our growth accounting framework builds on an approach proposed by Tang and Wang (2004). However, we extend their framework in three directions. First, Tang and Wang focus on industry-level labour productivity contributions to aggregate labour productivity. Our framework accounts for contributions across both industries and geographic regions. As a result our framework can be used to provide a quantitative answer to, for example, the contribution of labour productivity growth in manufacturing in Ontario to Canada's aggregate labour productivity growth.

Second, while Tang and Wang (2004) used their decomposition to examine industrial contributions to aggregate labour productivity growth, they note that such an approach is actually valid for a variety of aggregate variables (see pg. 428). We develop a similar decomposition for aggregate real wages in Canada. To our knowledge, such a decomposition has not been used to examine contributions to aggregate real wages and real unit labour cost.[4]

Third, since aggregate real unit labour cost growth reflects the difference between aggregate real wage growth and labour productivity growth, our decompositions of wages and labour productivity can be used to decompose aggregate real unit labour cost. We begin by outlining how we decompose real wages and labour productivity, which ultimately allow us to decompose real unit labour cost.

2.1 Decomposing real wages

Our decomposition begins by noting that the aggregate real wage rate, w_t, can be defined by

$$w_t = \frac{W_t H_t}{P_t H_t}, \tag{1}$$

[4] A number of papers have studied decompositions of aggregate labour productivity in Canada. E.g., see Sharpe and Bradley (2009); Sharpe (2010); Baldwin and Willox (2016); Blit et al. (2020).

where $W_t H_t$ is total nominal compensation, H_t is total hours worked, and P_t is the aggregate price level. Equation (1) can also be used to show that aggregate nominal compensation is equal to the real wage rate multiplied by the price level and hours worked. That is, $W_t H_t = w_t P_t H_t$. Further, while (1) is expressed for the aggregate economy, this relationship also holds at disaggregated levels such that $W_{i,j,t} H_{i,j,t} = w_{i,j,t} P_{i,j,t} H_{i,j,t}$, where j refers to a geographic region and i to an industry.

Let the economy be comprised of J total geographic regions and I industries. Then (1) can be expressed as

$$w_t = \frac{\sum_J \sum_I P_{ijt} H_{i,j,t} w_{i,j,t}}{P_t H_t}, \tag{2}$$

where $P_{i,j,t}$, $H_{i,j,t}$, and $w_{i,j,t}$ are the price level, number of hours worked, and real wage rate of industry i in geographic region j at time t. Next, let $s_{i,j,t} = \frac{P_{i,j,t}}{P_t} \frac{H_{i,j,t}}{H_t}$, which represents the labour input share in industry i in region j adjusted for relative output prices. Then equation (2) can be expressed in growth rates by

$$\Delta w_t = \sum_J \sum_I \frac{w_{i,j,t-1}}{w_{t-1}} \left(s_{i,j,t} \Delta w_{i,j,t} + \Delta s_{i,j,t} \right), \tag{3}$$

where Δw_t is the growth rate of the aggregate real wage rate, $\Delta w_{i,j,t}$ is the growth rate of the real wage rate in industry i in region j, $\frac{w_{i,j,t-1}}{w_{t-1}}$ is the relative level of real wage rates at the beginning of the period, and $\Delta s_{i,j,t}$ is the absolute change in the labour input share adjusted for relative output prices ($\Delta s_{i,j,t} \equiv s_{i,j,t} - s_{i,j,t-1}$).

Equation (3) shows that the growth rate of the aggregate real wage rage can be decomposed into two distinct effects from each industry in each region. The first captures what we call a *pure growth effect*, i.e., how real wage rate growth in a specific industry in a region contributes to the aggregate. The contribution to the aggregate depends on (1) the relative real wage rate at the beginning of the period; and (2) the labour input share adjusted for relative output prices.

The second effect captures what we call a *reallocation effect*. An increase in the labour input share or an increase in relative output prices can lead to changes in the aggregate even in the absence of real wage growth. Whether this change increases or decreases the aggregate real

wage rate depends on the relative real wage rate in that industry/region.

The above decomposition has a few important properties worth mentioning. First, the decomposition is valid for chained-Fisher indices, for which most of our data is based on. Second, it is exact over long time periods, which is important considering we examine contributions over more than 20 years.

2.2 Decomposing labour productivity

The decomposition of aggregate labour productivity is similar to that of wages. To start, let the level of aggregate labour productivity be defined as the ratio of aggregate nominal value added (Q) divided by the aggregate price level (P) and the number of hours worked (H)

$$LP_t = \frac{Q_t}{P_t H_t}. \tag{4}$$

Since nominal value added is measured in common units across regions and industries, aggregate nominal value added is simply the sum of value added across regions J and industries I

$$Q_t = \sum_J \sum_I Q_{i,j,t}. \tag{5}$$

Further, the expression for labour productivity in equation (4) also holds at the disaggregated level. Then it is straightforward to express aggregate nominal value added as

$$Q_t = \sum_J \sum_I P_{i,j,t} H_{i,j,t} LP_{i,j,t}, \tag{6}$$

where $P_{i,j,t}$, $H_{i,j,t}$, and $LP_{i,j,t}$ are the price level, number of hours worked, and labour productivity in industry i in region j at time t. Substituting this expression for aggregate nominal value added into equation (4) gives

$$LP_t = \frac{\sum_J \sum_I P_{i,j,t} H_{i,j,t} LP_{i,j,t}}{P_t H_t}. \tag{7}$$

Let $s_{ijt} = \frac{P_{ijt}}{P_t} \frac{H_{ijt}}{H_t}$ again be the labour input share adjusted for relative output prices in industry i in region j at time t. The growth rate of aggregate labour productivity can then be expressed as

$$\Delta LP_t = \sum_J \sum_I \frac{LP_{i,j,t-1}}{LP_{t-1}} \left(s_{i,j,t} \Delta LP_{i,j,t} + \Delta s_{i,j,t} \right), \tag{8}$$

where ΔLP_t is the growth rate of aggregate labour productivity, $\Delta LP_{i,j,t}$ is the growth rate of labour productivity in industry i in region j, $\frac{LP_{i,j,t-1}}{LP_{t-1}}$ is the relative level of labour productivity, and $\Delta s_{i,j,t}$ is the absolute change in the labour input share adjusted for relative output prices.

Equation (8) has an identical interpretation to the decomposition for wages. There are two effects, a pure growth effect and a reallocation effect. Labour productivity growth at the industry-level influences aggregate labour productivity growth with weights given by the relative level of labour productivity and the labour input share adjusted for relative output prices. Changes in the labour input share adjusted for relative output prices lead to reallocation effects. If the reallocation is towards industries in regions with high relative levels of labour productivity, then the aggregate will increase even in the absence of labour productivity growth at the industry-level.

2.3 Decomposing unit labour cost

Unit labour cost, which represents the labour cost per unit of real output, is most commonly expressed as the ratio of total labour income and real value added. But this is equivalent to the ratio of the wage rate and labour productivity. That is

$$ULC_t = \frac{W_t H_t}{Y_t} \equiv \frac{W_t}{LP_t}, \tag{9}$$

where ULC is aggregate nominal unit labour costs, W is the nominal wage rate, H is hours worked, Y is real value added, and LP is labour productivity. Unit labour cost can be ex-

pressed in real units by using a price deflator to convert the nominal wage rate to a real wage rate. This gives

$$ulc_t = \frac{w_t}{LP_t}, \tag{10}$$

where ulc is real unit labour cost and w is the real wage rate. This relationship can be expressed approximately in growth rates by

$$\Delta ulc_t \approx \Delta w_t - \Delta LP_t, \tag{11}$$

which states that at the aggregate level the growth rate in real unit labour costs is equal to the growth rate in aggregate real wages minus the growth rate in aggregate labour productivity. Using equations (3) and (8), we can express aggregate real unit labour cost growth as contributions coming from each industry in each region. This yields

$$\Delta ulc_t \approx \sum_I \sum_J \frac{w_{i,j,t-1}}{w_{t-1}} \left(s_{i,j,t} \Delta w_{i,j,t} + \Delta s_{i,j,t}\right) - \frac{LP_{i,j,t-1}}{LP_{t-1}} \left(s_{i,j,t} \Delta LP_{i,j,t} + \Delta s_{i,j,t}\right). \tag{12}$$

Equations (3), (8), and (12) are used for the analysis in the rest of the paper.

3 Decomposing ULC, Wages, and Labour Productivity

3.1 Data

All of the data used in the following section (unless otherwise stated) was obtained from Statistics Canada Table 36-10-0480-01.[5] The table provides industry level information at a annual frequency for nominal and real value added, total hours worked, and total compensation across provinces and territories. We use a compensation-based measure of wages, defined as total compensation divided by total hours worked. Our sample spans the years 1997 to 2019 and our industry analysis is conducted for 19 industries, corresponding to 2-digit NAICS levels. Our price deflator is an implicit price index, computed as the ratio of

[5] See https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=3610048001.

nominal and real value added. Finally, while the previous section showed that our framework can decompose aggregates across industries and regions, here we focus only on industry disaggregation for brevity.

3.2 Results

Our growth accounting framework from the previous section showed that the contribution of a particular industry to Canada's aggregate real unit labour cost, real wages, and labour productivity depends on two different effects, which we refer to as *pure growth effects* and *reallocation effects*. For each industry the quantitative magnitude of each effect depends on relative levels of wages, labour productivity, and the labour input share adjusted for relative output prices at each time period.

To provide a sense of which industries are quantitatively most important for aggregates, and likely to operate through which effect, Table 1 reports averages for relative wages, labour productivity, and labour input shares adjusted for relative output prices over our sample period. Columns 4 through 6 in the table report within industry averages of real wage growth, labour productivity growth, and real unit labour cost growth.

The first two columns of Table 1 show that across Canadian industries there is a wide range of relative wages and labour productivity. Average relative real wages range from 0.52 (agriculture and accommodation) to 2.24 (mining, oil, and gas). Average relative levels of labour productivity range from 0.41 (education services) to 5.44 (mining, oil, and gas). Industries with higher relative wages also tend to have higher relative levels of labour productivity (correlation close to 0.9). All else equal, industries which have higher relative levels of wages and labour productivity will have larger contributions to aggregate wages, labour productivity, and ULC.

The third column reports the average labour input share adjusted for relative output prices. The distribution of shares is quite skewed, with nearly 50% of the total input share accounted for by the non-business sector, construction, manufacturing, and retail trade. Consequently, within industry growth in real wages and labour productivity are likely to be most apparent in macroeconomic aggregates for these industries.

While Table 1 emphasizes averages over the sample period 1997 to 2019, it is worth mentioning that some industries have exhibited quite significant shifts over this period, which may be obscured by averages. Most notable is manufacturing, which had a labour input share adjusted for relative output prices of nearly 0.18 in 1997, and declined to 0.09 by 2019. The decline in the input share adjusted for output prices reflects both a decline in relative

Table 1: Average relative wages, labour productivity, input shares adjusted for relative output prices, and within industry growth rates

	Weights			Within Industry		
	$\overline{\frac{w_i}{w}}$	$\overline{\frac{LP_i}{LP}}$	$\overline{s_i}$	$\overline{\Delta w_i}$	$\overline{\Delta LP_i}$	$\overline{\Delta ulc_i}$
Non-business sector	1.19	1.21	0.210	0.8	0.89	-0.08
Agriculture [BS11]	0.52	0.7	0.028	2.7	3.99	-0.94
Mining and oil [BS21]	2.24	5.44	0.013	2.74	-0.17	2.49
Utilities [BS22]	1.79	3.33	0.007	1.66	0.79	0.97
Construction [BS23]	1.19	0.92	0.073	-0.15	0.33	-0.47
Manufacturing [BS3A]	1.09	1.05	0.127	1.8	1.64	0.19
Wholesale trade [BS41]	1.03	0.94	0.056	2.39	2.98	-0.55
Retail trade [BS4A]	0.63	0.49	0.103	1.84	2.33	-0.46
Transportation [BS4B]	0.89	0.81	0.053	0.74	0.96	-0.18
Information [BS51]	1.25	1.6	0.019	2.46	2.1	0.38
Finance [BS5B]	1.2	1.13	0.061	2.05	2.3	-0.23
Real estate [BS53]	0.87	2.39	0.019	1.34	1.05	0.35
Professional services [BS54]	1.23	0.92	0.058	0.56	0.94	-0.36
Administrative [BS56]	0.75	0.6	0.044	0.13	0.32	-0.14
Education [BS61]	0.64	0.41	0.004	-0.33	0.67	-0.97
Health care [BS62]	0.96	0.86	0.031	-0.38	-0.29	-0.05
Arts [BS71]	0.68	0.56	0.013	0.34	-0.67	1.09
Accommodation [BS72]	0.52	0.38	0.056	0.68	0.79	-0.1
Other services [BS81]	0.57	0.42	0.036	1.1	1.43	-0.32
All industries	1	1	1	1.09	1.21	-0.11

Notes: Bars indicate sample averages over the period 1997 to 2019. It is not a requirement that the labour input share adjusted for relative output prices sum to one. Unit labour cost averages may differ slightly from average real wage growth minus average labour productivity growth due to rounding.

prices and hours worked shares. Additionally, almost all of this decline occurred from 2000 to 2010, and shares have remained relative constant since.

Columns 4 through 6 in Table 1 report within industry average growth rates of real wages, labour productivity, and unit labour cost. Once again there is very disparate trends in growth across industries. For most industries labour productivity has increased over time, with agriculture exhibiting the largest productivity gains, averaging nearly 4% annually. However, there remain a few industries (mining, oil, & gas, construction, health care services, and arts) with very low or even negative labour productivity growth over this period. There is again a strong correlation (0.7) between average labour productivity growth and real wage growth over this period. As a result, most industries do not exhibit a large change in real unit labour cost from 1997 to 2019. A notable exception is mining, oil, & gas, where real wages averaged nearly 3% growth annually, while labour productivity growth has been negative. Consequently there has been a significant rise in real unit labour cost in the industry.

Table 2: Contributions to aggregate wages, labour productivity and unit labour cost growth by industry

	Real Wages			Labour Productivity			Real Unit Labour Cost		
	Pure	Realloc.	Total	Pure	Realloc.	Total	Pure	Realloc.	Total
Non-business sector	4.25	2.88	7.13	4.8	2.9	7.7	-0.55	-0.02	-0.56
Agriculture [BS11]	0.51	-1.09	-0.58	1.56	-1.66	-0.09	-1.05	0.57	-0.48
Mining and oil [BS21]	-1.18	2.61	1.43	-3.85	7.23	3.38	2.67	-4.62	-1.95
Utilities [BS22]	0.29	-0.08	0.21	0.14	-0.26	-0.12	0.16	0.18	0.34
Construction [BS23]	-0.49	4.46	3.96	0.2	3.48	3.68	-0.69	0.98	0.28
Manufacturing [BS3A]	5.73	-9.19	-3.46	5.3	-8.92	-3.62	0.43	-0.27	0.16
Wholesale trade [BS41]	2.9	-2.21	0.68	3.25	-1.98	1.27	-0.36	-0.24	-0.59
Retail trade [BS4A]	2.54	-1.09	1.45	2.44	-0.85	1.59	0.1	-0.24	-0.14
Transportation [BS4B]	0.67	0.58	1.24	0.74	0.6	1.34	-0.07	-0.02	-0.09
Information [BS51]	1.22	-0.31	0.91	1.4	-0.46	0.94	-0.18	0.15	-0.03
Finance [BS5B]	3.12	-1.36	1.77	3.26	-1.25	2.01	-0.14	-0.11	-0.24
Real estate [BS53]	0.41	0.16	0.57	0.96	0.3	1.26	-0.55	-0.14	-0.69
Professional services [BS54]	0.82	3.17	3.98	1.05	2.36	3.41	-0.23	0.8	0.57
Administrative [BS56]	0.04	1.61	1.65	0.16	1.27	1.43	-0.11	0.34	0.22
Education [BS61]	-0.03	0.19	0.15	0.02	0.12	0.14	-0.05	0.07	0.01
Health care [BS62]	-0.25	1.41	1.16	-0.15	1.24	1.09	-0.1	0.17	0.07
Arts [BS71]	0.06	0.24	0.3	-0.13	0.22	0.09	0.2	0.02	0.21
Accommodation [BS72]	0.41	0.52	0.93	0.34	0.38	0.71	0.07	0.14	0.21
Other services [BS81]	0.44	0.07	0.51	0.43	0.05	0.48	0.01	0.02	0.03
All industries	21.47	2.55	24.02	21.92	4.77	26.69	-0.45	-2.22	-2.67

Notes: Contributions for each industry are computed as the sum of contributions from 1998 to 2019. Columns labelled pure refer to pure growth effects and columns labelled realloc. refer to reallocation effects. The sum of pure effects and reallocation effects equals total effects. Any minor discrepancies are due to rounding.

Based on our growth accounting framework, Table 2 reports the long-run contributions of each industry to aggregate wages, labour productivity, and unit labour cost from 1998 to 2019. The values in the table are computed by summing contributions to aggregates across the sample period. Additionally, the table distinguishes between contributions resulting from within industry growth and contributions from changes in the labour input share adjusted for relative output prices (i.e., reallocation effects).

Given the large fraction of input shares adjusted for relative output prices accounted for by the non-business sector, construction, and manufacturing, it is perhaps unsurprising to see that these industries are quantitatively important for understanding aggregate growth over this period. Through both pure growth effects and reallocation effects, the non-business sector contributed to around 7% of aggregate growth in real wages and labour productivity, while construction contributed to around 4%. However, the reasons for large positive contributions to the aggregate are quite different. The non-business sector contributions result largely from within industry increases in wages and productivity, while construction

contributions are almost entirely from reallocation effects.

In contrast to the non-business sector and construction, manufacturing contributed negatively to both aggregate real wages and labour productivity. This negative contribution to the aggregate occurred in spite of within industry growth in real wages and labour productivity, as reallocation effects dominated pure growth effects.

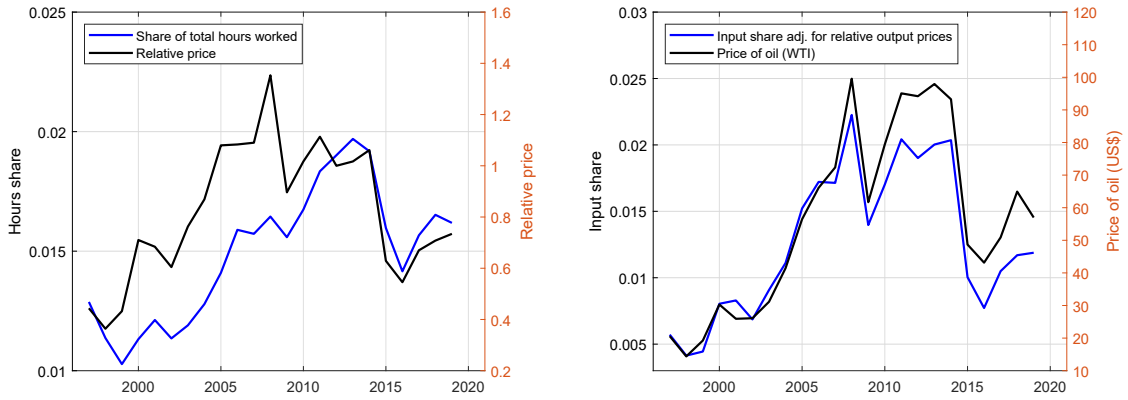
A more surprising contribution (based on input shares adjusted for output prices) to the aggregate comes from mining, oil, & gas. This industry contributed over 3% to aggregate labour productivity and roughly 1.5% to aggregate wages, even though within industry labour productivity growth was on average negative over this period. The explanation for this result resides in the fact that the mining, oil, & gas industry has high relative real wages and levels of labour productivity. As a result, reallocation effects from this industry can be particularly strong. Further, our framework suggests that these reallocation effects reflect movements in both relative prices and shares of hours worked, which coincide with booms in the price of oil.

To illustrate this phenomena, Figure 1 plots the share of total hours worked and relative price of the mining, oil, and gas industry in the left panel, and the labour input share adjusted for relative output prices with the West Texas Intermediate price of oil in the right panel. With rising oil prices beginning in the early 2000s, both the share of hours worked and relative prices begin to rise, with peaks around 2008 and 2015. Together these yield an input share adjusted for relative output prices which is also rising throughout this period. This rise, coupled with the high relative levels of wages and labour productivity, produces a strong reallocation effect, outweighing the negative within industry labour productivity growth in terms of contributions to aggregate labour productivity (see also Sharpe and Bradley (2009)).

For real unit labour cost, industry contributions from pure growth effects and reallocation effects are mostly small in magnitude. This is consistent with the previous facts which documented that industries exhibited strong correlations in both relative levels of real wages and labour productivity, and within industry average wage growth and labour productivity growth.

One exception is the mining, oil, and gas industry, which contributed roughly -2% to aggregate real unit labour cost over this period. Even though real unit labour cost in the mining, oil, and gas industry grew on average by 2.5% per year, illustrated by the pure growth effect contributing 2.67% to aggregate ULC, in total the industry contributed to lower aggregate real unit labour cost due to strong reallocation effects. In part because of the

Figure 1: Mining, oil, and gas: Share of hours worked, relative prices, input shares, and the price of oil



Notes: The share of hours worked corresponds to $\frac{H_{i,t}}{H_t}$, the relative price corresponds to $\frac{P_{i,t}}{P_t}$, and the input share adjusted for relative output prices corresponds to $\frac{H_{i,t}}{H_t}\frac{P_{i,t}}{P_t}$, where i is the mining, oil, and gas industry. The West Texas Intermediate price of oil is the monthly spot price converted to annual by averaging. The data was obtained from the Federal Reserve Bank of St. Louis database, and the series code is WTISPLC.

rise in the input share adjusted for output prices documented above, but also because the level of real unit labour cost in the industry is low. This is easiest to see by comparing the average levels of relative real wages and labour productivity reported in Table 1. Average relative real wages in the industry are high (2.24), however the average relative level of labour productivity is significantly higher (5.44), implying that the level of real unit labour cost in the industry are quite low. Together these produce a strong negative reallocation effect on aggregate real unit labour cost.

Finally, while our previous discussion has centered on long-run industry contributions over the period 1997 to 2019, our framework can also be used to understand short-run changes in aggregate wage, labour productivity, and unit labour cost growth. To facilitate this part of our analysis, we collapse our industry-level data into more coarse definitions, corresponding to goods producing sectors, service producing sectors, and the non-business sector. Industries included in goods production are: (1) Agriculture; (2) Mining, Oil, and Gas; (3) Utilities; (4) Construction; and (5) Manufacturing. Industries in service production are the remaining industries excluding the non-business sector.[6]

Figure 2 depicts the contribution of goods producing industries, service producing industries, and the non-business sector to aggregate real unit labour cost growth, real wage growth, and labour productivity growth from 1998 to 2019. The contribution for each sec-

[6]These definitions follow those used by Statistics Canada in Table 36-10-0480-01.

tor is the total contribution, that is, they combine both pure growth effects and reallocation effects. The length of the bar pertaining to a colour indicates that sectors magnitude of contribution (e.g., a bar mostly in blue indicates that goods producing sectors contributed to the majority of change in that period). Bars extending below 0 on the y-axis indicate a negative contribution. The sum of the bar equals the aggregate variable growth, which is indicated by the solid black line in the figures.

The top panel in Figure 2 shows that the large majority of year-over-year changes in aggregate real unit labour cost growth comes from goods producing industries. Notably in 2009 and 2015 there is large upward contributions to aggregate ULC. An inspection of contributions by goods producing sectors shows that almost all of this is mining, oil, and gas, and 2009 and 2015 are years where oil prices fell dramatically.

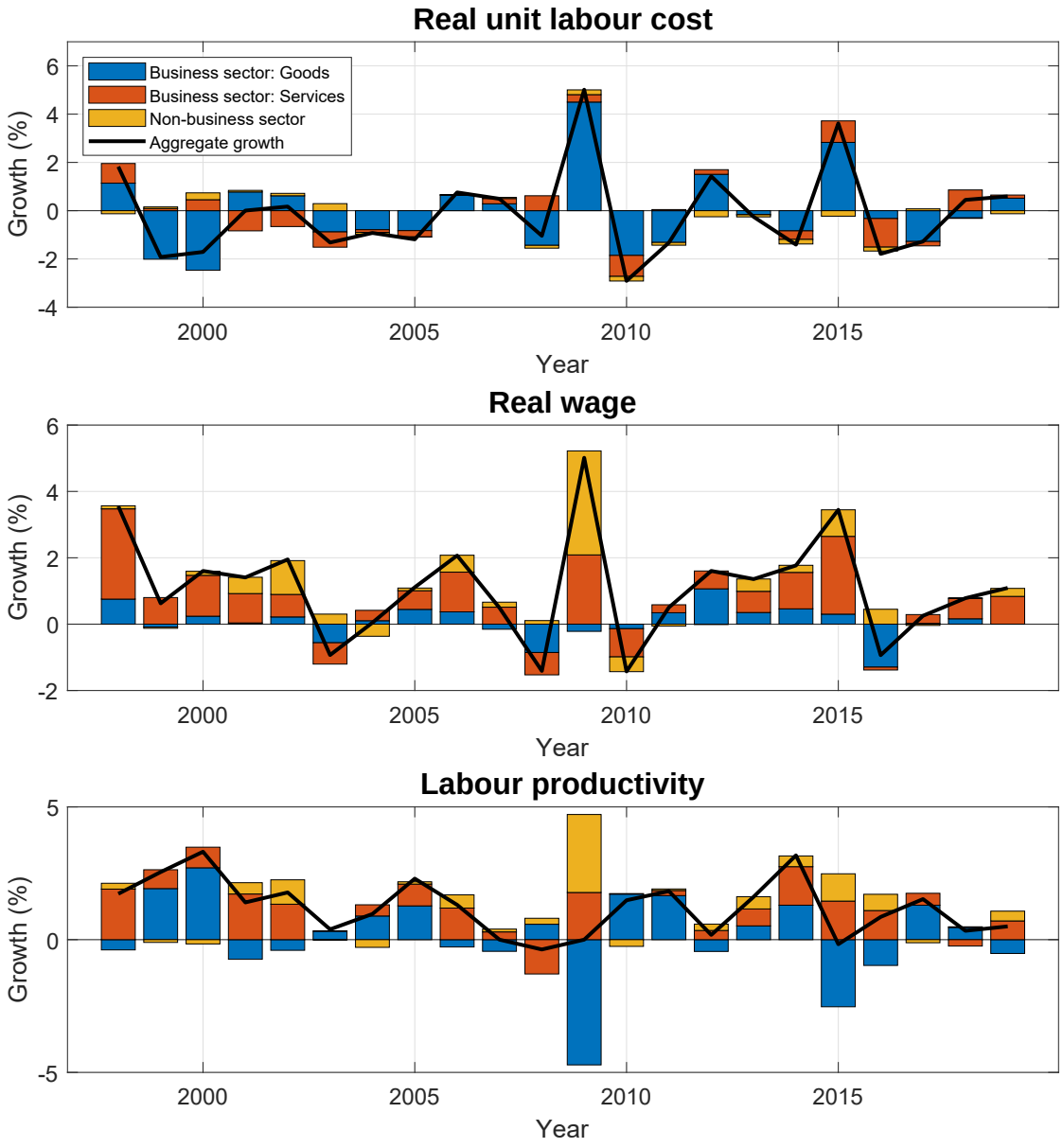
The middle panel in Figure 2 shows that, in contrast to ULC, much of aggregate real wage growth changes comes from the service producing sector and the non-business sector. There is a particularly large contribution from services and the non-business sector in 2009, where the aggregate real wage grew at roughly 5%. Part of this growth can be explained by a fall in the aggregate price level, which declined by 2.35%, and raised real wages. However, aggregate nominal compensation per hour also grew by about 2.5%, increasing from 28.86 per hour in 2008 to 29.59 in 2009.

Finally, turning to contributions to aggregate labour productivity we find that goods producing sectors, service producing sectors, and the non-business sector all contribute to changes in aggregate growth at points throughout the sample period. Goods producing sectors pushed up aggregate labour productivity growth in 1999 and 2000, and this is mostly driven by gains in manufacturing.[7] Notably, in 2009 and 2015 we again see the role played by oil prices as large negative contributions appear for goods producing sectors. However this negative contribution is offset by positive contributions from services and the non-business sector, netting out to yield roughly zero aggregate labour productivity growth.

The plots in Figure 2 can be informative about underlying trends and the future behaviour of macroeconomic aggregates. Using our decomposition it is possible to further examine short-run contributions emanating from pure growth effects and reallocation effects (not shown here). For example, the large increase in real unit labour cost in 2009 and 2015 reflects primarily reallocation from changes in the price of oil. From a policymakers perspective, such an observed change in real unit labour cost from reallocation may call for

[7] Baldwin and Willox (2016) find evidence consistent with this, noting that manufacturing, along with finance, insurance, & real estate and mining, oil, & gas can account for much of the labour productivity growth slowdown in Canada in the 2000s.

Figure 2: Contributions to aggregate unit labour cost, real wages, and labour productivity over time



Notes: Contributions begin in 1998 because our framework uses growth rates and one observation is lost at the beginning of the sample.

different policy prescriptions than if the observed change in real unit labour cost resulted from primarily within industry changes. Our framework offers the possibility to consider such scenarios.

4 Conclusion

In this paper we have decomposed Canadian aggregate real unit labour cost and its two components, real wages and labour productivity, into contributions from 19 industries over the period 1997 to 2019. Importantly, our method disentangles contributions to the aggregate arising from within industry growth and reallocation. We find that goods producing industries play an important role in real unit labour cost short- and long-run growth.

Our paper provides an example of utilizing granular data to understand the behaviour of macroeconomic aggregates. This approach can help policymakers to pinpoint the sources of changes in aggregates of unit labour cost, wages, and labour productivity, while simultaneously observing underlying trends at a disaggregated level.

References

Baldwin, J. R. and Willox, M.: 2016, The Industry Origins of Canada's Weaker Labour Productivity Performance and the Role of Structural Adjustment in the Post-2000 Period, *International Productivity Monitor* **31**, 19–36.

Barlevy, G. and Hu, L.: 2023, Unit labor costs and inflation in the non-housing service sector, *Chicago Fed Letter* **No 477**.

Blit, J., Skuterud, M. and Veall, M. R.: 2020, The Pandemic and Short-Run Changes in Output, Hours Worked and Labour Productivity: Canadian Evidence by Industry, *International Productivity Monitor* **39**, 16–32.

Foerster, A. T., Hornstein, A., Sarte, P.-D. G. and Watson, M. W.: 2022, Aggregate implications of changing sectoral trends, *Journal of Political Economy* **130**(12), 3286–3333.

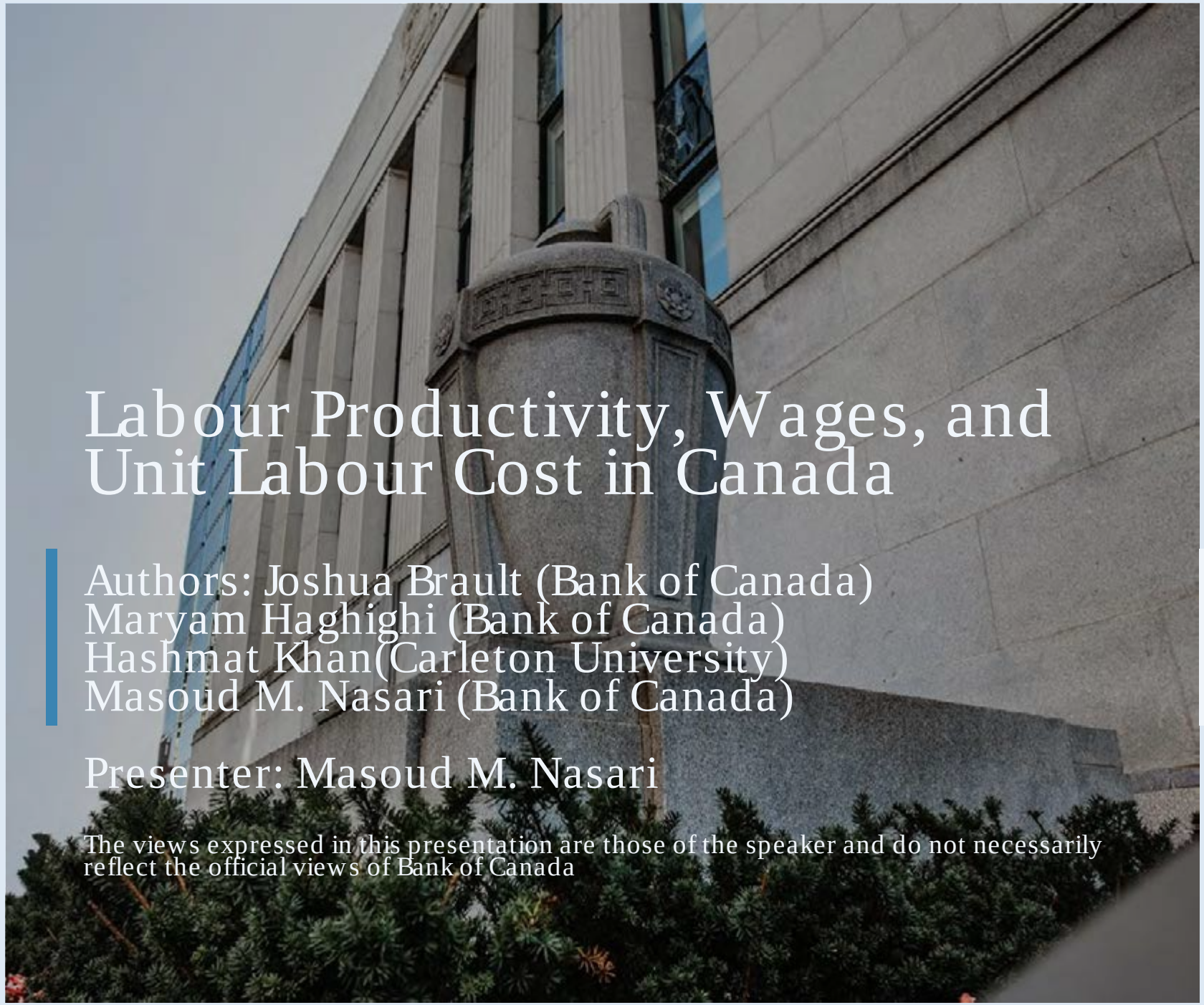
Kaci, M. and Maynard, J.-P.: 2002, Unit labour costs and the competitiveness of canadian businesses, *Productivity Growth in Canada Annual Catelogue 15-204 159-183*.

Perrier, P. and Gu, L.: 2023, Wage Growth—Canadian Inflation's Last Leg?, *Scotiabank Insights and Views 1-5*.

Sharpe, A.: 2010, Can Sectoral Reallocations of Labour Explain Canada's Absymal Productivity Performance?, *International Productivity Monitor* **19**, 40–49.

Sharpe, A. and Bradley, C.: 2009, A Detailed Analysis of the Productivity Performance of Oil and Gas Extraction in Canada, *CSLS Research Reports 2009-08*, Centre for the Study of Living Standards.

Tang, J. and Wang, W.: 2004, Sources of aggregate labour productivity growth in Canada and the United States, *Canadian Journal of Economics* **37**(2), 421–444.



Labour Productivity, Wages, and Unit Labour Cost in Canada

Authors: Joshua Brault (Bank of Canada)
Maryam Haghighi (Bank of Canada)
Hashmat Khan(Carleton University)
Masoud M. Nasari (Bank of Canada)

Presenter: Masoud M. Nasari

The views expressed in this presentation are those of the speaker and do not necessarily reflect the official views of Bank of Canada

BANK OF CANADA
BANQUE DU CANADA

July 15, 2023

IDAS, ITS

Background

- Aggregate labour productivity developments are an important consideration for central banks and policy makers
- In relation to wage developments, they can provide useful information on the inflation outlook
- For example, in April 2023, the Governor of Bank of Canada Tiff Macklem said ,“Wage growth has been running at four to five per cent and unless there’s a surprising acceleration in productivity, that’s not consistent with two per cent inflation.”

Motivations for creating a platform for labour productivity data

- Motivated by the importance of timely and dynamic labour productivity data for researchers, academics, executives and curious minds to conduct analysis and draw educated conclusions, we created the Canadian Labour Productivity Platform (CLaPP)
- CLaPP is a publicly available platform that hosts and visualizes aggregate and disaggregate labour productivity data
- Aggregate data are also available from Canadian Statistical Agencies and some international institutions such as OECD
- What distinguishes CLaPP is that it provides aggregate and disaggregate data alongside with interactive visualizations at sectoral and regional levels on the same platform

CLaPP

Canadian Labour Productivity Platform

CLaPP

- is a joint project with Carleton University
- has its dedicated domain, server and website hosted by Carleton University
  - www.carleton.ca/clapp
- uses the latest publicly available data from Statistics Canada
- provides easy to understand and interactive story telling visualizations along sectoral and regional dimensions
- has the flexibility and capacity of expanding to include more economic measures and to host deep diving results, insights, research and analysis

carleton.ca/clapp/

Carleton.ca | About | Admissions | Academics | Research | Campus | Future Students | Current Students | Faculty/Staff | Alumni

Canadian Labour Productivity Platform

Canadian Labour Productivity Platform (CLaPP)

LABOUR PRODUCTIVITY TRENDS

CURRENT SNAPSHOT

RESEARCH AND ANALYSIS

HOW TO CITE CLAPP DATA

CONTACT US

DEPARTMENT OF ECONOMICS

Labour productivity measures the efficiency of workers in producing goods and services. It is an important indicator of economic performance for a variety of reasons. Labour productivity is closely linked to businesses' costs of production (unit labour cost), international competitiveness, the output gap, and inflationary pressures in the economy. Over longer time horizons labour productivity is also tied to wages and the average standard of living.

About the Project

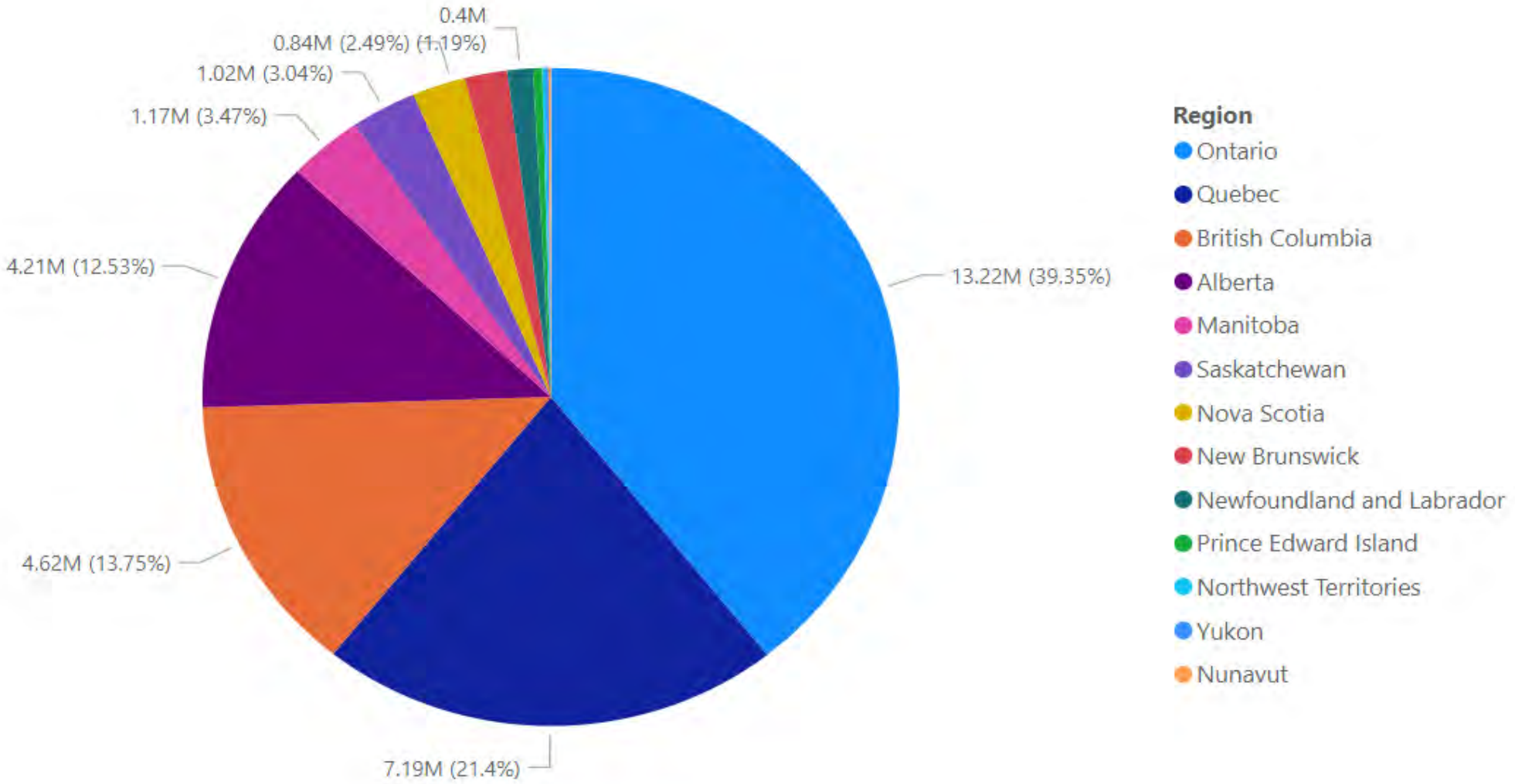
CLaPP is a Carleton University project led by the Department of Economics with collaborative support from the Bank of Canada. The platform provides data on Canadian labour productivity which is updated monthly. A distinguishing feature of CLaPP is that it provides both aggregate and disaggregated (sectoral/regional) labour productivity data alongside with rich visualizations.

Current Snapshot - Canadian L

Microsoft Power BI

app.powerbi.com/view?r=eyJrIjoiYTU2ZTBkZjgtZDZmZS00YjliLWIwMzktN2RjMTY3YjEzM2FkIiwidCI6IjZhZDkxODk1LWRlMDYtNDg1ZS1iYzUxLWZjZTEyNmNjODUzMCIsImMiOjN9&pageNa

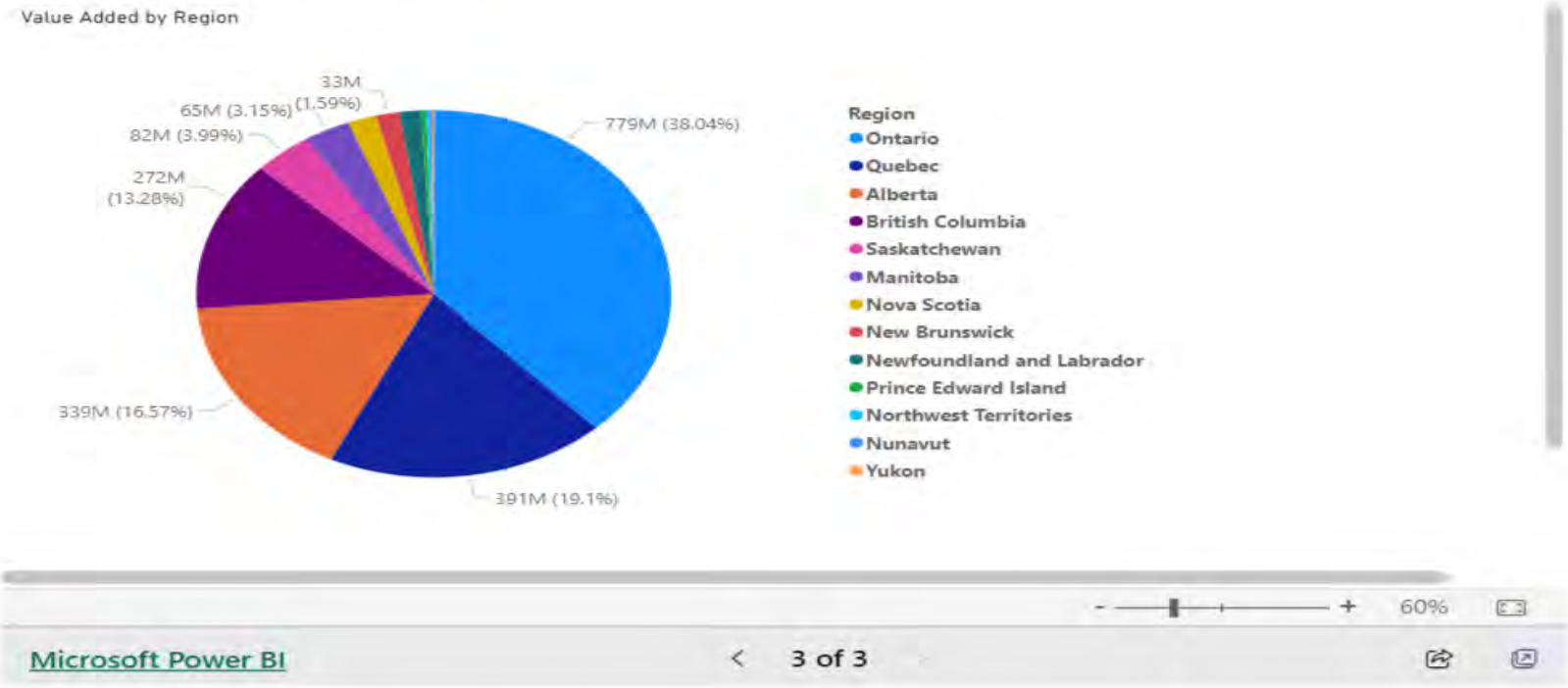
Hours by Region



TRENDS
CURRENT SNAPSHOT
RESEARCH AND ANALYSIS
HOW TO CITE CLAPP DATA
CONTACT US
DEPARTMENT OF ECONOMICS

The four most populated provinces, Alberta, British Columbia, Ontario, and Quebec, represent a substantial share of Canada's hours worked, roughly equal to 86%. The data is obtained from Statistics Canada Table 36-10-0480-01 (https://doi.org/10.25318/3610048001-eng) and the industry disaggregation is at level 5.

Value Added by Region



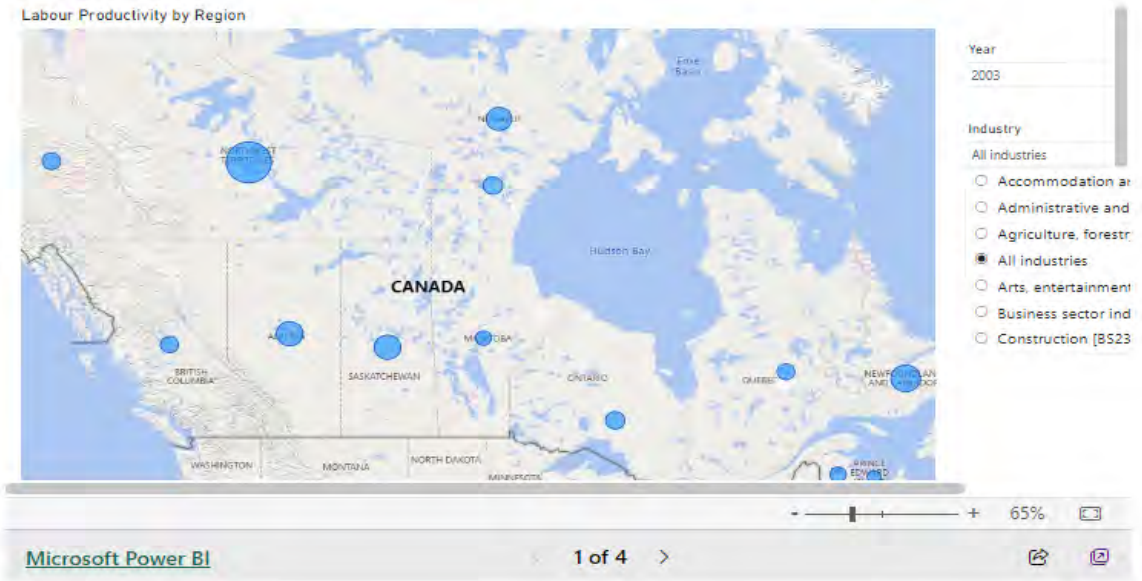
The above figure reports nominal value added across provinces and territories for 2022 in a pie chart.

LABOUR PRODUCTIVITY TRENDS
CURRENT SNAPSHOT
RESEARCH AND ANALYSIS
HOW TO CITE CLAPP DATA
CONTACT US
DEPARTMENT OF ECONOMICS

and the average standard of living.

For most industries and the aggregate, the level of labour productivity tends to grow positively over time. However, it can also fluctuate over shorter horizons for a variety of reasons. These reasons include changes in total/multi- factor productivity, capital utilization rates, and the capital-labour ratio. Canada's aggregate measure of labour productivity is also influenced by the allocation of economic activity across regions and industries.

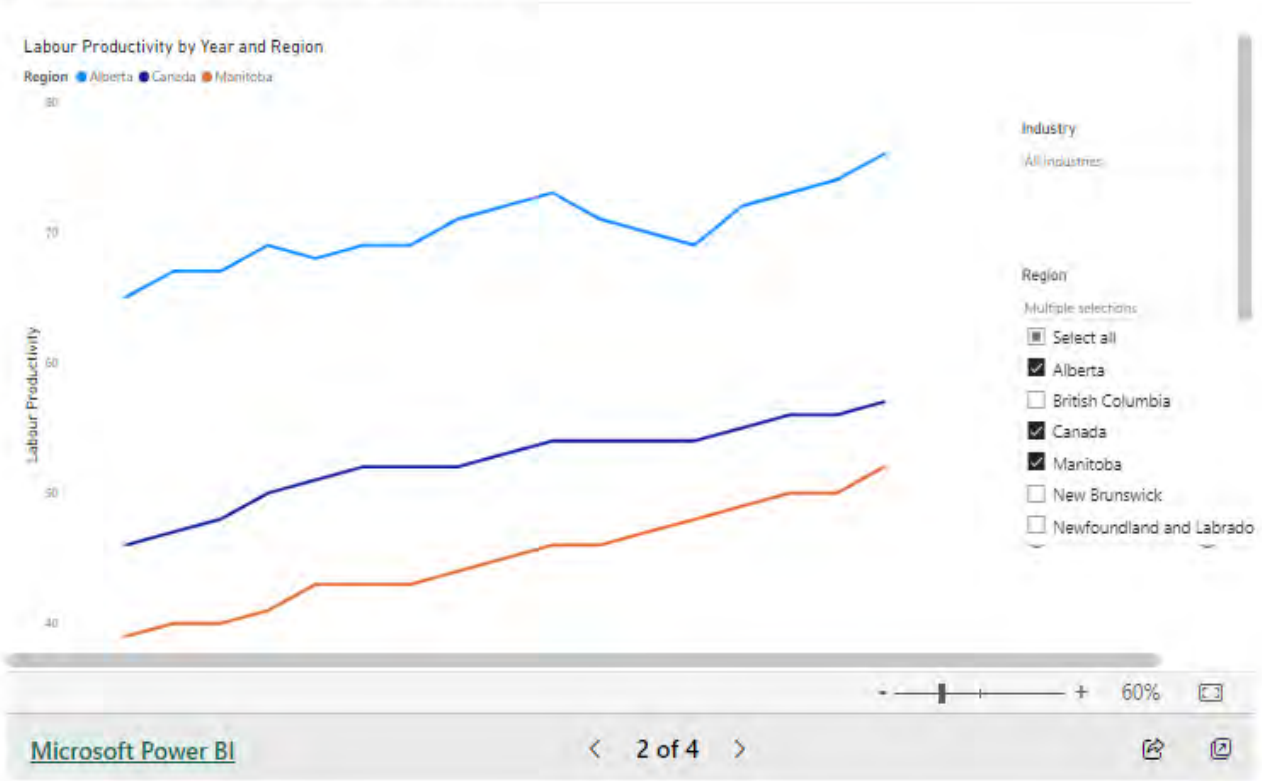
Bubble Map



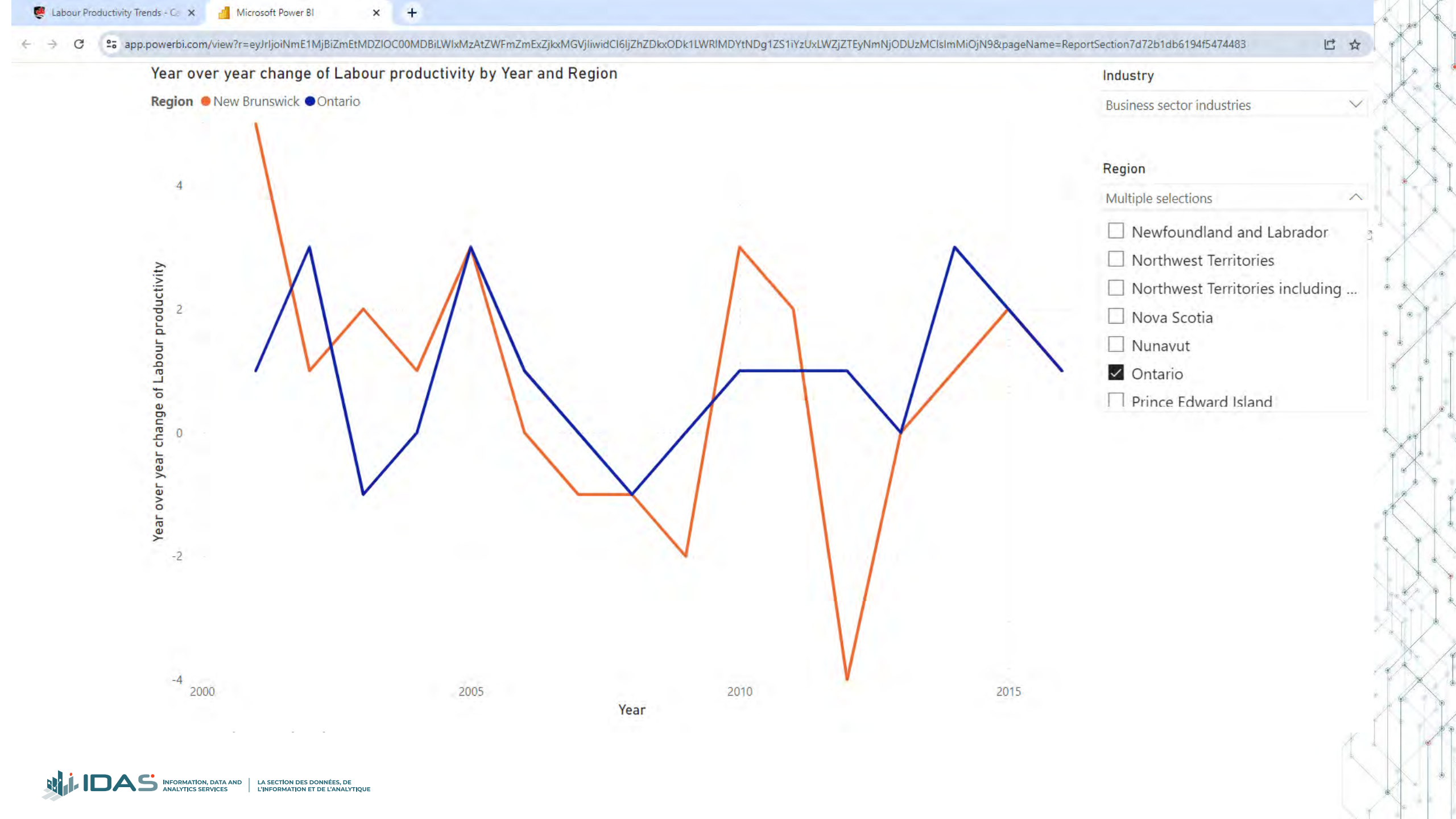
Labour productivity is defined as real output per hour worked. Real output is a value added measure and at the disaggregated level is constructed from a Fisher chain index evaluated in 2012 chained dollars. Hours worked is the total number of hours worked in all jobs, including overtime. The data is obtained from Statistics Canada Table 36-10-0480-01 (https://doi.org/10.25318/3610048001-eng) and the industry disaggregation is at level 5.

LABOUR PRODUCTIVITY TRENDS
CURRENT SNAPSHOT
RESEARCH AND ANALYSIS
HOW TO CITE CLAPP DATA
CONTACT US
DEPARTMENT OF ECONOMICS

Productivity Over Time



Since measures of real value added are constructed from Fisher chain indexes they are not additive, implying that summing real value added across industries does not yield Canada's real value added. Subsequently, it is often more informative to compare the growth rate of labour productivity across industries over time. The growth rate is computed as the log difference of the level of labour productivity multiplied by 100.



IDAS INFORMATION, DATA AND ANALYTICS SERVICES | LA SECTION DES DONNÉES, DE L'INFORMATION ET DE L'ANALYTIQUE

A growth accounting analysis of unit labour cost in Canada

- Unit labour cost (ULC) is a common metric central banks use to gauge inflationary pressures in the economy and directly related to data in CLaPP
- ULC also features prominently in determining inflation dynamics of modern structural macroeconomic models used at central banks for policy analysis and forecasts

Definition of ULC

$$ULC_t = \frac{w_t}{lp_t} = \frac{Real\ wage}{labour\ productivity}$$

$w_t = \frac{W_t}{P_t}$: Real wage at time t

$lp_t = \frac{y_t}{H_t}$: Labour productivity at time t

P_t: Price level at time t

W_t: Nominal wage at time t

H_t: Number of hours worked at time t

y_t: Output at time t

A Growth accounting analysis of unit labour cost in Canada--Continued

- Aggregate values of ULC at the National level conceal
  - (i) the contributions of the sub-National components
  - (ii) the allocation of economic activities
- We present a decomposition of the growth rate of ULC by decomposing its two components, i.e., wages and labour productivity, across geographical regions and industry sectors
- The decomposition assesses the contributions of
  - (i) the pure growth
  - (ii) the reallocations

Decomposition of ULC's growth

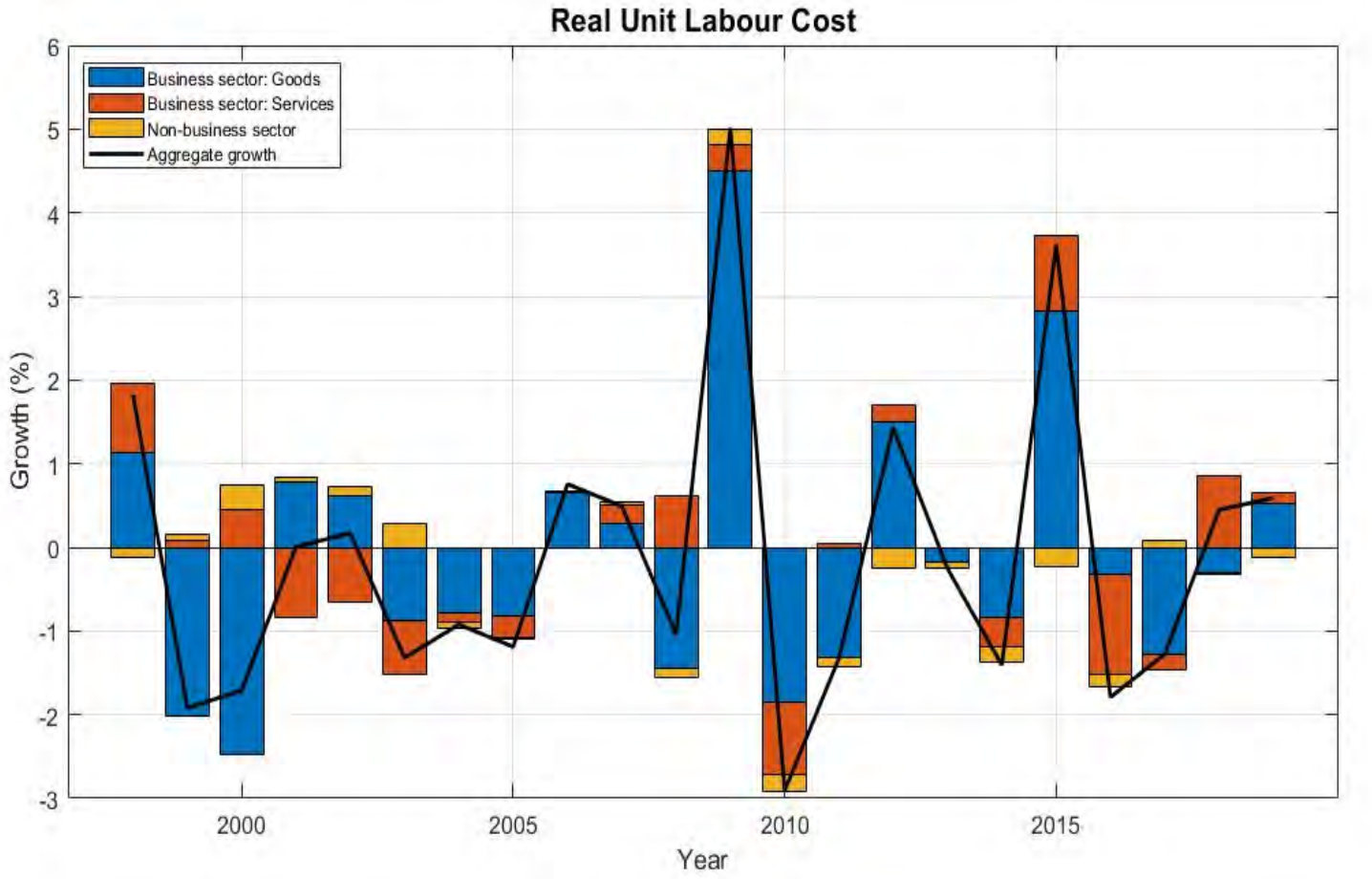
$$\frac{\Delta w_t}{w_{t-1}} = \sum_{j \in J} \sum_{i \in I} \left(\overbrace{\frac{w_{i,j,t-1}}{w_{t-1}} s_{i,j,t} \, \Delta w_{i,j,t}}^{Pure\ growth} + \overbrace{\frac{w_{i,j,t-1}}{w_{t-1}} \Delta s_{i,j,t}}^{Re-allocation} \right)$$

$$\frac{\Delta lp_t}{lp_{t-1}} = \sum_{j \in J} \sum_{i \in I} \left(\underbrace{\frac{lp_{i,j,t-1}}{lp_{t-1}} s_{i,j,t} \, \Delta lp_{i,j,t}}_{Pure\ growth} + \underbrace{\frac{lp_{i,j,t-1}}{lp_{t-1}} \Delta s_{i,j,t}}_{Re-allocation} \right)$$

- $s_{ijt} = \frac{H_{ijt} P_{ijt}}{P_t H_t}$: Labour input share adjusted for relative output prices in region j and industry i at time t
- Even in the absence of wage/productivity growth, the aggregate can increase if labour input share adjusted for relative output prices increases in areas/industries with higher relative wages/productivity

Some findings and results

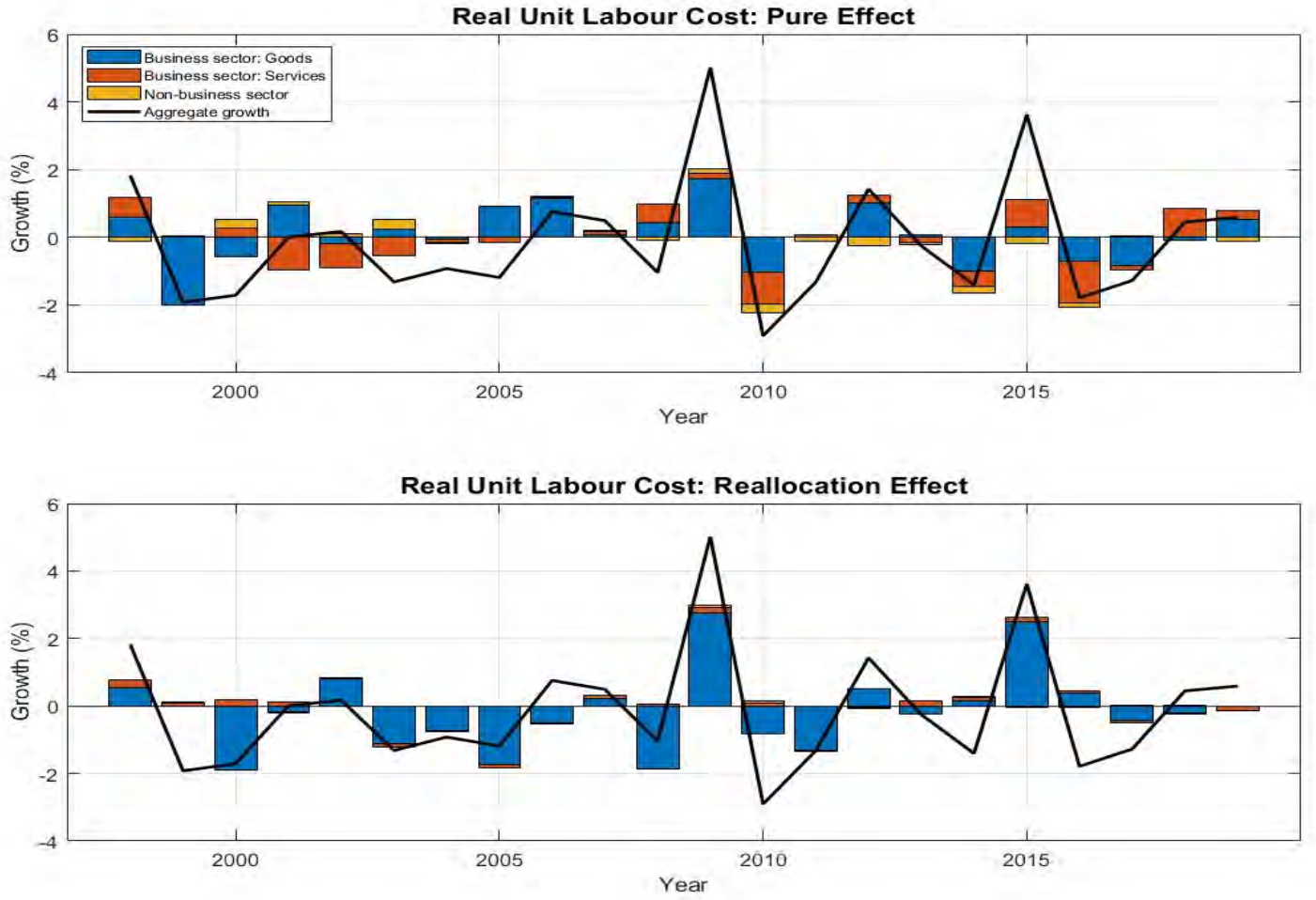
Data: Annual values from 1997 to 2019



The goods sectors account for a large proportion of ULC variation, even though it only accounts for roughly one quarter of the labour input share adjusted for output prices

Some findings and results (continued)

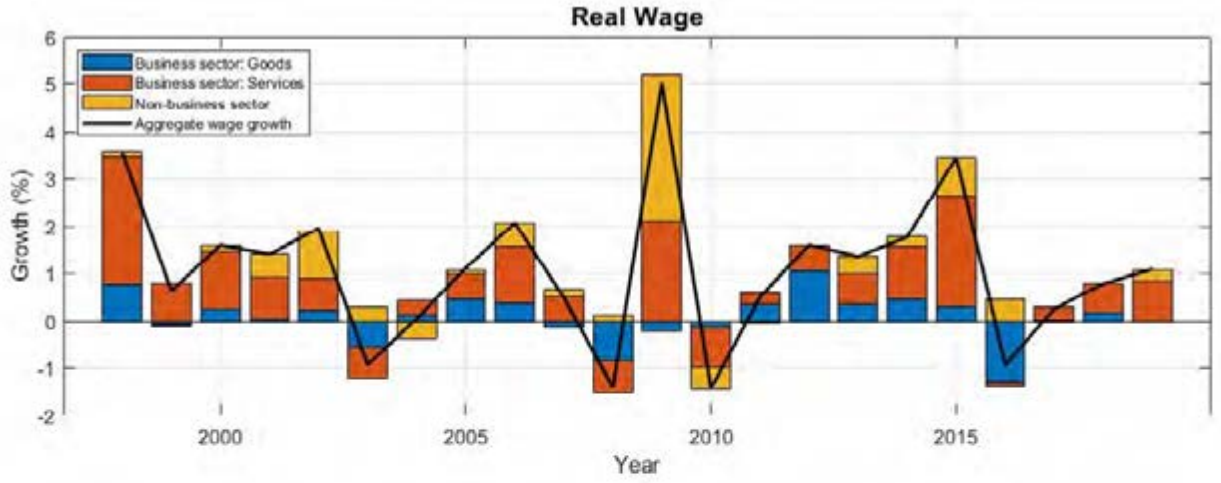
Data: Annual values from 1997 to 2019

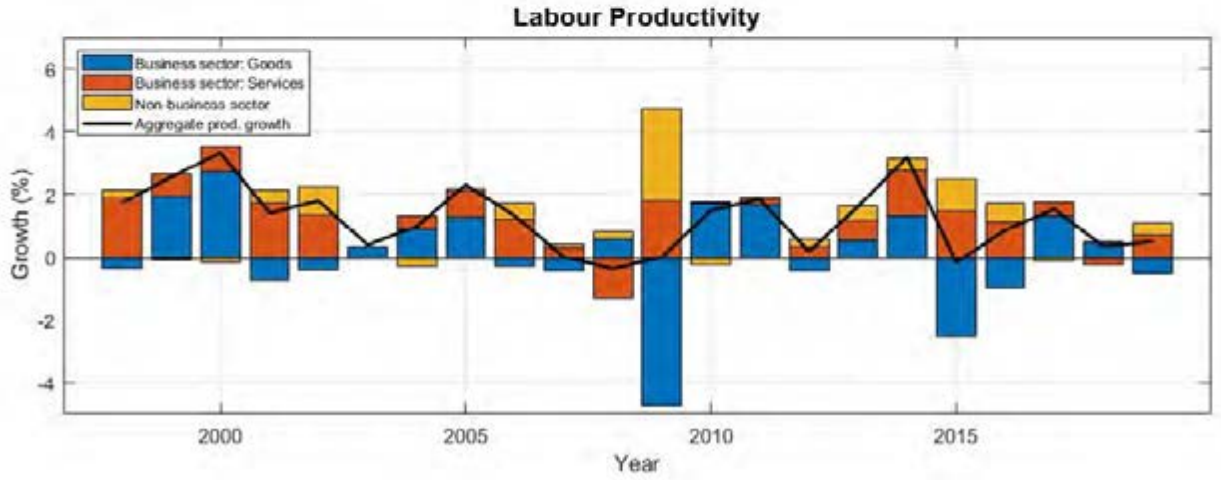


Reallocation effects are playing a large role

Decompositions of wages and labour productivity

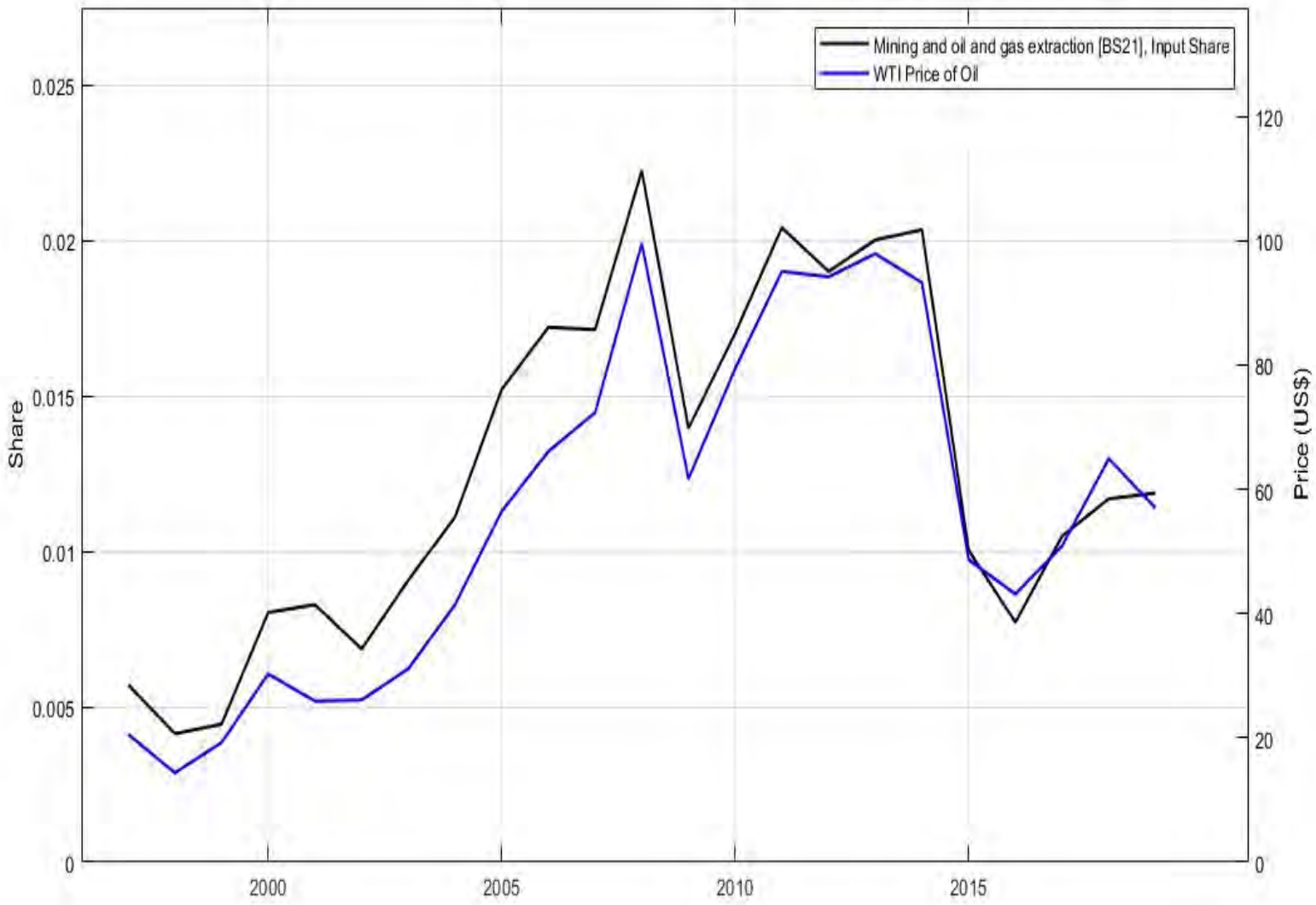
Data: Annual values from 1997 to 2019





The input share of Mining, Oil and Gas

Data: Annual values from 1997 to 2019



Future plans

- Identifying common trend(s) and cyclic co-movement of labour productivity data across geographical regions and sectors
- Determining convergence or divergence of labour productivity across regions/sectors relative to the aggregate values
- Producing decompositions of wages, labour productivity and ULC at a higher frequency than annual
- Studying the effect of AI on labour productivity, wage dispersion and, income distribution
- Improving data visualization and analysis on CLaPP

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Sectoral uncertainty[1]

Efrem Castelnuovo,
University of Padova,

and

Kerem Tuzcuoglu and Luis Uzeda,
Bank of Canada

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel—2022-38

Last updated: September 9, 2022



Sectoral Uncertainty

by Efrem Castelnuovo,[1] Kerem Tuzcuoglu[2] and Luis Uzeda[3]

[1]University of Padova, CESifo and Centre for Applied Macroeconomic Analysis
efrem.castelnuovo@unipd.it

[2]Financial Stability Department
Bank of Canada
ktuzcuoglu@bank-banque-canada.ca

[3]Canadian Economic Analysis Department
Bank of Canada
luzedagarcia@bank-banque-canada.ca

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

DOI: https://doi.org/10.34989/swp-2022-38 | **ISSN** 1701-9397

©2022 Bank of Canada

Acknowledgements

For their comments and suggestions, we would like to thank Carola Binder, Siddhartha Chib, Todd Clark, Gary Koop, Sharon Kozicki, André Kurmann, Rodrigo Sekkel, Soojin Jo, Giovanni Pellegrino, Ivan Petrella, Juan F. Rubio-Ramírez and conference and seminar participants at the 2021 Padova Macro Talks, ESCoE Conference on Economic Measurement 2021, 2020 NBER-NSF SBIES Conference, Econometric Society World Congress 2020, European Winter Meetings of the Econometric Society 2020, 14th International Conference on Computational and Financial Econometrics, 24th Central Bank Macroeconomic Modeling Workshop, The 56th Annual Conference of the Canadian Economics Association and Bank of Canada. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Bank of Canada.

Abstract

We propose a new empirical framework that jointly decomposes the conditional variance of economic time series into a common and a sector-specific uncertainty component. We apply our framework to a large dataset of disaggregated industrial production series for the US economy. Our results indicate that common uncertainty and uncertainty linked to non-durable goods both recorded their pre-pandemic global peaks during the 1973-75 recession. In contrast, durable goods uncertainty recorded its pre-pandemic peak during the global financial crisis of 2008-09. Vector autoregression exercises identify unexpected changes in durable goods uncertainty as drivers of downturns that are both economically and statistically significant, while unexpected hikes in non-durable goods uncertainty are expansionary. Our findings suggest that: (i) uncertainty is heterogeneous at a sectoral level; and (ii) durable goods uncertainty may drive some business cycle effects typically attributed to aggregate uncertainty.

Topics: Business fluctuations and cycles; Econometric and statistical methods; Monetary policy and uncertainty
JEL codes: E32, E44, C51, C55

Résumé

Nous proposons un nouveau cadre empirique qui permet de décomposer simultanément la variance conditionnelle des séries chronologiques de données économiques en deux facteurs : l'incertitude agrégée et l'incertitude sectorielle. Nous appliquons notre cadre à un vaste ensemble de données désagrégées relatives à la production industrielle aux États-Unis. Nos résultats indiquent qu'avant la pandémie, l'incertitude agrégée et l'incertitude liée aux biens non durables ont toutes deux atteint leur sommet durant la récession de 1973-1975. L'incertitude liée aux biens durables a quant à elle culminé pendant la crise financière mondiale de 2008-2009. Des exercices d'autorégression vectorielle permettent d'établir que les variations imprévues de l'incertitude liée aux biens durables sont des facteurs de ralentissement économiquement et statistiquement significatifs, tandis que les hausses inattendues de l'incertitude liée aux biens non durables ont une action expansionniste. Nos résultats donnent à penser que 1) l'incertitude est hétérogène au niveau sectoriel, et 2) l'incertitude liée aux biens durables peut être à l'origine de certains effets du cycle économique qui sont habituellement attribués à l'incertitude agrégée.

Sujets : Cycles et fluctuations économique; Méthodes économétriques et statistiques; Incertitude et politique monétaire
Codes JEL : E32, E44, C51, C55

1 Introduction

Recessions are typically associated with bouts of uncertainty, a stylized fact that has generated considerable research on the role of uncertainty as a driver of the business cycle.[1] More recently, policy institutions such as the International Monetary Fund have raised the importance of implementing sectoral policies to combat the recessionary effects of uncertainty.[2] While such policies have been advocated at the sectoral level, sectoral uncertainty itself remains a theme that has been little explored to date. This may in part be attributed to the technical challenges a researcher has to face to discern what portion of broader (aggregate) uncertainty reflects common (or across-sector) and sectoral (or sector-specific) components. Conceptually, such a task entails (i) the estimation of uncertainty at different layers of economic data; and (ii) the modeling of a reasonably large dataset to pin down all the implications that different dimensions of the economic system may have on uncertainty.

To address points (i) and (ii), this paper proposes a new dynamic factor model that allows for the joint estimation of common and sectoral uncertainty within a data-rich environment. Our strategy is based on a hierarchical factor setting in the spirit of, e.g., Kose et al. (2003), Del Negro and Otrok (2007), Moench et al. (2013), and Gorodnichenko and Ng (2017) that accommodates the different degrees of volatility pervasiveness in economic data. We define uncertainty as the conditional volatility of the unpredictable components of economic indicators. In particular, we parameterize *common* uncertainty as a dynamic factor that is common to the time-varying volatility of *all* variables in a large system. In contrast, *sectoral* uncertainty is modeled as a dynamic factor that is common to the time-varying volatility of a *subset* of variables corresponding to a particular sector. As a result, our methodology breaks down the (conditional) variance of each time series in our setting into two components, a common and sectoral one.

[1] Theoretical frameworks formalizing channels that capture the effects of uncertainty shocks on real activity have been proposed by, among others, Bloom (2009), Fernández-Villaverde et al. (2011), Fernández-Villaverde et al. (2015), Leduc and Liu (2016), Basu and Bundick (2017), Bloom et al. (2018), Bianchi et al. (2021), and Born and Pfeifer (2021). Empirical investigations identifying uncertainty as a driver of the business cycles include Caggiano et al. (2014), Jurado et al. (2015), Baker et al. (2016), Caldara et al. (2016), Angelini et al. (2019), Carriero et al. (2016, 2018), Ludvigson et al. (2021), and Coibion et al. (2021). For surveys, see Bloom (2014), Fernández-Villaverde and Guerron-Quintana (2020), and Cascaldi-Garcia et al. (2021).

[2] See, e.g., https://blogs.imf.org/2020/11/19/continued-strong-policy-action-to-combat-uncertainty/.

We apply our new empirical framework to a rich dataset of U.S. industrial production that comprises 185 industry-level production series grouped into two sectors, namely durable and non-durable goods over a sample period from 1972Q1 to 2019Q4.[3] In accordance with conventional wisdom, we document common uncertainty to be countercyclical, display sharp increases during recessions, and feature a positive correlation with popular uncertainty measures such as those of Jurado et al. (2015), Baker et al. (2016), and Ludvigson et al. (2021). Our sectoral uncertainty measures, however, are found to behave heterogeneously. Uncertainty in the nondurable goods sector displays its global peak during the 1973–75 recession, a feature shared with our common uncertainty factor. In contrast, uncertainty in the durable goods sector peaks during the Great Recession of 2008–09, an episode in recent U.S. history clearly characterized by a massive increase in uncertainty (Blanchard, 2009). Furthermore, similar to common uncertainty, we find durables uncertainty to be more tightly linked to a variety of business cycle indicators.

The last part of the paper carries out a VAR analysis that involves a battery of standard macroeconomic indicators (similar to those modeled by Bloom (2009), Fernández-Villaverde et al. (2015), Jurado et al. (2015), Basu and Bundick (2017), and Born and Pfeifer (2021)) and our novel measures of common and sector-specific uncertainty. We document unexpected hikes in uncertainty to generate different responses of real activity depending on the type of uncertainty one considers. In particular, we find three distinct patterns. First, unexpected changes in *common* uncertainty trigger a "drop-rebound-overshoot" dynamic response to real activity in line with the one documented by Bloom (2009).[4] Second, shocks to *durables* uncertainty generate an inverse hump-shaped (or contractionary) response to real activity. Such a result connects our analysis with the literature on the "wait-and-see" (or "real options") behavior of firms and consumers due to non-convex adjustment

[3]Given the set of unique circumstances that characterize the COVID-19 pandemic period, extending our analysis to address post-2019Q4 developments arguably warrants a separate study to focus exclusively on uncertainty (common and sectoral) within this period. That said, we stress that our framework exhibits features, such as stochastic volatility and outlier adjustment, that are desirable to model the extreme time-series values witnessed during the COVID-19 pandemic. Also, while we apply our methodology to disaggregated U.S. industrial production data (as, e.g., Foerster et al. (2011) and Garin et al. (2018)), our framework lends itself to a variety of applications in macroeconomics and finance where separating common and sector-specific volatility is important.

[4]Bloom (2009) employs Hodrick-Prescott filtered data in his VAR. Caggiano et al. (2022) show that Bloom's (2009) results are also robust to using unfiltered data.

costs, typically associated with durable goods, as in Bloom (2009) and Bloom et al. (2018). Third, and against conventional wisdom, shocks to *nondurable* goods uncertainty are found to generate an expansionary response in production and employment. Taken together, our results suggest that the countercyclicality of uncertainty, as is usually documented, may in fact be due to a specific subset of goods in the economy, namely durables, rather than the manifestation of a broader phenomenon. Consequently, a complete account of the business cycle effects of uncertainty seems to call for the modeling of sectoral heterogeneities.

We also conduct the above-described VAR analysis on a shorter sample that begins in 1984, which is often referred to as the beginning of the Great Moderation period. Such an exercise reveals the effects of sectoral uncertainty on real activity to be even stronger when compared to results from a sample that also encompasses the 1970s. This last finding, albeit related to second-moment sectoral shocks, matches the timing identified by Foerster et al. (2011), Atalay (2017), and Garin et al. (2018), who document the increased role of first-moment sectoral shocks for the determination of the U.S. business cycle since the mid-1980s. Our VAR results also echo the findings put forth by Horvath (1998, 2000), Conley and Dupor (2003), Acemoglu et al. (2012), Acemoglu et al. (2017), and Baqaee and Farhi (2019), who assert that sectoral shocks can lead to aggregate fluctuations due to network effects, input-output linkages between sectors, nonlinearities, and the presence of large and crucial sectors.

Our paper joins recent research that has also taken a sectoral view on uncertainty. Choi and Loungani (2015) employ stock market data to construct both an aggregate measure of time-series volatility in stock returns and a sectoral measure of cross-industry uncertainty in stock returns. They find sectoral uncertainty to have a different impact on unemployment than common uncertainty. Segal (2019) measures sectoral uncertainty as the shocks to the volatility of sectoral total factor productivity. Akin to our study, the author also provides evidence on the heterogeneous effects of sectoral uncertainty on economic aggregates. Specifically, uncertainty originating in the consumption sector has contractionary effects on macroeconomic growth rates, while investment-related uncertainty has expansionary economic effects. Ma and Samaniego (2019) measure uncertainty as the median absolute forecast errors drawn from a large firm-level dataset. In their setting, median absolute

forecast errors obtained from the full dataset are called common uncertainty, while those obtained from a selected subset of industries are interpreted as sectoral uncertainty.

We differ from these three studies in two key dimensions. First, and to the best of our knowledge, we are the first study to estimate common and sectoral uncertainty *jointly*. Hence, by explicitly parameterizing sectoral uncertainty as a net-of-common component, our setting provides an avenue to sharpen the inference of uncertainty in contexts where defining the latter requires a higher level of granularity. Put differently, our approach reduces the likelihood of confounding common and sectoral uncertainty dynamics, which is a risk one runs when estimating each measure independently (as in Ma and Samaniego (2019)), or sequentially via two-step procedures (as in Choi and Loungani (2015) and Segal (2019)).[5] Second, unlike previous studies, our take on sectoral uncertainty focuses on the goods sector, which is traditionally a sector that exhibits salient dynamics during recessions and downturns (see, e.g, Wynne and Balke (1993) and Kehoe et al. (2018)).

Our paper also adds to recent investigations that combine factor and stochastic volatility methods to deal with uncertainty measurement, such as Carriero et al. (2016, 2018), Jo and Sekkel (2019), and Mumtaz and Musso (2019). In this regard, our contribution is twofold. First, we extend the parametrization for common stochastic volatility models proposed in Carriero et al. (2016, 2018) by combining hierarchical-factor and outlier-adjustment methods that allow for a more flexible representation of second moments in our model. Second, to carry out estimation, we develop an efficient Markov Chain Monte Carlo (MCMC) algorithm that relies on precision sampling methods, as in, e.g., Chan and Jeliazkov (2009). As pointed out in McCausland et al. (2011), precision-based methods typically reduce computational complexity and expedite estimation relative to Kalman filter-based approaches, which are used for state-space inference more broadly.

The remainder of this paper proceeds as follows. Section 2 introduces our empirical framework. Section 3 discusses the estimation of our proposed model. Empirical results are given in Section 4. Section 5 concludes. An Online Appendix contains further details on the dataset, estimation technique, and robustness checks.

[5]Regarding two-step methods, while certainly a convenient strategy, they are prone to be affected by the generated regressors problem. Such a problem, as pointed out by Pagan (1984) and Kim and Kim (2011) (also in the context of estimating unobservable series), could ultimately lead to invalid inferences due to potential cumulative measurement errors from previous estimation steps.

2 A Hierarchical Common Stochastic Volatility Model

This section describes our novel empirical framework designed to measure uncertainty in the goods sector at different levels of aggregation. The first issue at hand is how to interpret uncertainty. A common and intuitive way to do so is to treat uncertainty as the volatility of the unpredictable component embedded in a dataset of interest. Such a characterization can be parameterized in the context of a linear regression setting, that is:

$$y_t = X_t\boldsymbol{\beta} + u_t, \tag{1}$$

where the conditional expectation $\mathbb{E}_t(y_t) = X_t\boldsymbol{\beta}$ provides an optimal prediction (in a mean squared error sense) for the $N \times 1$ vector y_t containing a set of economic time-series with observations going from 1 to T, where N identifies the number of industries in our dataset, and T the time-series dimension. Therefore, unpredictability (and, consequently, uncertainty) is associated with the error term u_t.

2.1 Factor-Based Controls

To ensure u_t is unpredictable (or approximately so), it is convenient to define X_t in a manner that extracts as much predictable content as possible from y_t while not overfitting the expression in Equation (1). In our exercise, y_t accounts for 185 (standardized) production growth series at the quarterly frequency from 1972Q1 until 2019Q4. These series are obtained from disaggregated IP data that cover a cross-section of industries in both the durable goods and nondurable goods sectors.[6] A detailed description of the data is provided in Section A3 of the Online Appendix.

Given the large dimension of y_t, adopting a more traditional framework such as a VAR to model X_t would represent a less tractable avenue computationally. Consequently, we

[6]Our definition of a sector is based on the North American Industry Classification System (NAICS) where the first two digits in the NAICS structure designate the sector to which an industry belongs. Also, we opt to work with quarterly data since monthly disaggregated IP series are substantially "noisier," and as pointed out by Miron and Zeldes (1989), possibly affected by significant measurement errors in the earlier part of the sample. At the quarterly frequency such errors may, in principle, become less consequential as monthly data is averaged over the quarter. Moreover, quarterly data is a conventional frequency for business cycle-related analysis.

follow studies such as Stock and Watson (2016) and McCracken and Ng (2016) that propose factor-based controls to construct X_t in data-rich environments. Specifically, we organize covariates in X_t into two types of factors, namely factors that capture economic conditions more broadly and factors that capture IP-specific dynamics.

For the first category, we follow McCracken and Ng (2016) and use principal-component techniques to extract common factors from a large cross-section of U.S. economic indicators. These indicators span a wide range of variables within broad categories (e.g., production, labor market, prices, financial markets), which we acquire from the Federal Reserve Economic Macroeconomic Database (FRED-MD hereafter).[7] The testing procedure of Bai and Ng (2002) is then employed to select the optimal number of factors (seven in our application). These factors are then integrated into X_t as "observable" measures. Given the large cross-section of data used to extract the principal components, the generated regressors problem that is commonly associated with treating principal components as actual data becomes less of an issue in the context of our exercise (see Bai and Ng (2008)).

Regarding the IP-specific factors, we treat them as latent factors that are estimated jointly with the other states and parameters in our model (estimation details are discussed in Section 3). To determine the number of IP-specific factors, we apply once again the testing procedure of Bai and Ng (2002), except now the method is applied to the 185 IP growth series in y_t. Since we model these latent factors as a VAR process, we check both the static and dynamic versions of the test statistic proposed by Bai and Ng (2002). We also employ a robust test statistic proposed by Alessi et al. (2010). All these criteria point to four IP-specific factors. Below we summarize how X_t is structured:

$$\texttt{Controls:} \begin{cases} X_t = [I_N \otimes z_t' \quad I_N \otimes f_t'], \\ z_t = (z_{1,t}, \cdots, z_{7,t})' \quad - \text{McCracken and Ng (2016) factors}, \\ f_t = (f_{1,t}, \cdots, f_{4,t})' \quad - \text{IP-specific factors}, \\ f_t = \Phi_f f_{t-1} + \eta_{f,t} \quad \eta_{f,t} \sim \mathcal{N}\left(0, \ \texttt{diag}\left(\sigma_{f_1}^2, \cdots, \sigma_{f_4}^2\right)\right), \end{cases} \tag{2}$$

where I_N and $\otimes$ denote an N-dimensional identity matrix and the Kronecker product,

[7] Details on the FRED-MD database can be found in Section A3 of the Online Appendix and at the Federal Reserve Bank of St. Louis website.

respectively. The seven factors from the FRED-MD database and the four IP-specific factors are collected by z_t and f_t, respectively. The latter, as shown in the last expression in (2), is modeled as a (four-variable) first-order VAR with homoskedastic innovations. The constant-variance assumption is relaxed in Section 4.5, where we allow for stochastic volatility to model the conditional volatility of $\eta_{f,t}$ in the context of robustness checks applied to our baseline framework. We note, however, that our key findings are largely unchanged when stochastic volatility is introduced to model second-moment dynamics for f_t.

2.2 Identifying Sectoral Uncertainty

The notion of uncertainty derived from (1) has represented the building block for various studies, as in, *inter alia*, Jurado et al. (2015), Carriero et al. (2016), Jo and Sekkel (2019), and Clark et al. (2020). We thus build on these papers to adopt a hierarchical structure that characterizes uncertainty across the entire goods sector as well as within its sectoral subcomponents, namely the durable and nondurable sectors. To this end, we first cast u_t as:

$$u_t = \Psi_t^{\frac{1}{2}} \Sigma_t^{\frac{1}{2}} e_t, \quad s.t.\ e_t \sim \mathcal{N}(0,\ I_N), \tag{3}$$

$$\Sigma_t^{\frac{1}{2}} = \texttt{diag}\left(\exp(h_{1,t}/2), \cdots, \exp(h_{N,t}/2)\right), \tag{4}$$

$$\Psi_t^{\frac{1}{2}} = \texttt{diag}\left(\psi_{1,t}^{\frac{1}{2}}, \cdots, \psi_{N,t}^{\frac{1}{2}}\right). \tag{5}$$

Equations (3)–(5) indicate that the evolution of the volatility of u_t is expressed in terms of two time-varying components, which we define as follows: (i) $\Sigma_t^{\frac{1}{2}}$ captures volatility changes that are associated with uncertainty dynamics across and within the subcomponents of the goods sector; such dynamics are modeled in the tradition of stochastic volatility models (see, e.g., Kim et al. (1998) and Omori et al. (2007)) and absorbed by the state variables (or log-volatilities) $h_{i,t}$ for $i = 1, ..., N$; and (ii) $\Psi_t^{\frac{1}{2}}$ captures changes in volatility that are idiosyncratic to each industry, i.e., to each element in u_t. We now turn to discuss the modeling of each of these time-varying volatility components in greater detail.

Our goal is to formulate a framework that distinguishes between uncertainty that is

common across and within the durables and nondurables sectors. Also, in keeping with uncertainty-based theories of the business cycle, which typically presuppose common variation in uncertainty across a large number of series, we want our framework to generate uncertainty measures that reflect pervasiveness. As a result, we adopt a factor-based approach that extracts comovement in second moments at different levels of aggregation. In particular, we build upon the work on common stochastic volatility models by Carriero et al. (2016, 2018) to parameterize the vector of log-volatilities $h_t = (h_{1,t}, \cdots, h_{N,t})'$ in (4) as follows:

$$h_t = \Lambda_c v_{c,t} + \Lambda_s v_{s,t}. \tag{6}$$

Equation (6) implies that uncertainty is decomposed into two parts: (i) a factor $v_{c,t}$ that captures the *common* component of uncertainty across all industries in the goods sector; and (ii) a vector $v_{s,t} = (v_{d,t}\ v_{nd,t})'$, where each element denotes a factor that captures the *sector-specific* component of uncertainty in the durables ($v_{d,t}$) and nondurables ($v_{nd,t}$) sectors. To sum up, $v_{c,t}$ and $v_{s,t}$ collect our estimates (in log-volatility form) of across- and within-sector uncertainty, respectively.[8]

To identify uncertainty at such different levels of aggregation, we parameterize the factor-loading matrices as follows:

$$\Lambda_c = \begin{bmatrix} \lambda_{c,1} \\ \lambda_{c,2} \\ \vdots \\ \vdots \\ \lambda_{c,N} \end{bmatrix} \quad \text{and} \quad \Lambda_s = \begin{bmatrix} \lambda_{d,1} & 0 \\ \vdots & \vdots \\ \lambda_{d,S_d} & 0 \\ 0 & \lambda_{nd,1} \\ \vdots & \vdots \\ 0 & \lambda_{nd,S_{nd}} \end{bmatrix}, \tag{7}$$

where S_d and S_{nd} denote the number of industries in the durable and nondurable sectors, respectively (i.e., $S_d + S_{nd} = N$). The characterization above implies that common uncertainty loads on all N industries in our dataset, while durable- and nondurable-specific

[8] Once we obtain volatility estimates in the logarithmic scale, we convert results to conditional variances by computing $\exp(v_{j,t})$ for $j = c,\ d,\ nd$.

uncertainty loads only on the industries corresponding to these two sectors. Also, as is standard in factor models, $\lambda_{c,1}$, $\lambda_{d,1}$, and $\lambda_{nd,1}$ are normalized to one in order to separately identify $v_{c,t}$ and $v_{s,t}$ from Λ_c and Λ_s, respectively. In particular, our normalization strategy for the sectoral uncertainty factors follows from multilevel-factor studies such as Moench et al. (2013).[9] In practice, the parameterization given by (6) and (7) implies that the level of overall uncertainty for each industry (i.e., $\exp(h_{i,t}/2)$ for $i = 1, ..., N$) is governed by an economy-wide and a sector-specific component. This nuanced representation is useful as it will allow us to examine potentially heterogeneous effects of goods-related uncertainty on economic activity, a theme we return to in Section 4.3.

Our framework also accommodates the possibility that uncertainty may "spillover" across and within sectors. Such interdependencies are formulated by letting the common and sector-specific factors described above evolve as a VAR process of order one, i.e.:

$$\underbrace{\begin{bmatrix} v_{c,t} \\ v_{d,t} \\ v_{nd,t} \end{bmatrix}}_{v_t} = \Phi_v \underbrace{\begin{bmatrix} v_{c,t-1} \\ v_{d,t-1} \\ v_{nd,t-1} \end{bmatrix}}_{v_{t-1}} + \underbrace{\begin{bmatrix} \eta_{c,t} \\ \eta_{d,t} \\ \eta_{nd,t} \end{bmatrix}}_{\eta_{v,t}}, \tag{8}$$

where we assume the errors in the vector η_t are normally distributed as well as mutually and serially orthogonal to each other.[10]

Note that the parameterization in (6) builds on the work of Carriero et al. (2016, 2018), but provides a key extension to them. Specifically, Carriero et al. (2016), in the context of a VAR model, allow for a single common factor driving the conditional volatility of all variables in the system, but do not explore the idea of group-specific common volatility. In contrast, Carriero et al. (2018) explore the latter idea to distinguish between macroeconomic and financial uncertainty, but do not account for a common factor between these two

[9]The block exogeneity restrictions imposed on the loadings provide an intuitive way to identify the latent factors in our model. We acknowledge, however, the existence of other data-driven approaches to interpret factors, such as checking the correlations between the factors and series (Ludvigson and Ng, 2009) and applying shrinkage methods to the loadings (Hacioglu Hoke and Tuzcuoglu, 2016).

[10]We assume the errors in η_t are mutually orthogonal to be consistent with our definition that the common uncertainty factor is the term that captures the (uncertainty-related) contemporaneous correlation amongst the durables and nondurables sectors. Consequently, the remaining errors driving the (sector-specific) factors in (8) should be orthogonal to the rest of the system.

groups. In other words, in light of (6), the proposal put forth by Carriero et al. (2016) can be interpreted as imposing the restriction $\Lambda_s = 0$, while Carriero et al. (2018) impose $\Lambda_c = 0$. Our framework nests both cases. We view the joint modeling of across- and within-group comovement as crucial to avoid potential distortions in the measurement of common and sector-specific uncertainty.

2.3 Idiosyncratic Volatility

Given the large cross-section of industries ($N = 185$) that make up the durables and nondurables sectors in our dataset, it is likely that the production from these industries may exhibit substantial variability amongst themselves and over time. This suggests that the volatility associated with the residual term u_t might be, in part, idiosyncratic rather than pervasive. As show in (3), we accommodate such a possibility by setting $u_t = \Psi_t^{\frac{1}{2}} \Sigma_t^{\frac{1}{2}} e_t$, where industry-specific volatility changes are captured by each element in the scale matrix $\Psi_t^{\frac{1}{2}} = \texttt{diag}\left(\psi_{1,t}^{\frac{1}{2}}, \cdots, \psi_{N,t}^{\frac{1}{2}}\right)$. More precisely, we define:

$$\psi_{i,t} \overset{i.i.d.}{\sim} \mathcal{IG}\left(\frac{\nu_\psi}{2}, \frac{\nu_\psi}{2}\right) \text{ for } i = 1, \cdots, N \text{ and } t = 1, \cdots, T, \tag{9}$$

where $\mathcal{IG}$ denotes the inverse-gamma distribution. Such a choice of probability distribution and its parameterization follows directly from, e.g., Jacquier et al. (2004) and Chib et al. (2006), where the authors adopt a similar decomposition of the error covariance matrix to the one in (3). Notably, in addition to absorbing series-specific volatility, $\psi_{i,t}$ turns the distribution of the composite error term $u_t = \Psi_t^{\frac{1}{2}} \Sigma_t^{\frac{1}{2}} e_t$, marginalized over $\Psi_t^{\frac{1}{2}}$, into a random variable that follows a Student-t distribution with ν_ψ degrees of freedom.[11] As a result, since the t-distribution is a fat-tailed distribution, the introduction of Ψ_t provides our framework with a form of outlier adjustment. This is empirically useful as it allows our framework to accommodate a distinction between what constitutes sizeable series-specific volatility and uncertainty dynamics. The latter, as previously argued, tend to viewed as a

[11] See, e.g., Koop et al. (2007), chapter 15, for a formal presentation of the above-mentioned result. To allow the error distribution to have (*a priori*) heavier tails than those of the Gaussian distribution, we set the degree of freedom parameter ν_ψ at six.

more widespread phenomenon.[12]

An alternative way to model idiosyncratic volatility is the one proposed by Carriero et al. (2018), who introduce a series-specific error term into the log-volatility state equation (6) to model 18 macroeconomic time-series. Unfortunately, this approach is unfeasible in our framework, because this would require us to augment our posterior simulation algorithm to estimate as many log-volatility states as there are variables in y_t, i.e., 185. Estimating such a large number of log-volatility states makes computation virtually prohibitive.

3 Estimation

Our framework constitutes a nonlinear state space model. As is common in such instances, we conduct estimation using Bayesian methods. In particular, we propose an estimation algorithm that is fully based on Gibbs sampling steps. Briefly, we develop an MCMC algorithm, whereby we retain 50,000 draws from the 60,000 runs of our MCMC sampler. The first 10,000 burn-in draws are discarded. Results for the mixing efficiency of the MCMC chain are presented in Section A1 of the Online Appendix. We note here that the mixing of our algorithm is quite good.

To facilitate the description of our estimation strategy, we make use of additional notation. Specifically, we define $\boldsymbol{\theta}$ and $\mathcal{Z}$ as the sets containing, respectively, the parameters and latent states in our model, while $\boldsymbol{\theta}_{-\mathbf{j}}$ and $\mathcal{Z}_{-\mathbf{j}}$ will be used as short notation when these sets contain all their elements except for j. Also, hereafter we make use of the following stacked representation for y_t, i.e., $\mathbf{y} = (y_1, \ \cdots, \ y_T)'$. At a high level, our estimation framework consists of two main blocks, namely parameter sampling and state simulation. Below we discuss these two blocks in turn.

3.1 Parameter Sampling

We begin by setting $\boldsymbol{\theta} = \{\tilde{\boldsymbol{\beta}}_z, \ \tilde{\boldsymbol{\beta}}_f, \ \tilde{\boldsymbol{\lambda}}_c, \ \tilde{\boldsymbol{\lambda}}_d, \ \tilde{\boldsymbol{\lambda}}_{nd}, \ \boldsymbol{\phi}_f, \ \boldsymbol{\phi}_v, \ \boldsymbol{\sigma}_f^2, \ \boldsymbol{\sigma}_v^2\}$, where the first five elements in $\boldsymbol{\theta}$ denote vectors collecting the loading coefficients in (1) and (6). To be clear,

[12] Following the COVID-19 pandemic, stochastic volatility models with outlier adjustment have witnessed a renewed interest. See, e.g., Antolín-Díaz et al. (2020), Carriero et al. (2020); Marcellino et al. (2021)), and Lenza and Primiceri (2021) for recent applications.

$\tilde{\boldsymbol{\beta}}_z$ collects the coefficients that are associated with the FRED-MD factors (z_t), which—as discussed in Section 2.1—are treated as observables. The vector $\tilde{\boldsymbol{\beta}}_f$ collects the coefficients that load on the IP-specific factors (f_t), which are treated as latent variables. As a result, $\tilde{\boldsymbol{\beta}}_f$ excludes coefficients that are normalized to one since these do not need to be sampled. The same rationale applies to $\tilde{\boldsymbol{\lambda}}_c$, $\tilde{\boldsymbol{\lambda}}_d$, and $\tilde{\boldsymbol{\lambda}}_{nd}$, which collect the loadings associated with the latent common, durables, and nondurables uncertainty factors, respectively. Next, let `vec` stand for the operator that transforms a matrix into a column vector. The terms $\boldsymbol{\phi}_f = \texttt{vec}(\Phi_f')$ and $\boldsymbol{\phi}_v = \texttt{vec}(\Phi_v')$ thus denote vectors collecting the coefficients in the VAR-based law of motion for f_t and v_t in (2) and (8), respectively. Lastly, $\boldsymbol{\sigma}_f^2$ and $\boldsymbol{\sigma}_v^2$ denote vectors containing the variances associated with the error terms in the state equations for f_t and v_t, respectively. The parameter-sampling block of our MCMC algorithm can thus be summarized by sequentially drawing from the following distributions:

$$\texttt{MCMC Steps:} \begin{cases} (1)\ p\left(\tilde{\boldsymbol{\beta}}_x|\mathbf{y},\mathcal{Z},\boldsymbol{\theta}_{-\tilde{\beta}_x}\right) & \text{for } x=z,\ f, \\ (2)\ p\left(\tilde{\boldsymbol{\lambda}}_j|\mathbf{y},\mathcal{Z},\boldsymbol{\theta}_{-\tilde{\boldsymbol{\lambda}}_j}\right) & \text{for } j=c,\ d,\ nd, \\ (3)\ p\left(\boldsymbol{\phi}_\ell|\mathbf{y},\mathcal{Z},\boldsymbol{\theta}_{-\boldsymbol{\phi}_\ell}\right), & \\ (4)\ p\left(\boldsymbol{\sigma}_\ell^2|\mathbf{y},\mathcal{Z},\boldsymbol{\theta}_{-\boldsymbol{\sigma}_\ell^2}\right) & \text{for } \ell=v,\ f. \end{cases} \tag{10}$$

To sample from each of the MCMC steps above, we adopt mutually independent priors. More specifically, for each element in $\tilde{\boldsymbol{\beta}}_x$, $\tilde{\boldsymbol{\lambda}}_j$, and $\boldsymbol{\phi}_\ell$ we elicit a Gaussian prior given by $\mathcal{N}\left(\hat{\mu}_l,\ \hat{\sigma}_l^2\right)$ for $l=\tilde{\beta}_x,\ \tilde{\lambda}_j,\ \phi_\ell$. For the coefficients in $\boldsymbol{\phi}_v$, $\boldsymbol{\phi}_f$, and $\tilde{\boldsymbol{\beta}}_z$ we set $\hat{\mu}_{\tilde{\beta}_z} = \hat{\mu}_{\phi_\ell} = 0$ and $\hat{\sigma}^2_{\tilde{\beta}_z} = \hat{\sigma}^2_{\phi_\ell} = 1$ such that $\ell = v,\ f$. Differently, for the loadings in $\tilde{\boldsymbol{\beta}}_f$, $\tilde{\boldsymbol{\lambda}}_c$, $\tilde{\boldsymbol{\lambda}}_d$, and $\tilde{\boldsymbol{\lambda}}_{nd}$, akin to Carriero et al. (2017), our priors are more tightly parameterized around the normalization condition, i.e., $\hat{\mu}_{\tilde{\beta}_f} = \hat{\mu}_{\tilde{\lambda}_j} = 1$ and $\hat{\sigma}^2_{\tilde{\beta}_f} = \hat{\sigma}^2_{\tilde{\lambda}_j} = 0.01^2$ for $j=c,\ d,\ nd$. For each variance parameter in $\boldsymbol{\sigma}_\ell^2$, we adopt an inverse-gamma prior given by $\mathcal{IG}\left(\nu_\ell,\ S_\ell\right)$ such that $l=v,\ f$. In particular, we set $S_v = 0.2^2(\nu_v - 1)$ and $S_f = (\nu_f - 1)$. The last two expressions imply that $\mathbb{E}\sigma_v^2 = 0.2^2$ and $\mathbb{E}\sigma_f^2 = 1$, respectively. These imply that we allow, *a priori*, for larger variations in f_t relative to v_t, consistent with our goal to let the former absorb as much predictable content in $\mathbf{y}$ as possible. Such prior beliefs are, however, diffuse

and reflected in the calibration of the shape hyperparameters $\nu_v = \nu_f = \frac{T}{10}$.[13] Notably, the priors discussed above lead to closed-form expressions for all the conditional posterior distributions in (10). These expressions are standard and, for the sake of brevity, are reported in Section A1 of the Online Appendix. Also, our main results carry over to a number of sensitivity checks to the above-described priors. Such checks are discussed in greater detail in Section 4.5.

3.2 State Simulation

Let $\psi_t = (\psi_{1,t}, \cdots \psi_{N,t})'$, while f_t and v_t are defined as in (2) and (8), respectively. We can then group the latent states in our model as $\mathcal{Z} = \{\boldsymbol{\psi},\ \mathbf{f},\ \mathbf{v}\}$, where $\boldsymbol{\psi} = (\psi_1, \cdots, \psi_T)$, $\mathbf{f} = (f_1, \cdots, f_T)$, and $\mathbf{v} = (v_1, \cdots, v_T)$. State simulation thus entails sampling from the following three distributions:

$$\texttt{MCMC Steps:} \begin{cases} (5)\ p\left(\boldsymbol{\psi}|\mathbf{y}, \mathcal{Z}_{-\boldsymbol{\psi}}, \boldsymbol{\theta}\right), \\ (6)\ p\left(\mathbf{f}|\mathbf{y}, \mathcal{Z}_{-\mathbf{f}}, \boldsymbol{\theta}\right), \\ (7)\ p\left(\mathbf{v}|\mathbf{y}, \mathcal{Z}_{-\mathbf{v}}, \boldsymbol{\theta}\right). \end{cases} \tag{11}$$

Step 5 entails sequentially drawing from an inverse-gamma distribution to obtain NT draws for $\psi_{i,t}$. This result follows directly from the inverse-gamma prior in (9). Draws for $\mathbf{f}$ are obtained using the precision sampling methods by Chan and Jeliazkov (2009). The advantage of following this route is that precision methods typically reduce computational complexity and expedite state simulation (McCausland et al., 2011). For brevity, we relegate the expressions for the conditional posterior distributions in Steps 5 and 6 to Section A1 of the Online Appendix. We now discuss how to generate draws for $\mathbf{v}$ from the distribution in Step 7, which is a key feature of our estimation strategy.[14]

[13] See, e.g., Kroese and Chan (2014), chapter 11, for details on the parametrization of the inverse-gamma distribution which we adopt.

[14] The initial conditions for f_t and v_t are treated as additional parameters to which we assign a diffuse Gaussian prior centered at zero. We augment our MCMC algorithm to sample such parameters accordingly. The idiosyncratic volatility state ($\psi_{i,t}$) is assumed to be serially uncorrelated and therefore does not require an initial condition.

Joint Sampling of Common and Sector-Specific Uncertainty

Note first that, using the factorization of the error term in (3), the expressions (1), (6), and (8) can be described using the following stack representation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{L}_{\Psi}\mathbf{L}_{h}\mathbf{e}, \tag{12}$$

$$\mathbf{h} = \boldsymbol{\Lambda}\mathbf{v}, \tag{13}$$

$$\mathbf{L}_{\Phi_v}\mathbf{v} = \widetilde{\mathbf{v}}_0 + \boldsymbol{\eta}_{\mathbf{v}}, \tag{14}$$

where:

$$\mathbf{X} = \begin{bmatrix} I_N \otimes z_1' & I_N \otimes f_1' \\ \vdots & \vdots \\ I_N \otimes z_T' & I_N \otimes f_T' \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_z \\ \boldsymbol{\beta}_f \end{bmatrix}, \quad \boldsymbol{\Lambda} = I_T \otimes \begin{bmatrix} \Lambda^c & \Lambda^s \end{bmatrix},$$

$$\mathbf{L}_h = \mathtt{diag}\left(\Sigma_1^{\frac{1}{2}}, \cdots, \Sigma_T^{\frac{1}{2}}\right), \quad \mathbf{L}_{\Psi} = \mathtt{diag}\left(\Psi_1^{\frac{1}{2}}, \cdots, \Psi_T^{\frac{1}{2}}\right), \quad \mathbf{L}_{\Phi_v} = \begin{bmatrix} I_3 & \mathbf{0} & \cdots & \mathbf{0} \\ -\Phi_v & I_3 & & \\ \mathbf{0} & -\Phi_v & \ddots & \vdots \\ \vdots & & \ddots & \\ \mathbf{0} & & \cdots -\Phi_v & I_3 \end{bmatrix}.$$

The vector $\widetilde{\mathbf{v}}_0 = \iota_0 \otimes (\Phi_v v_0)$ collects the initial conditions for the common and sectoral uncertainty measures, where ι_0 denotes a $T \times 1$ vector given by $\iota_0 = (1,\ 0, \cdots, 0)'$.

To estimate $\mathbf{v}$, we combine the auxiliary mixture sampler approach of Omori et al. (2007) with precision sampling techniques. To apply the former, we first recast (12) as $(\mathbf{L}_{\Psi})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{L}_h\mathbf{e}$, or, more compactly, $\widetilde{\mathbf{y}} = \mathbf{L}_h\mathbf{e}$. Squaring and subsequently taking natural logarithms of each element in both sides of the previous expression yields:

$$\widetilde{\mathbf{y}}^* = \mathbf{h} + \widetilde{\mathbf{e}}^*, \tag{15}$$

where $\widetilde{\mathbf{y}}^* = (\log(\widetilde{y}_1^2), \cdots, \log(\widetilde{y}_T^2))$ and $\widetilde{\mathbf{e}}^* = (\log(e_1^2), \cdots, \log(e_T^2))$.

Next, plugging (13) into the right-hand side of (15) and appending (14) returns:

$$\widetilde{\mathbf{y}}^* = \mathbf{\Lambda}\mathbf{v} + \widetilde{\mathbf{e}}^*, \tag{16}$$

$$\mathbf{L}_{\Phi_v}\mathbf{v} = \widetilde{\mathbf{v}}_0 + \boldsymbol{\eta}_{\mathbf{v}}, \tag{17}$$

which denotes a linear state space form in $\mathbf{v}$. Nevertheless, the squaring and log-transformation applied to the error vector $\mathbf{e}$ implies that each entry in $\widetilde{\mathbf{e}}^*$ now follows a log chi-square distribution with one degree of freedom. To bring (16)–(17) back to a Gaussian (and hence more tractable) form, Omori et al. (2007) suggest approximating the distribution of $\widetilde{\mathbf{e}}^*$ as a ten-component weighted sum (or finite mixture) of Gaussian densities:[15]

$$\widetilde{\mathbf{e}}^* \sim p_1\mathcal{N}\left(\boldsymbol{\alpha}_1,\ \mathbf{\Sigma}_1\right) + \cdots + p_{10}\mathcal{N}\left(\boldsymbol{\alpha}_{10},\ \mathbf{\Sigma}_{10}\right). \tag{18}$$

Notably, the values for $\boldsymbol{\alpha}_k$, $\mathbf{\Sigma}_k$, and p_k for $k = 1, \cdots, 10$ are predetermined and given in Table 1 of Omori et al. (2007). Therefore, conditional on a particular density in (18), the state space in (16)–(17) can be recast in (conditionally) Gaussian form as:

$$\widetilde{\mathbf{y}}^* = \mathbf{\Lambda}\mathbf{v} + \boldsymbol{\alpha}_k + \widetilde{\mathbf{e}}_k^*, \tag{19}$$

$$\mathbf{L}_{\Phi_v}\mathbf{v} = \widetilde{\mathbf{v}}_0 + \boldsymbol{\eta}_{\mathbf{v}}, \tag{20}$$

$$\begin{bmatrix} \widetilde{\mathbf{e}}_k^* \\ \boldsymbol{\eta} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathbf{v}} \end{bmatrix}\right), \tag{21}$$

where $\mathbf{\Sigma}_{\mathbf{v}} = I_T \otimes \texttt{diag}(\sigma^2_{v_c},\ \sigma^2_{v_d},\ \sigma^2_{v_{nd}})$. Given the parameterization above, our posterior sampler needs to be augmented to sample a discrete state variable $k_t \in \{1, \cdots, 10\}$ for $t = 1, ..., T$ that serves as the mixture component indicator in (18) and thus determines the values for $\boldsymbol{\alpha}_k$ and $\mathbf{\Sigma}_k$ in (19)–(21). More precisely, estimation of $\mathbf{v}$ requires sequentially

[15] Their approach extends the seven-component auxiliary mixture sampling from Kim et al. (1998).

sampling from the following two full conditional posterior distributions:[16]

$$\texttt{Step 1}\ p(\mathbf{k}|\widetilde{\mathbf{y}}^*, \mathcal{Z}, \boldsymbol{\theta})\ s.t.\ \mathbf{k} = (k_1, \cdots, k_T)\,,$$

$$\texttt{Step 2}\ p(\mathbf{v}|\widetilde{\mathbf{y}}^*, \mathcal{Z}_{-\mathbf{v}}, \mathbf{k}, \boldsymbol{\theta}).$$

Posterior draws for $\mathbf{k}$ are obtained by independently sampling each element in $\mathbf{k}$ from a multinomial distribution via the inverse transform method as discussed in Kroese and Chan (2014).[17]

Conditional on $\mathbf{k}$ (and the remaining states and parameters), we then combine likelihood- and prior-based information from (19) and (20), respectively, using Bayes rule to obtain the following closed-form expression for $p(\mathbf{v}|\widetilde{\mathbf{y}}^*, \mathcal{Z}_{-\mathbf{v}}, \boldsymbol{\theta})$:

$$\mathbf{v}|\widetilde{\mathbf{y}}^*, \mathcal{Z}_{-\mathbf{v}}, \boldsymbol{\theta} \sim \mathcal{N}\left(\overline{\mathbf{d}}_v,\ \overline{\mathbf{D}}_v\right),\ \text{where}\ \begin{cases} \overline{\mathbf{d}}_v = \overline{\mathbf{D}}_v\left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_k^{-1}(\widetilde{\mathbf{y}}^* - \boldsymbol{\alpha}_k) + \mathbf{L}_{\Phi_v}'\boldsymbol{\Sigma}_{\mathbf{v}}^{-1}\widetilde{\mathbf{v}}_0\right), \\ \overline{\mathbf{D}}_v = \left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Lambda} + \mathbf{L}_{\Phi_v}'\boldsymbol{\Sigma}_{\mathbf{v}}^{-1}\mathbf{L}_{\Phi_v}\right)^{-1}. \end{cases} \tag{22}$$

Draws from $\mathcal{N}\left(\overline{\mathbf{d}}_v,\ \overline{\mathbf{D}}_v\right)$ are obtained using the precision sampling techniques as in Chan and Jeliazkov (2009) to construct $\overline{\mathbf{d}}_v$ and $\overline{\mathbf{D}}_v$. Once draws for $\mathbf{v}$ (and for the loadings in $\boldsymbol{\Lambda}_v$) are generated, we plug them into (13) to back out the log-volatilities for the appropriate levels of aggregation.

4 Empirical Results

In this section we present the empirical results obtained from the estimation of the model discussed in Sections 2 and 3. First, we analyze the evolution of common and sectoral uncertainty and discuss how they correlate with other macroeconomic and uncertainty indicators. Second, we conduct a VAR exercise akin to Jurado et al. (2015) to examine the role of uncertainty (both common and sectoral) as drivers of economic activity. In the last

[16] The ordering of Steps 1 and 2 shown above is consistent with the discussion in Del Negro and Primiceri (2015) on estimating stochastic volatility models with auxiliary mixture methods.

[17] Further details on how to sample from $p(\mathbf{k}|\widetilde{\mathbf{y}}^*, \mathcal{Z}, \boldsymbol{\theta})$ are provided in Section A1 of the Online Appendix.

part of this section we discuss some robustness checks.

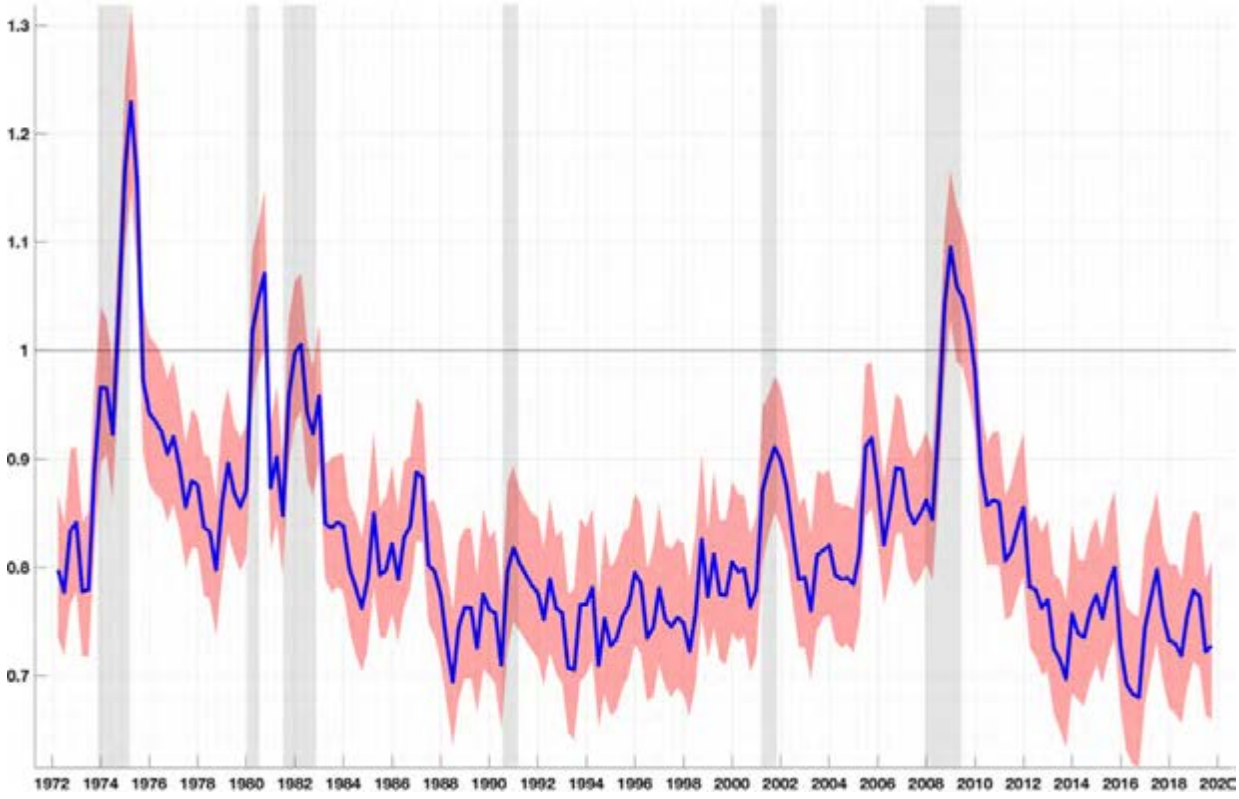
To avoid confusion, it is worth reiterating that we obtain measures of uncertainty from a large cross-section of standardized growth rates for 185 industries allocated evenly across the durable and nondurables sectors in the U.S. economy from 1972Q1 to 2019Q4. Therefore, our common uncertainty measure can be perceived as metric for uncertainty in the goods sector as a whole, while sectoral uncertainty denotes uncertainty that is specific to the durables and nondurables sectors.

4.1 Common Uncertainty

Figure 1 shows the evolution of our estimated common uncertainty factor. In keeping with conventional wisdom (Bloom, 2014), common uncertainty peaks during recessions. In particular, according to our model, (pre-pandemic) uncertainty peaked in the mid-1970s during the first energy crisis of that decade. The second highest peak is associated with the Great Recession of 2008–09. Other distinctive peaks can be seen in the late 1970s and early 1980s, periods respectively characterized by the second energy crisis of the 1970s and restrictive monetary policy—the latter in light of the decade-long period of double-digit inflation that characterized the 1970s. We also document that common uncertainty has become, on average, substantially lower since the mid-1980s. This reinforces the idea put forward by various authors (e.g., Stock and Watson (2002), Sims and Zha (2006), and Galí and Gambetti (2009)) that the Great Moderation is a period predominantly characterized by low overall volatility.

Next, we contrast our estimated measure of common uncertainty against other common measures of broader uncertainty. For this exercise, we adopt the macroeconomic uncertainty indicator of Jurado et al. (2015) (JLN Macro, hereafter), the news data-based economic policy uncertainty index of Baker et al. (2016), the real activity uncertainty index of Ludvigson et al. (2021) (LMN Real, hereafter), and the VIX, used, e.g., by Bloom (2009).

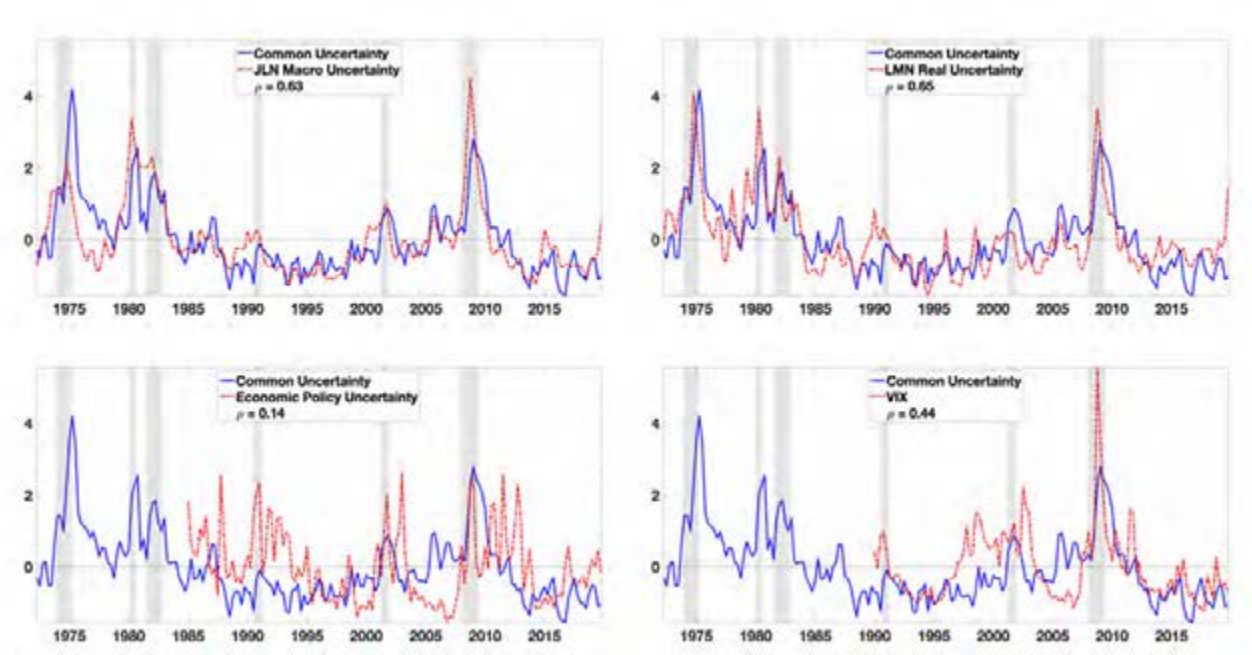
Figure 1: Common Uncertainty



Notes: Common uncertainty is extracted from disaggregated IP data for 185 industries using the modeling approach discussed in Sections 2 and 3. The solid line denotes the posterior median for $\exp(v_{c,t}/2)$. The shaded region denotes the 67% equal-tailed posterior credible interval. The vertical shaded bars correspond to the NBER recession dates.

Figure 2 documents these comparisons. While we find our common uncertainty factor (positively) correlates with all the above-mentioned measures, correlation is stronger with the JLN Macro and LMN Real uncertainty indices. This is perhaps not too surprising, given that we extract uncertainty from real activity data, which have different characteristics from financial and news data. In this sense, taking our common uncertainty measure as an indicator of uncertainty in the goods sector as a whole, Figure 2 then suggests that broader uncertainty in such a sector has a component that moves in line with macroeconomic uncertainty.

Figure 2: Common and Other Measures of Uncertainty

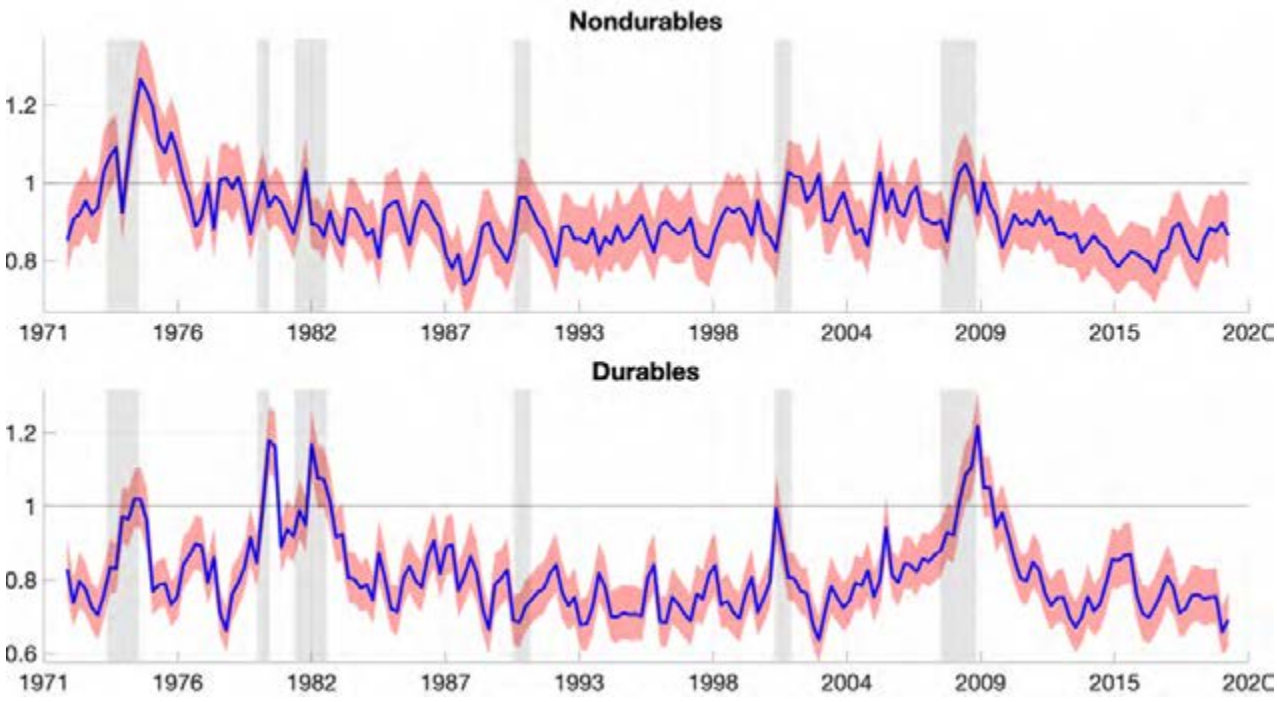


Notes: JLN Macro denotes uncertainty as estimated by Jurado et al. (2015). LMN Real denotes uncertainty as in Ludvigson et al. (2021). Economic Policy denotes uncertainty as as in Baker et al. (2016). VIX is the Financial Volatility Index produced by the Chicago Board Options Exchange. Measures are normalized for ease of comparison. ρ stands for the unconditional correlation coefficient between two measures of uncertainty. The vertical shaded bars correspond to NBER recession dates.

4.2 Sectoral Uncertainty

We now turn to our results for sectoral uncertainty. Figure 3 shows our measures of uncertainty that are specific to the durable and nondurable goods sectors. The two series exhibit different dynamics, which is manifested in the weak correlation (0.20) between them. Like common uncertainty, nondurable uncertainty displays its highest peak during the mid-1970s recession, suggesting that the first oil crisis of the 1970s also had repercussions at the sector-specific level in addition to its economy-wide effects. Interestingly, nondurable uncertainty is relatively stable without any discernible peaks over the remainder of our sample. This is consistent with nondurable consumption often being characterized by its excess smoothness (see, e.g., Luengo-Prado (2006)). Durable uncertainty, on the other hand, peaks during most recessions, the exception being the early-1990s recession.

Figure 3: Sectoral Uncertainty



Notes: Nondurable and durable uncertainty are obtained from two subsets of disaggregated IP data consisting of 80 and 105 industries in the nondurable and durable goods sectors, respectively. The series are estimated using the methodology discussed in Sections 2 and 3. Solid lines denote the posterior median for $\exp(v_{nd,t}/2)$ (nondurables) and $\exp(v_{d,t}/2)$ (durables). Shaded regions around solid lines denote the 67% equal-tailed posterior credible interval. Vertical shaded bars correspond to NBER recession dates.

Also, unlike nondurable and common uncertainty, durable uncertainty exhibits its highest peak during the Great Recession of 2008–09, reflecting the disproportionate impact this recession had on durable-spending decisions. Such a result also echoes panel-based evidence, as documented by Berger and Vavra (2015), on the sluggish readjustment of durable expenditure to economic stimulus during the Great Recession, hence potentially exacerbating nondurable uncertainty.

Table 1 reports the correlation between our measures of uncertainty, both common and sectoral, and other macroeconomic indicators of real activity. A few results stand out. First, our common uncertainty measure is clearly countercyclical, as indicated by the negative correlation with the growth rate of variables such as real GDP, IP, and employment and the positive correlation with the NBER recession indicator. Furthermore, as pointed

out above, common uncertainty has a high positive correlation with the JLN Macro and LMN Real uncertainty measures, while somewhat less (positively) correlated with financial uncertainty measures such as the financial uncertainty index by Ludvigson et al. (2021) and the VIX.[18]

Table 1: Correlation between Our Proposed Measures of Uncertainty and Selected Economic Indicators.

Variable	**Common**	**Nondurables**	**Durables**
Real GDP	-0.23	-0.12	-0.19
Industrial Production	-0.30	-0.14	-0.26
Investment	-0.24	-0.12	-0.19
Employment	-0.37	-0.15	-0.38
Real PCE	0.33	0.29	0.20
CPI-Inflation	0.23	0.23	0.12
NBER Recession Indicator	0.55	0.28	0.51
JLN Macroeconomic Uncertainty	0.63	0.32	0.60
LMN Real Uncertainty	0.65	0.39	0.53
LMN Financial Uncertainty	0.36	0.24	0.29
Economic Policy Uncertainty	0.14	0.09	0.11
Monetary Policy Uncertainty	0.07	0.09	0.02
Fiscal Policy Uncertainty	0.13	0.08	0.10
VIX	0.44	0.28	0.35

Notes: The first five series above are measured as annualized growth rates.

Among the sectoral uncertainty measures, the one associated with durables is more tightly correlated with real activity indicators than nondurable uncertainty. This evidence is in line with the micro-data analysis by Kehrig (2015), who finds a more pronounced countercyclical pattern of productivity dispersion in durable goods industries than in nondurable goods industries. Also, uncertainty (common and sectoral) is positively correlated with consumer price index (CPI) inflation and real personal consumption expenditure (PCE) growth. The former correlation is consistent with firms changing their prices upward in presence of uncertainty, a mechanism known as "upward pricing bias" (Fernández-Villaverde et al., 2015; Born and Pfeifer, 2021). The positive correlation with personal consumption is consistent with closed economy models of the business cycle in which the "wait-and-see" channel—combined with the time-to-build assumption that implies a lagged impact

[18] Our framework also allows us to construct a measure of uncertainty by taking the average of volatilities of all industries, which would be more akin to the JLN uncertainty measure. As a matter of fact, this average measure is slightly more countercyclical and more correlated with the JLN measure.

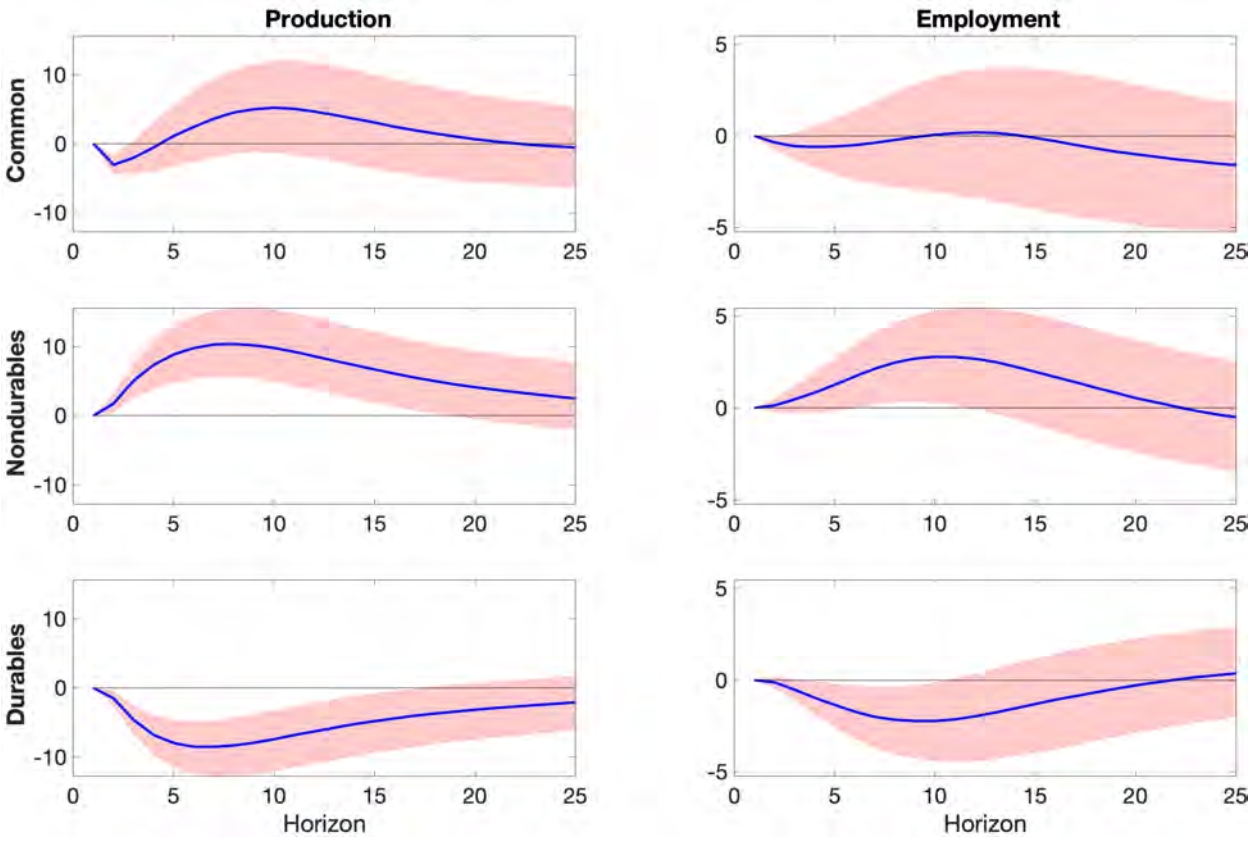
of investment on productive capital—implies an increase in households' spending (see, e.g., Bloom et al. (2018)).

4.3 Sectoral Uncertainty and Macroeconomic Dynamics

The empirical literature has predominantly found contractionary effects after surprise increases in various (economy-wide) uncertainty measures (see, e.g., Bloom (2009), Bachmann et al. (2013), Jurado et al. (2015), and Caldara et al. (2016)). But there is no evidence if these results hold when the separation between common and sectoral uncertainty is considered. To address this, we revisit the VAR model in Jurado et al. (2015) to measure the real macroeconomic effects of our sectoral uncertainty measures. Like the authors, our vector of macroeconomic indicators is given by $\widetilde{Y}_M$=[IP, EMP, CON, PCE, NOR, WAGE, HOURS, FFR, SP500, M2]$'$, where we use IP as short notation for real (aggregate) industrial production, EMP for employment, CON for real consumption, PCE for PCE deflator, NOR for real new orders, WAGE for real wage, HOURS for hours, FFR for the federal funds rate, SP500 for the Standard & Poor's stock market 500 index, and M2 for money supply. The variables IP, EMP, CON, PCE, NOR, WAGE, and SP500 enter the model in log-levels, while M2 is in growth rates.

Our uncertainty measures are appended to $\widetilde{Y}_M$ and we then run three different VARs with the following vectors of endogenous variables: $\widetilde{Y}_C = [\widetilde{Y}_M', \ \exp(v_{c,t}/2)]'$, $\widetilde{Y}_{ND} = [\widetilde{Y}_M', \ \exp(v_{c,t}/2), \ \exp(v_{nd,t}/2)]'$, and $\widetilde{Y}_D = [\widetilde{Y}_M', \ \exp(v_{c,t}/2), \ \exp(v_{d,t}/2)]'$. In the first VAR, we place our proxy for common uncertainty $\exp(v_{c,t}/2)$ last, whereas for the other two VARs sectoral uncertainty measures $\exp(v_{nd,t}/2)$ and $\exp(v_{d,t}/2)$ are positioned last in the vector, right after the common uncertainty measure. Similar to Jurado et al. (2015), we orthogonalize the VAR residuals via a Cholesky decomposition of their covariance matrix. Hence, we focus on impulse responses to unexpected changes in our uncertainty measures—for brevity, we label such changes as "shocks." For future reference, the results associated with common uncertainty come from the VAR model with $\widetilde{Y}_C$, while those associated with nondurables and durables uncertainty are based on the VAR models with $\widetilde{Y}_{ND}$ and $\widetilde{Y}_D$, respectively. All VARs are estimated using Bayesian methods with the Minnesota prior of Doan et al. (1984).

Figure 4: Responses of Real Activity to Uncertainty Shocks. Full Sample: 1972Q1–2019Q4.



Notes: Impulse responses denote a one standard deviation shock to the common, nondurables, and durables uncertainty measures. Solid lines denote posterior medians, and the shaded areas represent the 67% equal-tailed posterior credible interval.

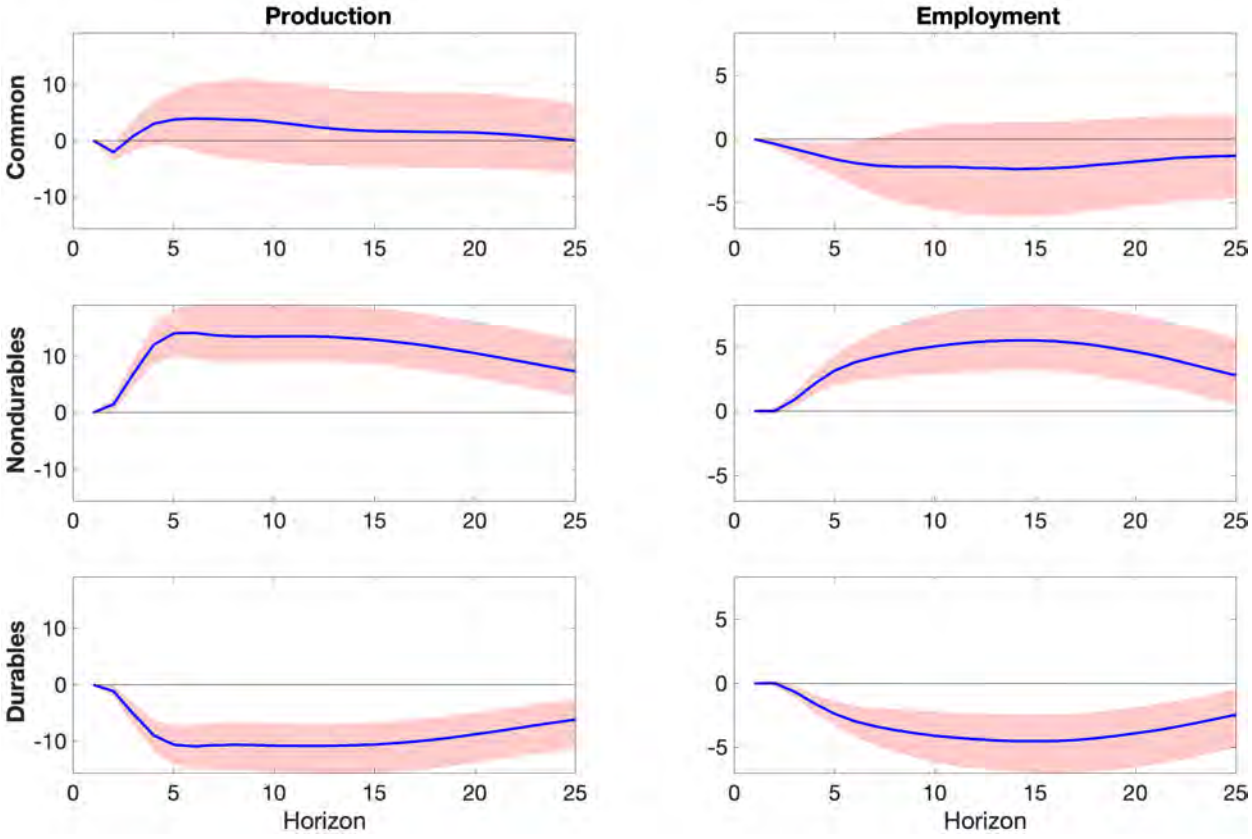
Figure 4 shows the responses of production and employment to a one standard deviation surprise movement in the common and sectoral uncertainty measures. Evidently, the definition of uncertainty matters. Shocks to common uncertainty generate the well-known "drop-rebound-overshoot" dynamic response of real activity after an uncertainty shock (Bloom, 2009). A different reaction is documented after an unexpected hike in durable goods uncertainty. In this case, the response of real activity is much more gradual, reaches a trough after one year, and follows a smoother (inverse) hump-shaped path. This negative response is interpretable in terms of the "wait-and-see" behavior by firms after an uncertainty shock. Such behavior is justified by non-convex adjustment costs that make it optimal to postpone investment until the "smoke clears" (Blanchard, 2009). In sharp contrast, the response of production and employment to a nondurable uncertainty shock

is positive. One plausible explanation for this result is that industries in the nondurable goods sector are typically not characterized by non-convex adjustment costs. As discussed by Bloom (2009) and Ludvigson et al. (2021), "growth options" theories of uncertainty predict that an increase in uncertainty might lead firms to invest and hire, given that such an increase in risk—coupled with the possibility of revert back investment decisions at low costs—increases expected profits due to the shape of the profit function. Another relevant difference in the transmission of uncertainty shocks in these two sectors may be related to price rigidity. In presence of higher price rigidity in the durables good sector, the real effects of uncertainty shocks might be more recessionary. This is admittedly a conjecture because, to our knowledge, there is not yet solid evidence on the relative importance of price stickiness in the durable vs. nondurable goods sector.

4.4 Sectoral Uncertainty During the Great Moderation

Recent research has documented that the onset of the Great Moderation was accompanied by an increasing role of first-moment sectoral shocks as drivers of the business cycle (Foerster et al., 2011; Atalay, 2017; Garin et al., 2018). It is thus tempting to complete the picture and verify if second-moment sectoral shocks have also become more relevant since the mid-1980s. To this end, Figure 5 documents the impulse responses of production and employment to common, durable, and nondurable uncertainty shocks for a sample running from 1984Q1 to 2019Q4. The responses to exogenous changes in common uncertainty are broadly similar to those documented for the full sample in Figure 4. A quick visual inspection, however, shows that the peak and trough responses for the nondurables and durables cases, respectively, are slightly more pronounced when fitting the VAR to the Great Moderation sample. Table 2 confirms such results by reporting the forecast error variance decomposition results based on a 24-quarter horizon. Our estimates indicate a considerably larger contribution of both sectoral uncertainty shocks to IP and employment when the sample is truncated to the Great Moderation period.

Figure 5: Responses of Real Activity to Uncertainty Shocks. Sample: 1984Q1–2019Q4.



Notes: Impulse responses denote a one standard deviation shock to the common, nondurables, and durables uncertainty measures. Solid lines denote posterior medians, and the shaded areas represent the 67% equal-tailed posterior credible interval.

Table 2: Forecast Error Variance Decomposition (FEVD) for Uncertainty Shocks to Industrial Production and Employment.

	72Q2–19Q4	**84Q1–19Q4**
Ind. Production: % explained by		
Common	1.4%	1.8%
Nondurables	4.0%	11.5%
Durables	4.1%	11.9%
Employment: % explained by		
Common	1.1%	2.3%
Nondurables	1.4%	6.8%
Durables	1.5%	7.4%

Notes: FEVD results are constructed over a 24-quarter horizon.

4.5 Robustness Checks

We assess the robustness of our proposed framework along several dimensions. In particular, we generate measures of common and sectoral uncertainty under various alternative specifications for the measurement equation in (1). These include (i) four different structures for X_t; and (ii) allowing for normally (in addition to t-) distributed residuals (u_t). Regarding (i), amongst other structures, we entertain the possibility that the latent factors (i.e., f_t in (2)) exhibit conditional stochastic volatility as in, e.g., Chib et al. (2006) and Mumtaz and Musso (2019).[19]

We also examine whether our impulse-response results would hold under different VAR specifications. To this end, we recompute impulse responses from VAR models of seven macroeconomic variables (similar to Bloom (2009)) and a smaller scale VAR with five macroeconomic variables (similar to Leduc and Liu (2016) and Alessandri and Mumtaz (2019)). In the interest of space, all robustness-related results are reported in Section A2 of the Online Appendix. We do stress, however, that all our main findings carry over to the above-mentioned checks.

5 Conclusion

This paper investigates sectoral uncertainty. It does so by proposing a novel empirical framework designed to conduct the joint estimation of common uncertainty (across different sectors) and sectoral uncertainty (which is specific to selected series belonging to sectors we focus on) in a data-rich environment. We apply such a framework to 185 industrial production series for the U.S. economy. We find that durable goods uncertainty displays a large peak in correspondence with the great recession, while nondurable uncertainty was more prominent during the 1973–75 recession. Durables uncertainty is also documented to be more tightly correlated with proxies for real activity than nondurables uncertainty. Working with vector autoregressions that account for common and sectoral uncertainty, we find shocks to the former to generate a quick drop-rebound-overshoot response of real

[19] A model comparison exercise based on the deviance information criteria further supported the baseline specification adopted in this study to be the one with the best fit amongst all competing variants of X_t. The results for this exercise are presented in Section A2 of the Online Appendix.

activity. In contrast, shocks to durable goods uncertainty are found to generate a persistent downturn, while shocks to nondurables are expansionary. These suggest that findings in the literature that point to aggregate uncertainty as a (recessionary) driver of the business cycle might actually conflate the heterogeneous effects of common, durable, and nondurable uncertainty.

Overall, our results point to the need of investigating uncertainty at a sectoral level, both from an empirical and a theoretical standpoint. From an empirical standpoint, our contribution offers an empirical framework that encompasses frameworks successfully applied to the investigation of the role of uncertainty shocks (Carriero et al., 2016, 2018)). From a theoretical standpoint, our findings call for theoretical models featuring durable and nondurable goods and characterized by different second-moment shocks with a heterogeneous impact on the business cycle.

References

Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "Microeconomic origins of macroeconomic tail risks," *American Economic Review*, 2017, *107* (1), 54–108.

_ , Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The network origins of aggregate fluctuations," *Econometrica*, 2012, *80* (5), 1977–2016.

Alessandri, Piergiorgio and Haroon Mumtaz, "Financial regimes and uncertainty shocks," *Journal of Monetary Economics*, 2019, *101*, 31–46.

Alessi, Lucia, Matteo Barigozzi, and Marco Capasso, "Improved penalization for determining the number of factors in approximate factor models," *Statistics & Probability Letters*, 2010, *80* (23-24), 1806–1813.

Angelini, Giovanni, Emanuele Bacchiocchi, Giovanni Caggiano, and Luca Fanelli, "Uncertainty across volatility regimes," *Journal of Applied Econometrics*, 2019, *34(3)*, 437–455.

Antolín-Díaz, Juan, Thomas Drechsel, and Ivan Petrella, "Advances in nowcasting economic activity: Secular trends, large shocks and new data," Technical Report, Working Paper 2020.

Atalay, Enghin, "How important are sectoral shocks?," *American Economic Journal: Macroeconomics*, 2017, *9* (4), 254–80.

Bachmann, Rüdiger, Steffen Elstner, and Eric R Sims, "Uncertainty and economic activity: Evidence from business survey data," *American Economic Journal: Macroeconomics*, 2013, *5* (2), 217–49.

Bai, Jushan and Serena Ng, "Determining the number of factors in approximate factor models," *Econometrica*, 2002, *70* (1), 191–221.

__ and __, "Extremum estimation when the predictors are estimated from large panels," *Annals of Economics and Finance*, 2008, *9* (2), 201–222.

Baker, Scott R, Nicholas Bloom, and Steven J Davis, "Measuring economic policy uncertainty," *The Quarterly Journal of Economics*, 2016, *131* (4), 1593–1636.

Baqaee, David Rezza and Emmanuel Farhi, "The macroeconomic impact of microeconomic shocks: Beyond Hulten's Theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

Basu, Susanto and Brent Bundick, "Uncertainty shocks in a model of effective demand," *Econometrica*, 2017, *85* (3), 937–958.

Berger, David and Joseph Vavra, "Consumption dynamics during recessions," *Econometrica*, 2015, *83* (1), 101–154.

Bianchi, Francesco, Howard Kung, and Mikhail Tirskikh, "The origins and effects of macroeconomic uncertainty," 2021, *available at https://sites.google.com/view/francescobianchi/home.*

Blanchard, O., "(Nearly) nothing to fear but fear itself," 2009, *The Economist, January 29.*

Bloom, Nicholas, "The impact of uncertainty shocks," *Econometrica*, 2009, *77* (3), 623–685.

__, "Fluctuations in uncertainty," *Journal of Economic Perspectives*, 2014, *28* (2), 153–76.

__, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J Terry, "Really uncertain business cycles," *Econometrica*, 2018, *86* (3), 1031–1065.

Born, Benjamin and Johannes Pfeifer, "Uncertainty-driven business cycles: Assessing the markup channel," *Quantitative Economics*, 2021, *12(2)*, 587–623.

Caggiano, G., E. Castelnuovo, and G. Nodari, "Uncertainty and monetary policy in good and bad times: A replication of the VAR investigation by Bloom (2009)," *Journal of Applied Econometrics*, 2022, *37*, 210–217.

Caggiano, Giovanni, Efrem Castelnuovo, and Nicolas Groshenny, "Uncertainty shocks and unemployment dynamics in US recessions," *Journal of Monetary Economics*, 2014, *67*, 78–92.

Caldara, Dario, Cristina Fuentes-Albero, Simon Gilchrist, and Egon Zakrajšek, "The macroeconomic impact of financial and uncertainty shocks," *European Economic Review*, 2016, *88*, 185–207.

Carriero, Andrea, Todd E Clark, and Massimiliano Marcellino, "Common drifting volatility in large Bayesian VARs," *Journal of Business & Economic Statistics*, 2016, *34* (3), 375–390.

__, __, and __, "Measuring uncertainty and its impact on the economy," *Review of Economics and Statistics*, 2017, (0).

__, __, and __, "Measuring uncertainty and its impact on the economy," *Review of Economics and Statistics*, 2018, *100* (5), 799–815.

__, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens, "Measuring uncertainty and its effects in the COVID-19 era," 2020, *Queen Mary - University of London, Federal Reserve Bank of*

Cleveland, Bocconi University, and Deutsche Bundesbank, mimeo.

Cascaldi-Garcia, Danilo, Cisil Sarisoy, Juan M. Londono, Bo Sun, Deepa Datta, Thiago Ferreira, Olesya Grishchenko, Mohammad R. Jahan-Parvar, Francesca Loria, Sai Ma, Marius Rodriguez, Ilknur Zer, and John Rogers, "What is certain about uncertainty?," *Journal of Economic Literature*, 2021.

Chan, J.C.C. and I. Jeliazkov, "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 2009, *1(1-2)*, 101–120.

Chib, Siddhartha, Federico Nardari, and Neil Shephard, "Analysis of high dimensional multivariate stochastic volatility models," *Journal of Econometrics*, 2006, *134* (2), 341–371.

Choi, Sangyup and Prakash Loungani, "Uncertainty and unemployment: The effects of aggregate and sectoral channels," *Journal of Macroeconomics*, 2015, *46*, 344–358.

Clark, Todd E., Michael W. McCracken, and Elmar Mertens, "Modeling time-varying uncertainty of multiple-horizon forecast errors," *Review of Economics and Statistics*, 2020, *102* (1), 17–33.

Coibion, Olivier, Dimitris Georgarakos, Yuriy Gorodnichenko, Geoff Kenny, and Michael Weber, "The effect of macroeconomic uncertainty on household spending," 2021, *National Bureau of Economic Research Working Paper No. 28625.*

Conley, Timothy G and Bill Dupor, "A spatial analysis of sectoral complementarity," *Journal of Political Economy*, 2003, *111* (2), 311–352.

Doan, Thomas, Robert Litterman, and Christopher Sims, "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 1984, *3* (1), 1–100.

Fernández-Villaverde, Jesus and Pablo Guerron-Quintana, "Uncertainty shocks in real business cycle models," *Review of Economic Dynamics*, 2020, *37*, 118–166.

_ , _ , Juan F. Rubio-Ramírez, and M. Uribe, "Risk matters: The real effects of volatility shocks," *American Economic Review*, 2011, *101(6)*, 2530–2561.

_ , _ , Keith Kuester, and Juan F. Rubio-Ramírez, "Fiscal volatility shocks and economic activity," *American Economic Review*, 2015, *105(11)*, 3352–3384.

Foerster, Andrew T, Pierre-Daniel G Sarte, and Mark W Watson, "Sectoral versus aggregate shocks: A structural factor analysis of industrial production," *Journal of Political Economy*, 2011, *119* (1), 1–38.

Galí, Jordi and Luca Gambetti, "On the sources of the great moderation," *American Economic Journal: Macroeconomics*, 2009, *1* (1), 26–57.

Garin, Julio, Michael J Pries, and Eric R Sims, "The relative importance of aggregate and sectoral shocks and the changing nature of economic fluctuations," *American Economic Journal: Macroeconomics*, 2018, *10* (1), 119–48.

Gorodnichenko, Yuriy and Serena Ng, "Level and volatility factors in macroeconomic data," *Journal*

of Monetary Economics, 2017, *91*, 52–68.

Hoke, Sinem Hacioglu and Kerem Tuzcuoglu, "Interpreting the latent dynamic factors by threshold FAVAR model," *Staff Working Paper Series 622, Bank of England*, 2016.

Horvath, Michael, "Cyclicality and sectoral linkages: Aggregate fluctuations from independent sectoral shocks," *Review of Economic Dynamics*, 1998, *1* (4), 781–808.

_ , "Sectoral shocks and aggregate fluctuations," *Journal of Monetary Economics*, 2000, *45* (1), 69–106.

Jacquier, Eric, Nicholas G Polson, and Peter E Rossi, "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 2004, *122* (1), 185–212.

Jo, Soojin and Rodrigo Sekkel, "Macroeconomic uncertainty through the lens of professional forecasters," *Journal of Business & Economic Statistics*, 2019, *37* (3), 436–446.

Jurado, Kyle, Sydney C Ludvigson, and Serena Ng, "Measuring uncertainty," *American Economic Review*, 2015, *105* (3), 1177–1216.

Kehoe, Patrick J, Virgiliu Midrigan, and Elena Pastorino, "Evolution of modern business cycle models: Accounting for the great recession," *Journal of Economic Perspectives*, 2018, *32* (3), 141–66.

Kehrig, Matthias, "The cyclical nature of the productivity distribution," *Earlier version: US Census Bureau Center for Economic Studies Paper No. CES-WP-11-15*, 2015.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib, "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 1998, *65(3)*, 361–393.

Kim, Yunmi and Chang-Jin Kim, "Dealing with endogeneity in a time-varying parameter model: Joint estimation and two-step estimation procedures," *The Econometrics Journal*, 2011, *14* (3), 487–497.

Koop, Gary, Dale J. Poirier, and Justin L. Tobias, "Bayesian econometric methods," 2007, *Cambridge University Press.*

Kose, M Ayhan, Christopher Otrok, and Charles H Whiteman, "International business cycles: World, region, and country-specific factors," *American Economic Review*, 2003, *93* (4), 1216–1239.

Kroese, Dirk P. and Joshua C.C. Chan, "Statistical modeling and computation," 2014, *Springer, New York, NY.*

Leduc, Sylvain and Zheng Liu, "Uncertainty shocks are aggregate demand shocks," *Journal of Monetary Economics*, 2016, *82(C)*, 20–35.

Lenza, Michele and Giorgio Primiceri, "How to estimate a VAR after March 2020," *Journal of Applied Econometrics*, 2021, *forthcoming.*

Ludvigson, Sydney C and Serena Ng, "Macro factors in bond risk premia," *The Review of Financial Studies*, 2009, *22* (12), 5027–5067.

Ludvigson, Sydney C., Sai Ma, and Serena Ng, "Uncertainty and business cycles: Exogenous impulse or endogenous response?," *American Economic Journal: Macroeconomics*, 2021, *13(4)*, 369–410.

Luengo-Prado, María José, "Durables, nondurables, down payments and consumption excesses," *Journal of Monetary Economics*, 2006, *53* (7), 1509–1539.

Ma, Xiaohan and Roberto Samaniego, "Deconstructing uncertainty," *European Economic Review*, 2019, *119*, 22–41.

Marcellino, Massimiliano, Todd Clark, and Andrea Carriero, "Nowcasting tail risk to economic activity at a weekly frequency," 2021, *CEPR Discussion Paper No. 16496.*

McCausland, William J., Shirley Millera, and Denis Pelletierc, "Simulation smoothing for state-space models: A computational efficiency analysis," *Computational Statistics and Data Analysis*, 2011, *55*, 199–212.

McCracken, Michael W. and Serena Ng, "FRED-MD: A monthly database for macroeconomic research," *Journal of Business and Economic Statistics*, 2016, *34(4)*, 574–589.

Miron, Jeffrey A and Stephen P Zeldes, "Production, sales, and the change in inventories: An identity that doesn't add up," *Journal of Monetary Economics*, 1989, *24* (1), 31–51.

Moench, Emanuel, Serena Ng, and Simon Potter, "Dynamic hierarchical factor models," *Review of Economics and Statistics*, 2013, *95* (5), 1811–1817.

Mumtaz, Haroon and Alberto Musso, "The evolving impact of global, region-specific, and country-specific uncertainty," *Journal of Business & Economic Statistics*, 2019, pp. 1–16.

Negro, Marco Del and Christopher Otrok, "99 Luftballons: Monetary policy and the house price boom across US states," *Journal of Monetary Economics*, 2007, *54* (7), 1962–1985.

__ **and Giorgio E Primiceri**, "Time varying structural vector autoregressions and monetary policy: A corrigendum," *The Review of Economic Studies*, 2015, *82* (4), 1342–1345.

Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima, "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 2007, *140(2)*, 425–449.

Pagan, Adrian, "Econometric issues in the analysis of regressions with generated regressors," *International Economic Review*, 1984, pp. 221–247.

Segal, Gill, "A tale of two volatilities: Sectoral uncertainty, growth, and asset prices," *Journal of Financial Economics*, 2019.

Sims, Christopher A. and Tao Zha, "Were there regime switches in U.S. monetary policy?," *American Economic Review*, 2006.

Stock, James H and Mark W Watson, "Has the business cycle changed and why?," *NBER Macroeconomics Annual*, 2002, *17*, 159–218.

Stock, James H. and Mark W. Watson, "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics," 2016, *in: J. B. Taylor and H. Uhlig (Eds.): Handbook of Macroeconomics, Vol. 2, North Holland*, 415–525.

Wynne, Mark A and Nathan S Balke, "Recessions and recoveries," *Federal Reserve Bank of Dallas, Economic Review, first quarter*, 1993, pp. 1–18.

Online Appendix to *Sectoral Uncertainty* *

Efrem Castelnuovo† Kerem Tuzcuoglu‡ Luis Uzeda§

Contents

A1 Estimation Details **1**

A1.1 Posterior Sampler 2

A1.1.1 State Simulation 2

A1.1.2 Parameter Sampling 8

A1.2 Computational Performance 12

A2 Robustness Checks **14**

A2.1 Uncertainty Measurement under Alternative Specifications 14

A2.2 A "Strictly" Gaussian Representation 20

A2.3 Alternative VAR Specifications 22

A3 Dataset Description **24**

*The views expressed are those of the authors and do not necessarily reflect the position of the Bank of Canada. All remaining errors are ours.

†University of Padova, CESifo, and CAMA. E-mail: efrem.castelnuovo@unipd.it

‡Bank of Canada. E-mail: ktuzcuoglu@bank-banque-canada.ca

§Bank of Canada and CAMA. E-mail: luzedagarcia@bank-banque-canada.ca

A1 Estimation Details

For convenience, below we reproduce the stacked representation of our model as presented in Section 3.2 of the main text:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{L}_{\Psi}\mathbf{L}_{h}\mathbf{e} \quad \mathbf{e} \sim \mathcal{N}\left(\mathbf{0},\ I_{NT}\right) \tag{A1}$$

$$\mathbf{h} = \boldsymbol{\Lambda}\mathbf{v}, \tag{A2}$$

$$\mathbf{L}_{\Phi_v}\mathbf{v} = \widetilde{\mathbf{v}}_0 + \boldsymbol{\eta}_{\mathbf{v}} \quad \mathbf{v} \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathbf{v}}\right), \tag{A3}$$

where:

$$\mathbf{X} = \begin{bmatrix} I_N \otimes z_1' & I_N \otimes f_1' \\ \vdots & \vdots \\ I_N \otimes z_T' & I_N \otimes f_T' \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_z \\ \boldsymbol{\beta}_f \end{bmatrix}, \quad \boldsymbol{\Lambda} = I_T \otimes \begin{bmatrix} \Lambda^c & \Lambda^s \end{bmatrix},$$

$$\mathbf{L}_h = \mathtt{diag}\left(\Sigma_1^{\frac{1}{2}}, \cdots, \Sigma_T^{\frac{1}{2}}\right), \quad \mathbf{L}_{\Psi} = \mathtt{diag}\left(\Psi_1^{\frac{1}{2}}, \cdots, \Psi_T^{\frac{1}{2}}\right), \quad \mathbf{L}_{\Phi_v} = \begin{bmatrix} I_3 & \mathbf{0} & \cdots & \mathbf{0} \\ -\Phi_v & I_3 & & \\ \mathbf{0} & -\Phi_v & \ddots & \vdots \\ \vdots & & \ddots & \\ \mathbf{0} & & \cdots -\Phi_v & I_3 \end{bmatrix},$$

$\boldsymbol{\Sigma}_{\mathbf{v}} = I_T \otimes \mathtt{diag}(\sigma^2_{v_c},\ \sigma^2_{v_d},\ \sigma^2_{v_{nd}})$, and the vector $\widetilde{\mathbf{v}}_0 = \iota_0 \otimes (\Phi_v v_0)$ collects the initial conditions for the common and sectoral uncertainty measures. ι_0 denotes a $T \times 1$ vector given by $\iota_0 = (1,\ 0, \cdots, 0)'$.

Next, recall from our discussion in Section 2 that z_t denotes a vector collecting the principal components (seven in total) extracted from a large set of macroeconomic indicators (i.e., the FRED-MD database). These are obtained following the same approach as in McCracken and Ng (2016). The other set of controls, f_t, denote a vector collecting latent dynamic factors (four in total) extracted from the 185 industries in our IP dataset. We define f_t as the vector collecting IP-specific factors. These are jointly modeled as a VAR(1)

process, i.e.:

$$f_t = \Phi_f f_{t-1} + \eta_{f,t}, \quad \eta_{f,t} \sim \mathcal{N}\left(0,\ \texttt{diag}\left(\sigma_{f_1}^2, \cdots, \sigma_{f_4}^2\right)\right). \tag{A4}$$

A1.1 Posterior Sampler

Let the parameters and states in our model be, respectively, organized into the following sets: $\boldsymbol{\theta} = \{\tilde{\boldsymbol{\beta}}_z,\ \tilde{\boldsymbol{\beta}}_f,\ \tilde{\boldsymbol{\lambda}}_c,\ \tilde{\boldsymbol{\lambda}}_d,\ \tilde{\boldsymbol{\lambda}}_{nd},\ \boldsymbol{\phi}_{\mathbf{f}},\ \boldsymbol{\phi}_{\mathbf{v}},\ \boldsymbol{\sigma}_{\mathbf{f}}^2,\ \boldsymbol{\sigma}_{\mathbf{v}}^2\}$ and $\mathcal{Z} = \{\boldsymbol{\psi},\ \mathbf{f},\ \mathbf{v}\}$.[1] Posterior draws are generated by sampling sequentially from:

$$\texttt{MCMC Steps:} \begin{cases} p\left(\mathbf{f}|\mathbf{y}, \mathcal{Z}_{-\mathbf{f}}, \boldsymbol{\theta}\right), \\ p\left(\boldsymbol{\psi}|\mathbf{y}, \mathcal{Z}_{-\boldsymbol{\psi}}, \boldsymbol{\theta}\right), \\ p\left(\mathbf{v}|\mathbf{y}, \mathcal{Z}_{-\mathbf{v}}, \boldsymbol{\theta}\right), \\ p\left(\tilde{\boldsymbol{\beta}}_x|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\tilde{\beta}_x}\right) \quad \text{for } x = z,\ f, \\ p\left(\tilde{\boldsymbol{\lambda}}_j|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\tilde{\boldsymbol{\lambda}}_j}\right) \quad \text{for } j = c,\ d,\ nd, \\ p\left(\boldsymbol{\phi}_{\boldsymbol{\ell}}|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\boldsymbol{\phi}_{\boldsymbol{\ell}}}\right), \\ p\left(\boldsymbol{\sigma}_{\boldsymbol{\ell}}^2|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\boldsymbol{\sigma}_{\boldsymbol{\ell}}^2}\right) \quad \text{for } \boldsymbol{\ell} = \mathbf{v},\ \mathbf{f}. \end{cases} \tag{A5}$$

In what follows we provide a detailed discussion on each of the steps above. For ease of exposition, we separate the discussion below into two parts, namely state simulation (i.e., the first three conditional posterior densities) and parameter sampling (i.e., the last four densities).

A1.1.1 State Simulation

(i) Drawing the IP-specific factors

Let $\mathbf{f} = (f_1, \cdots, f_T)'$ and $\mathbf{z} = (z_1, \cdots, z_T)'$ denote the stacked representation for the IP-specific and McCracken and Ng (2016) factors, respectively. By a simple change of variable,

[1] A detailed description of each element in these sets is provided in Sections 3.1 and 3.2 of the main text.

we can recast the controls in (A1) as $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_{\beta_f}\mathbf{f} + \mathbf{X}_{\beta_z}\mathbf{z}$. This yields:

$$\mathbf{y_f} = \mathbf{X}_{\beta_f}\mathbf{f} + \mathbf{u} \qquad \mathbf{u} \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_u\right), \tag{A6}$$

where:

$$\mathbf{y_f} = \mathbf{y} - \mathbf{X}_{\beta_z}\mathbf{z}, \quad \mathbf{u} = \mathbf{L}_\Psi\mathbf{L}_h\mathbf{e}, \quad \boldsymbol{\Sigma}_u = \mathbf{L}_\Psi\mathbf{L}_h\mathbf{L}_h'\mathbf{L}_\Psi', \quad \mathbf{X}_{\beta_j} = I_T \otimes \Lambda_{\beta_x} \quad \text{for } x = z,\ f,$$

$$\Lambda_{\beta_z} = \begin{bmatrix} \beta^z_{1,1} & \cdots & \beta^z_{7,1} \\ \vdots & \vdots & \vdots \\ \beta^z_{1,N} & \cdots & \beta^z_{7,N} \end{bmatrix}, \quad \Lambda_{\beta_f} = \begin{bmatrix} \beta^f_{1,1} & \cdots & \beta^f_{4,1} \\ \vdots & \vdots & \vdots \\ \beta^f_{1,N} & \cdots & \beta^f_{4,N} \end{bmatrix}.$$

Next, from (A4) we obtain the following stacked representation for $\mathbf{f}$:

$$\mathbf{L}_{\Phi_f}\mathbf{f} = \widetilde{\mathbf{f}}_0 + \boldsymbol{\eta}_\mathbf{f} \qquad \boldsymbol{\eta}_\mathbf{f} \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_\mathbf{f}\right), \tag{A7}$$

where:

$$\mathbf{L}_{\Phi_f} = \begin{bmatrix} I_4 & \mathbf{0} & \cdots & \mathbf{0} \\ -\Phi_f & I_4 & & \\ \mathbf{0} & -\Phi_f & \ddots & \vdots \\ \vdots & & \ddots & \\ \mathbf{0} & & \cdots -\Phi_f & I_4 \end{bmatrix}, \quad \widetilde{\mathbf{f}}_0 = \iota_0 \otimes \left(\Phi_f f_0\right), \quad \boldsymbol{\Sigma}_\mathbf{f} = I_T \otimes \mathtt{diag}\left(\sigma^2_{f_1}, \cdots, \sigma^2_{f_4}\right),$$

such that ι_0 denotes a $T \times 1$ vector given by $\iota_0 = (1,\ 0, \cdots, 0)'$ and $\sigma^2_{f_i}$ for $i = 1, \cdots, 4$ denotes the variance of the ith innovation driving the ith IP-specific factor. Combining (A6)–(A7) and applying standard linear regression results yields:

$$\mathbf{f}|\mathbf{y}, \mathcal{Z}_{-\mathbf{f}}, \boldsymbol{\theta} \sim \mathcal{N}\left(\overline{\mathbf{d}}_\mathbf{f},\ \overline{\mathbf{D}}_\mathbf{f}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_\mathbf{f} = \overline{\mathbf{D}}_\mathbf{f}\left(\mathbf{X}'_{\beta_f}\boldsymbol{\Sigma}_u^{-1}\mathbf{y_f} + \mathbf{L}'_{\Phi_f}\boldsymbol{\Sigma}_\mathbf{f}^{-1}\widetilde{\mathbf{f}}_0\right), \\ \overline{\mathbf{D}}_\mathbf{f} = \left(\mathbf{X}'_{\beta_f}\boldsymbol{\Sigma}_u^{-1}\mathbf{X}_{\beta_f} + \boldsymbol{\Sigma}_\mathbf{f}^{-1}\right)^{-1}. \end{cases} \tag{A8}$$

Note that to construct $\overline{\mathbf{D}}_\mathbf{f}$ one is required to invert an $NT \times NT$ matrix. Carrying out

such an operation via brute-force methods is computationally cumbersome. Therefore, we adopt a more efficient approach suggested by Chan and Jeliazkov (2009), often referred to as precision sampling. To illustrate how we employ their method, we first introduce the following notation: given a lower triangular $NT \times NT$ non-singular matrix $\mathbf{C}$ and a $NT \times 1$ vector $\mathbf{b}$, let $\mathbf{C} \setminus \mathbf{b}$ denote the unique solution to the triangular system $\mathbf{Cx} = \mathbf{b}$ obtained by forward substitution, i.e., $\mathbf{x} = \mathbf{C} \setminus \mathbf{b} = \mathbf{C}^{-1}\mathbf{b}$. Sampling $\mathbf{f}$ is then conducted by following the four operations below:

$$
\begin{aligned}
&(1) \quad Chol(\overline{\mathbf{D}}_{\mathbf{f}}^{-1}) = \mathbf{CC}', \\
&(2) \quad \mathbf{x} = \mathbf{C} \setminus \left(\mathbf{X}_{\beta_f}' \boldsymbol{\Sigma}_u^{-1} \mathbf{y}_{\mathbf{f}} + \mathbf{L}_{\Phi_f}' \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} \widetilde{\mathbf{f}}_0 \right), \\
&(3) \quad \overline{\mathbf{d}}_{\mathbf{f}} = \mathbf{C}' \setminus \mathbf{x}, \\
&(4) \quad \mathbf{f} = \overline{\mathbf{d}}_{\mathbf{f}} + \mathbf{C}' \setminus \boldsymbol{\zeta} \qquad \boldsymbol{\zeta} \sim \mathcal{N}\left(\mathbf{0},\ I_{NT}\right).
\end{aligned}
$$

The first step describes the Cholesky decomposition of the inverse covariance (or precision) matrix $\overline{\mathbf{D}}_{\mathbf{f}}^{-1}$. Step 2 requires solving a triangular system by forward substitution, given that $\mathbf{C}$ is a lower triangular matrix. Step 3 is analogous to Step 2, except that the solution of the triangular system, $\mathbf{C}' \setminus \mathbf{x}$, is now obtained by backward substitution, since $\mathbf{C}'$ is an upper triangular matrix. It is then straightforward to see that Steps 2 and 3 combined, by construction, yield:

$$
\begin{aligned}
\overline{\mathbf{d}}_{\mathbf{f}} &= \mathbf{C}'^{-1} \left(\mathbf{C}^{-1} \left(\mathbf{X}_{\beta_f}' \boldsymbol{\Sigma}_u^{-1} \mathbf{y}_{\mathbf{f}} + \mathbf{L}_{\Phi_f}' \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} \widetilde{\mathbf{f}}_0 \right) \right) = \left(\mathbf{CC}' \right)^{-1} \left(\mathbf{X}_{\beta_f}' \boldsymbol{\Sigma}_u^{-1} \mathbf{y}_{\mathbf{f}} + \mathbf{L}_{\Phi_f}' \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} \widetilde{\mathbf{f}}_0 \right) \\
&= \overline{\mathbf{D}}_{\mathbf{f}} \left(\mathbf{X}_{\beta_f}' \boldsymbol{\Sigma}_u^{-1} \mathbf{y}_{\mathbf{f}} + \mathbf{L}_{\Phi_f}' \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} \widetilde{\mathbf{f}}_0 \right).
\end{aligned}
$$

Finally, Step 4 describes an affine transformation of a standard normal random vector that ensures, by definition, that the expression in such a step returns a $NT \times 1$ random vector $\mathbf{f}|\mathbf{y}, \mathcal{Z}_{-\mathbf{f}}, \boldsymbol{\theta} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\mathbf{f}},\ \overline{\mathbf{D}}_{\mathbf{f}}\right)$.

• *Initial Conditions*

We treat the vector containing the state initialization conditions (i.e., f_0) as an additional parameter to our model. We thus extend our MCMC algorithm to sample f_0. Accordingly,

we rewrite (A7) as:

$$\mathbf{L}_{\Phi_f}\mathbf{f} = \mathbf{L}_{0f} f_0 + \boldsymbol{\eta}_{\mathbf{f}} \qquad \boldsymbol{\eta}_{\mathbf{f}} \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathbf{f}}\right), \tag{A9}$$

where:

$$\mathbf{L}_{0f} = \begin{bmatrix} \Phi_f \\ 0 \times I_4 \\ \vdots \\ 0 \times I_4 \end{bmatrix}.$$

Assuming a Gaussian prior $f_0 \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{f_0},\ \hat{\boldsymbol{\Sigma}}_{f_0}\right)$ and using standard regression results gives:

$$f_0|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{f_0},\ \overline{\mathbf{D}}_{f_0}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{f_0} = \overline{\mathbf{D}}_{f_0}\left(\mathbf{L}_{0f}'\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\mathbf{L}_{\Phi_f}\mathbf{f} + \hat{\boldsymbol{\Sigma}}_{f_0}^{-1}\hat{\boldsymbol{\mu}}_{f_0}\right), \\ \overline{\mathbf{D}}_{f_0} = \left(\mathbf{L}_{0f}'\boldsymbol{\Sigma}_{\mathbf{v}}^{-1}\mathbf{L}_{0f} + \hat{\boldsymbol{\Sigma}}_{f_0}^{-1}\right)^{-1}. \end{cases} \tag{A10}$$

(ii) Drawing the outlier-adjustment component

Recall from Section 2.3 that we elicit an inverse-gamma prior to (the square of) each element in the scale matrix $\Psi_t^{\frac{1}{2}} = \texttt{diag}\left(\psi_{1,t}^{\frac{1}{2}}, \cdots, \psi_{N,t}^{\frac{1}{2}}\right)$ collected by $\mathbf{L}_\Psi = \texttt{diag}\left(\Psi_1^{\frac{1}{2}}, \cdots, \Psi_T^{\frac{1}{2}}\right)$, i.e.:

$$\psi_{i,t} \overset{i.i.d.}{\sim} \mathcal{IG}\left(\frac{\nu_\psi}{2},\ \frac{\nu_\psi}{2}\right) \text{ for } i = 1, \cdots, N \text{ and } t = 1, \cdots, T. \tag{A11}$$

Since $\psi_{1,1}, \cdots, \psi_{N,T}$ are conditionally independent given the model parameters, states, and the data, we can draw each of them sequentially from the natural-conjugate inverse-gamma conditional posterior. In other words, we have:

$$\psi_{i,t}|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\psi_{i,t}} \sim \mathcal{IG}\left(\overline{\nu}_{i,t}^{\psi},\ \overline{S}_{i,t}^{\psi}\right), \text{ where } \begin{cases} \overline{\nu}_{i,t}^{\psi} = \frac{\nu_\psi + 1}{2}, \\ \overline{S}_{i,t}^{\psi} = \frac{\tilde{e}_{i,t}^2 + \nu_\psi}{2} \end{cases} \text{ for } i = 1, \cdots, 7 \text{ and } t = 1, \cdots, T, \tag{A12}$$

and $\tilde{e}_{i,t}$ denotes each element in the $NT \times 1$ vector of innovations given by $\tilde{\mathbf{e}} = \mathbf{L}_h\mathbf{e}$.

(iii) Drawing the common and sector-specific uncertainty factors

Recall from Section 3.2 that to sample $\mathbf{v}$ we adopt the auxiliary mixture-sampling approach by Omori et al. (2007). This entails a two-step procedure:

Step 1 $p(\mathbf{k}|\widetilde{\mathbf{y}}^*, \mathcal{Z}, \boldsymbol{\theta})$ $s.t.$ $\mathbf{k} = (k_1, \cdots, k_T)$,

Step 2 $p(\mathbf{v}|\widetilde{\mathbf{y}}^*, \mathcal{Z}_{-\mathbf{v}}, \mathbf{k}, \boldsymbol{\theta})$.

•Step 1

Each element of $\mathbf{k} = (k_1, \cdots, k_T)$ is drawn independently from a multinomial distribution parameterized by the full conditional posterior probabilities $\mathbb{P}\mathrm{r}\left(k_t = i|\widetilde{y}^*_{\tau,t}, z_t, \theta\right)$ given by:

$$\mathbb{P}\mathrm{r}\left(k_t = i|\widetilde{y}^*_t, \mathcal{Z}, \boldsymbol{\theta}\right) = \frac{\phi\left(h_t + \alpha_{k_t=i},\ \sigma^2_{s=i}\right) p_{k_t=i}}{\sum_{j=1}^{10} \phi\left(\Lambda_c v_{c,t} + \Lambda_s v_{s,t} + \alpha_{k=j},\ \sigma^2_{k_t=j}\right) p_{k_t=i}} \quad \text{for } i = 1, \cdots, 10,$$

where $\phi\left(\Lambda_c v_{c,t} + \Lambda_s v_{s,t} + \alpha_{k_t=j},\ \sigma^2_{k_t=j}\right)$ denotes a Gaussian density evaluated at mean $\Lambda_c v_{c,t} + \Lambda_s v_{s,t} + \alpha_{k_t=i}$ and variance $\sigma^2_{k_t=i}$. Again, the values for $\alpha_{k_t=i}$ and $\sigma^2_{k_t=i}$ are given in Table 1 in Omori et al. (2007) and $v_{c,t}$ and $v_{d,t}$ denote posterior draws for the uncertainty factors (in log form).

Given $\mathbb{P}\mathrm{r}\left(k_t = i|\widetilde{y}^*_t, \mathcal{Z}, \boldsymbol{\theta}\right)$, posterior draws for k_t can then be generated via the inverse transform method for $t = 1, \cdots, T$ as follows:[2]

(a) Generate $\zeta_t \sim Uniform(0,\ 1)$

(b) Find the smallest $i \in \{1, 2, \cdots, 10\}$ that satisfies $\sum_{j=1}^{i} \mathbb{P}\mathrm{r}\left(k_t = j|\widetilde{y}^*_t, \mathcal{Z}, \boldsymbol{\theta}\right) \geq \zeta_t$

(c) Return $\left(k_t|\widetilde{y}^*_t, \mathcal{Z}, \boldsymbol{\theta}\right) = i$

[2]See algorithm 3.2 in Kroese et al. (2013) for a more detailed discussion of the inverse transform method for discrete random variables.

•Step 2

As discussed in Section 3.2, conditional on $\mathbf{k}$ (the data, remaining states, and parameters), we use Bayes rule to obtain the following closed-form expression for $p(\mathbf{v}|\widetilde{\mathbf{y}}^*, \mathcal{Z}_{-\mathbf{v}}, \boldsymbol{\theta})$:

$$\mathbf{v}|\widetilde{\mathbf{y}}^*, \mathcal{Z}_{-\mathbf{v}}, \boldsymbol{\theta} \sim \mathcal{N}\left(\overline{\mathbf{d}}_v,\ \overline{\mathbf{D}}_v\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_v = \overline{\mathbf{D}}_v \left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_k^{-1}(\widetilde{\mathbf{y}}^* - \boldsymbol{\alpha}_k) + \mathbf{L}'_{\Phi_v}\boldsymbol{\Sigma}_v^{-1}\widetilde{\mathbf{v}}_0\right), \\ \overline{\mathbf{D}}_v = \left(\boldsymbol{\Lambda}'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\Lambda} + \mathbf{L}'_{\Phi_v}\boldsymbol{\Sigma}_v^{-1}\mathbf{L}_{\Phi_v}\right)^{-1}. \end{cases} \tag{A13}$$

Draws from $\mathcal{N}\left(\overline{\mathbf{d}}_v,\ \overline{\mathbf{D}}_v\right)$ are obtained using the same precision sampling techniques discussed for drawing $\mathbf{f}$.

• *Initial Conditions*

We treat the vector containing the state initialization conditions (i.e., v_0) as an additional parameter to our model. We thus extend our MCMC algorithm to sample v_0. Accordingly, we rewrite (A3) as:

$$\mathbf{L}_{\Phi_v}\mathbf{v} = \mathbf{L}_{0v}v_0 + \boldsymbol{\eta}_{\mathbf{v}} \qquad \boldsymbol{\eta}_{\mathbf{v}} \sim \mathcal{N}\left(\mathbf{0},\ \boldsymbol{\Sigma}_{\mathbf{v}}\right), \tag{A14}$$

where:

$$\mathbf{L}_{0v} = \begin{bmatrix} \Phi_v \\ 0 \times I_3 \\ \vdots \\ 0 \times I_3 \end{bmatrix}.$$

Assuming a Gaussian prior $v_0 \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{v_0},\ \hat{\boldsymbol{\Sigma}}_{v_0}\right)$ and using standard regression results gives:

$$v_0|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{v_0},\ \overline{\mathbf{D}}_{v_0}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{v_0} = \overline{\mathbf{D}}_{v_0} \left(\mathbf{L}'_{0v}\boldsymbol{\Sigma}_{\mathbf{v}}^{-1}\mathbf{L}_{\Phi_v}\mathbf{v} + \hat{\boldsymbol{\Sigma}}_{v_0}^{-1}\hat{\boldsymbol{\mu}}_{v_0}\right), \\ \overline{\mathbf{D}}_{v_0} = \left(\mathbf{L}'_{0v}\boldsymbol{\Sigma}_{\mathbf{v}}^{-1}\mathbf{L}_{0v} + \hat{\boldsymbol{\Sigma}}_{v_0}^{-1}\right)^{-1}. \end{cases} \tag{A15}$$

A1.1.2 Parameter Sampling

(iv) Drawing the loadings for the IP-specific and McCracken and Ng (2016) factors

As discussed in Section 3.1, not all loadings in $\boldsymbol{\beta}$ need to sampled. Specifically, we set the usual unit-lower-triangular-matrix normalization strategy to the separately identify $\boldsymbol{\beta}_f$ from the f_t. Since f_t is 4×1 vector, we apply the above-mentioned normalization to the loadings associated with the first four industries in y_t. We reiterate that no normalization is applied to $\boldsymbol{\beta}_z$, given z_t is obtained prior to estimation via principal-component techniques and is thus treated as an "observable" within our estimation framework.

With these in mind, let $\tilde{\boldsymbol{\beta}} = \begin{bmatrix} \tilde{\boldsymbol{\beta}}_z \\ \tilde{\boldsymbol{\beta}}_f \end{bmatrix}$ denote the vector that contains only non-normalized loadings and $\odot$ be the Hadamard (or element-wise) product. We can thus recast the controls in (A1) as $\mathbf{X}\boldsymbol{\beta} = \tilde{\mathbf{w}}_f + \mathbf{Z}\tilde{\boldsymbol{\beta}}_z + \mathbf{F}\tilde{\boldsymbol{\beta}}_f$, which gives:

$$\mathbf{y} = \tilde{\mathbf{w}}_f + \mathbf{Z}\tilde{\boldsymbol{\beta}}_z + \mathbf{F}\tilde{\boldsymbol{\beta}}_f + \mathbf{L}_h\mathbf{L}_\Psi\mathbf{e}, \tag{A16}$$

where:

$$\tilde{\mathbf{w}}_f = \left(I_T \otimes \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix} \right) \times \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_T \end{bmatrix}, \; \mathbf{Z} = \begin{bmatrix} I_N \otimes z_1' \\ I_N \otimes z_2' \\ \vdots \\ I_N \otimes z_T' \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} (I_N \otimes f_1') \odot \left(\tilde{I}_N \otimes (1,\ 1,\ 1,\ 1) \right) \\ (I_N \otimes f_2') \odot \left(\tilde{I}_N \otimes (1,\ 1,\ 1,\ 1) \right) \\ \vdots \\ (I_N \otimes f_T') \odot \left(\tilde{I}_N \otimes (1,\ 1,\ 1,\ 1) \right) \end{bmatrix} \text{ and } \tilde{I}_N = \begin{bmatrix} 0 & 0 & \cdots & & & 0 \\ 0 & 0 & & & & 0 \\ \vdots & & 0 & & & \vdots \\ & & & 0 & & \\ & & & & 1 & \\ & & & & & \ddots \\ 0 & 0 & \cdots & & & 1 \end{bmatrix}.$$

Next, defining $\mathbf{y}_{\tilde{\boldsymbol{\beta}}} = \mathbf{y} - \tilde{\mathbf{w}}_f,\ \mathbf{W} = [\mathbf{Z}\ \mathbf{F}]$, and $\mathbf{u} = \mathbf{L}_h \mathbf{L}_\Psi \mathbf{e}$, allows us to rewrite (A16) more compactly as:

$$\mathbf{y}_{\tilde{\boldsymbol{\beta}}} = \mathbf{W}\tilde{\boldsymbol{\beta}} + \mathbf{u} \quad \mathbf{u} \sim \mathcal{N}\left(0,\ \boldsymbol{\Sigma}_u\right). \tag{A17}$$

Assuming a Gaussian prior $\tilde{\boldsymbol{\beta}} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{\beta}}},\ \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}}\right)$ and applying Bayes rule yields:

$$\tilde{\boldsymbol{\beta}} | \mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\tilde{\beta}} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\boldsymbol{\beta}},\ \overline{\mathbf{D}}_{\boldsymbol{\beta}}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{\boldsymbol{\beta}} = \overline{\mathbf{D}}_{\boldsymbol{\beta}} \left(\mathbf{W}' \boldsymbol{\Sigma}_u^{-1} \mathbf{y}_{\tilde{\boldsymbol{\beta}}} + \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}}^{-1} \hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{\beta}}} \right), \\ \overline{\mathbf{D}}_{\boldsymbol{\beta}} = \left(\mathbf{W}' \boldsymbol{\Sigma}_u^{-1} \mathbf{W} + \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}}^{-1} \right)^{-1}. \end{cases} \tag{A18}$$

(v) Drawing the loadings for common and sectoral uncertainty factors

Akin to $\boldsymbol{\beta}$, we only need to draw the non-normalized loadings in $\boldsymbol{\Lambda}$. To do so, recall first from (19) in the main text that, conditional on a draw for the vector $\mathbf{k}$ (i.e., the

auxiliary state variable in the mixture-sampling approach by Omori et al. (2007)), we obtain a representation for the measurement equation that casts $\mathbf{v}$ in conditional linear Gaussian form given by:

$$\widetilde{\mathbf{y}}^* = \mathbf{\Lambda v} + \boldsymbol{\alpha}_k + \widetilde{\mathbf{e}}_k^*. \tag{A19}$$

Next, by a change of variable we can express $\mathbf{\Lambda v} = \tilde{\mathbf{w}}_v + \mathbf{V}\tilde{\boldsymbol{\lambda}}$, which gives:

$$\widetilde{\mathbf{y}}^* = \tilde{\mathbf{w}}_v + \mathbf{V}\tilde{\boldsymbol{\lambda}} + \boldsymbol{\alpha}_k + \widetilde{\mathbf{e}}_k^*, \tag{A20}$$

where:

$$\tilde{\mathbf{w}}_v = \left(I_T \otimes \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix} \right) \times \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \end{bmatrix}, \quad v_t = \begin{bmatrix} v_{c,t} \\ v_{d,t} \\ v_{nd,t} \end{bmatrix} \text{ for } t = 1, ..., T,$$

$$\mathbf{V} = [\mathbf{V}_c \ \mathbf{V}_s], \text{ s.t. } \mathbf{V}_c = \mathtt{diag}\left(V_{c,1}, \cdots, V_{c,T}\right), \ \mathbf{V}_s = \mathtt{diag}\left(V_{s,1}, \cdots, V_{s,T}\right),$$

$$V_{c,t} = v_{c,t} I_N, \quad V_{s,t} = \begin{bmatrix} 0 & 0 & \cdots & & & 0 & \cdots & & 0 \\ 0 & v_{d,t} & & & & 0 & \ddots & & 0 \\ \vdots & & \ddots & & & \vdots & & & \vdots \\ 0 & \cdots & & v_{d,t} & & 0 & \cdots & & 0 \\ 0 & \cdots & & & 0 & 0 & \cdots & & 0 \\ 0 & \ddots & & & 0 & v_{nd,t} & \cdots & & 0 \\ \vdots & & & & & \vdots & \ddots & 0 & \vdots \\ 0 & \cdots & & & & 0 & \cdots & & 0 \\ 0 & \cdots & & & & 0 & \cdots & & v_{nd,t} \end{bmatrix} \text{ and } \tilde{\boldsymbol{\lambda}} = \begin{bmatrix} \tilde{\boldsymbol{\lambda}}_c \\ \tilde{\boldsymbol{\lambda}}_d \\ \tilde{\boldsymbol{\lambda}}_{nd} \end{bmatrix}.$$

In other words, $\tilde{\mathbf{w}}_v$ collects (in stacked form) the normalization conditions; i.e., the first loading corresponding to the common, durable, and nondurable uncertainty factors is set to one, and $\tilde{\boldsymbol{\lambda}}$ denotes the $(3N-3)\times 1$ vector containing the non-normalized loadings associated with our common and sectoral uncertainty factors.

Setting $\widetilde{\mathbf{y}}^*_\lambda = \widetilde{\mathbf{y}}^* - \tilde{\mathbf{w}}_v - \boldsymbol{\alpha}_k$, assuming a Gaussian prior $\tilde{\boldsymbol{\lambda}} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{\lambda}}},\ \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\lambda}}}\right)$ and applying Bayes rule yields:

$$\tilde{\boldsymbol{\lambda}}|\mathbf{y}, \mathcal{Z}, \mathbf{k}, \boldsymbol{\theta}_{-\tilde{\boldsymbol{\lambda}}} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\tilde{\boldsymbol{\lambda}}},\ \overline{\mathbf{D}}_{\tilde{\boldsymbol{\lambda}}}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{\tilde{\boldsymbol{\lambda}}} = \overline{\mathbf{D}}_{\tilde{\boldsymbol{\lambda}}}\left(\mathbf{V}'\boldsymbol{\Sigma}_{\mathbf{k}}^{-1}\widetilde{\mathbf{y}}^*_\lambda + \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\lambda}}}^{-1}\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{\lambda}}}\right), \\ \overline{\mathbf{D}}_{\tilde{\boldsymbol{\lambda}}} = \left(\mathbf{V}'\boldsymbol{\Sigma}_{\mathbf{k}}^{-1}\mathbf{V} + \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\lambda}}}^{-1}\right)^{-1}. \end{cases} \tag{A21}$$

Again, the values for $\boldsymbol{\alpha}_k$ and $\boldsymbol{\Sigma}_k$ in the expression for the full conditional posterior density above are predetermined and given in Table 1 of Omori et al. (2007).

(vi) Drawing VAR coefficients for factor law of motion

We assume a VAR(1) law of motion for the vector collecting both the common and sectoral uncertainty factors (v_t) and the four IP-specific latent factors in the measurement equation (f_t). Stacking each of these over t yields:

$$\mathbf{v} = \mathbf{X}_\mathbf{v}\boldsymbol{\phi}_\mathbf{v} + \boldsymbol{\eta}_\mathbf{v} \qquad \boldsymbol{\eta} \sim \mathcal{N}\left(0,\ \boldsymbol{\Sigma}_\mathbf{v}\right), \tag{A22}$$

$$\mathbf{f} = \mathbf{X}_\mathbf{f}\boldsymbol{\phi}_\mathbf{f} + \boldsymbol{\eta}_\mathbf{f} \qquad \boldsymbol{\eta}_\mathbf{f} \sim \mathcal{N}\left(0,\ \boldsymbol{\Sigma}_\mathbf{f}\right), \tag{A23}$$

where:

$$\mathbf{X_v} = \begin{bmatrix} I_3 \otimes v_0' \\ \vdots \\ I_3 \otimes v_{T-1}' \end{bmatrix} \quad \mathbf{X_f} = \begin{bmatrix} I_4 \otimes f_0' \\ \vdots \\ I_4 \otimes f_{T-1}' \end{bmatrix}.$$

Assuming a Gaussian prior $\boldsymbol{\phi_\ell} \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{\phi_\ell}},\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\phi_\ell}}\right)$ and applying Bayes rule yields:

$$\boldsymbol{\phi_\ell}|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\boldsymbol{\phi_\ell}} \sim \mathcal{N}\left(\overline{\mathbf{d}}_{\boldsymbol{\phi_\ell}},\ \overline{\mathbf{D}}_{\boldsymbol{\phi_\ell}}\right), \text{ where } \begin{cases} \overline{\mathbf{d}}_{\boldsymbol{\phi_\ell}} = \overline{\mathbf{D}}_{\boldsymbol{\phi_\ell}}\left(\mathbf{X}_{\boldsymbol{\ell}}'\boldsymbol{\Sigma}_{\boldsymbol{\ell}}^{-1}\boldsymbol{\ell} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\phi_\ell}}^{-1}\hat{\boldsymbol{\mu}}_{\boldsymbol{\phi_\ell}}\right), \\ \overline{\mathbf{D}}_{\boldsymbol{\phi_\ell}} = \left(\mathbf{X}_{\boldsymbol{\ell}}'\boldsymbol{\Sigma}_{\boldsymbol{\ell}}^{-1}\mathbf{X}_{\boldsymbol{\ell}} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\phi_\ell}}^{-1}\right)^{-1} \quad \text{for } \boldsymbol{\ell} = \mathbf{v},\ \mathbf{f}. \end{cases} \tag{A24}$$

(vii) **Drawing the conditional variance of first- and second-moment states**

We collect the variance parameters for the innovations driving $\mathbf{v}$ and $\mathbf{f}$ in the vectors $\boldsymbol{\sigma}_{\mathbf{v}}^2 = \left(\sigma_{v_c}^2,\ \sigma_{v_d}^2,\ \sigma_{v_{nd}}^2\right)'$ and $\boldsymbol{\sigma}_{\mathbf{f}}^2 = \left(\sigma_{f_1}^2,\ \sigma_{f_2}^2,\ \sigma_{f_3}^2,\ \sigma_{f_4}^2\right)'$. Assuming an inverse-gamma prior, $\mathcal{IG}\left(\nu,\ S\right)$, such that $S = 0.2^2(\nu - 1)$, $S = (\nu - 1)$ and $\nu = \frac{T}{10}$ and applying Bayes rule yields:

$$\sigma_{v_i}^2|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\sigma_{v_i}^2} \sim \mathcal{IG}\left(\overline{\nu}_{v_i},\ \overline{S}_{v_i}\right), \text{ where } \begin{cases} \overline{\nu}_{v_i} = \frac{T}{2} + \nu, \\ \overline{S}_{v_i} = \frac{\sum_{t=1}^{T}\left(\eta_{v_i,t}\right)^2}{2} + S \quad \text{for } i = c, d, nd \end{cases} \tag{A25}$$

and

$$\sigma_{f_i}^2|\mathbf{y}, \mathcal{Z}, \boldsymbol{\theta}_{-\sigma_{f_i}^2} \sim \mathcal{IG}\left(\overline{\nu}_{f_i},\ \overline{S}_{f_i}\right), \text{ where } \begin{cases} \overline{\nu}_{f_i} = \frac{T}{2} + \nu, \\ \overline{S}_{f_i} = \frac{\sum_{t=1}^{T}\left(\eta_{f_i,t}\right)^2}{2} + S \quad \text{for } i = 1, 2, 3, 4. \end{cases} \tag{A26}$$

A1.2 Computational Performance

We assess the performance of the MCMC algorithm discussed in the previous section based on two criteria: (i) its mixing properties and (ii) computational speed.

• *Inefficiency Factors*

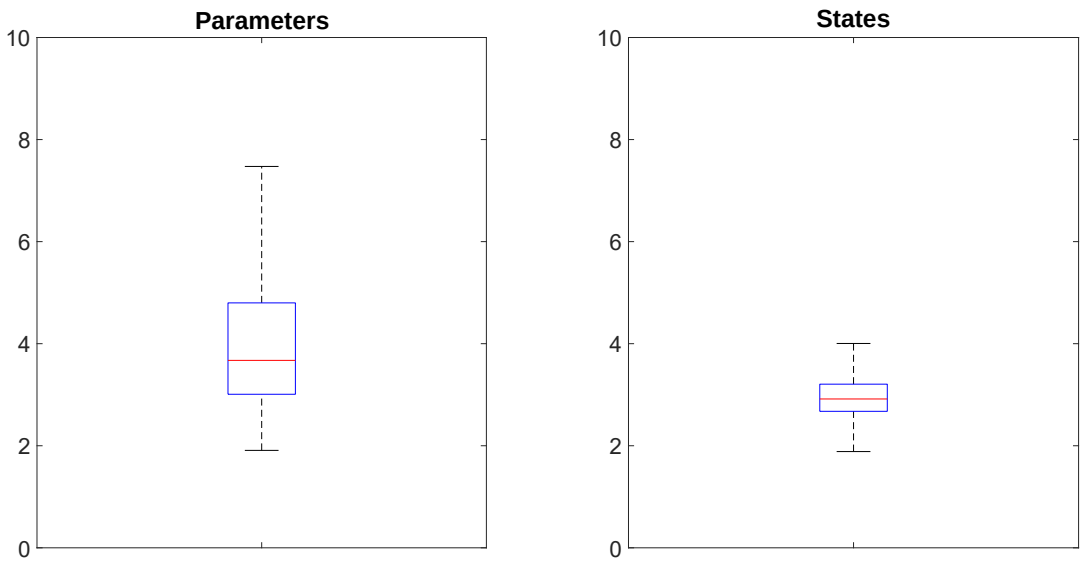
To address point (i), below we report two sets of inefficiency factors: one associated with the draws for the states and one associated with the draws for the parameters. Such metrics are computed using a common approach (see, e.g., Chib (2001)) given by:

$$1 + 2\sum_{j=1}^{J} \rho_j,$$

where ρ_j is the sample autocorrelation at lag j through lag J. In our empirical application we set J to be large enough until autocorrelation tapers off. In an ideal setting where MCMC draws are virtually independent draws, inefficiency factors should be one. As a rule of thumb, inefficiency factors around twenty are typically interpreted as an indication of fast mixing.[3] Figure A1 reports boxplots to summarize inefficiency factor results. The middle line denotes the median inefficiency factor. The lower and upper lines, respectively, represent the 25 and 75 percentiles, while whiskers extend to the maximum and minimum inefficiency factors. All in all, results in Figure A1 demonstrate that our posterior sampler exhibits good mixing properties.

[3]Another way to interpret the inefficiency factor adopted here is to think that an inefficiency factor of 100 means that approximately 10,000 posterior draws are required to convey the same information as 100 independent draws.

Figure A1: Inefficiency Factors



• *Computational Speed*

It takes approximately 30 minutes to generate 10,000 MCMC draws for our baseline model. Estimation routines were implemented on a desktop with an Intel Xeon E5-2690 v2 @3.00 GHz processor.

A2 Robustness Checks

In this section we report results for the robustness checks we performed. These examined the strength of our proposed measu res of common and sectoral uncertainty to (i) different parameterizations of the hierarchical common stochastic volatility model discussed in Section 2; and (ii) different VAR specifications relative to the one presented in Section 4. Overall, our findings documented in the main text carry over to all such checks.

A2.1 Uncertainty Measurement under Alternative Specifications

We begin by showing the estimated measures of common and sectoral uncertainty under different specifications of X_t in Equation (1) in the paper, i.e., the matrix collecting the

controls in the measurement equation. In particular, in addition to the baseline specification in the main text (labeled as M1 hereafter), we allowed for four alternative representations of X_t. These are summarized below in Table A1 and formalized in greater detail in Equations (A27)–(A31).

Table A1: Different Model Specifications

Model Identifier	Control Variables in X_t
M1	7 factors obtained from the FRED-MD database 4 IP-specific factors
M2	7 factors obtained from the FRED-MD database 4 IP-specific factors featuring stochastic volatility
M3	7 factors obtained from the FRED-MD database
M4	4 IP-specific factors
M5	4 IP-specific factors featuring stochastic volatility

$$\texttt{Controls for M1:}\begin{cases} X_t = [I_N \otimes z_t' \quad I_N \otimes f_t'], \\ z_t = (z_{1,t}, \cdots, z_{7,t})' \quad \text{– McCracken and Ng factors,} \\ f_t = (f_{1,t}, \cdots, f_{4,t})' \quad \text{– IP-specific factors,} \\ f_t = \Phi_f f_{t-1} + \eta_{f,t} \quad \eta_{f,t} \sim \mathcal{N}\left(0,\ \texttt{diag}\left(\sigma^2_{f_1}, \cdots, \sigma^2_{f_4}\right)\right) \end{cases} \tag{A27}$$

$$\texttt{Controls for M2:}\begin{cases} X_t = [I_N \otimes z_t' \quad I_N \otimes f_t'], \\ z_t = (z_{1,t}, \cdots, z_{7,t})' \quad \text{– McCracken and Ng factors,} \\ f_t = (f_{1,t}, \cdots, f_{4,t})' \quad \text{– IP-specific factors,} \\ f_t = \Phi_f f_{t-1} + \eta_{f,t}, \quad \eta_{f,t} \sim \mathcal{N}\left(0,\ \texttt{diag}(\exp(h^f_{i,t}))\right), \\ h^f_{i,t} = h^f_{i,t-1} + \zeta_{i,t}, \quad \zeta_{i,t} \sim \mathcal{N}\left(0,\ \sigma^2_{\zeta_i}\right) \text{ for } i = 1, ..., 4 \end{cases} \tag{A28}$$

$$\texttt{Controls for M3:} \begin{cases} X_t = I_N \otimes z_t', \\ z_t = (z_{1,t}, \cdots, z_{7,t})' \quad - \text{McCracken and Ng factors} \end{cases} \tag{A29}$$

$$\texttt{Controls for M4:} \begin{cases} X_t = I_N \otimes f_t', \\ f_t = (f_{1,t}, \cdots, f_{4,t})' \quad - \text{IP-specific factors}, \\ f_t = \Phi_f f_{t-1} + \eta_{f,t} \quad \eta_{f,t} \sim \mathcal{N}\left(0,\ \texttt{diag}\left(\sigma_{f_1}^2, \cdots, \sigma_{f_4}^2\right)\right) \end{cases} \tag{A30}$$

$$\texttt{Controls for M5:} \begin{cases} X_t = I_N \otimes f_t', \\ f_t = (f_{1,t}, \cdots, f_{4,t})' \quad - \text{IP-specific factors}, \\ f_t = \Phi_f f_{t-1} + \eta_{f,t}, \quad \eta_{f,t} \sim \mathcal{N}\left(0,\ \texttt{diag}(\exp(h_{i,t}^f))\right), \\ h_{i,t}^f = h_{i,t-1}^f + \zeta_{i,t}, \quad \zeta_{i,t} \sim \mathcal{N}\left(0,\ \sigma_{\zeta_i}^2\right) \text{ for } i = 1, ..., 4 \end{cases} \tag{A31}$$

Figures A2 and A3 present the measures of common and sectoral uncertainty, respectively, obtained from the five variants described in Table A1. All specifications deliver similar results both qualitatively and quantitatively. Such similarity is reflected in the measures from all models being virtually perfectly correlated. We do note, however, that models M4 and M5 find higher uncertainty peaks during recessions. To select which specification suggests a more suitable fit, we conducted a model comparison based on the deviance information criteria (DIC). Details on this exercise are provided below.

Figure A2: **Common Uncertainty (Alternative Models).** Measure of uncertainty that is common across 185 industries from disaggregated Industrial Production data. Estimates denote series estimated with the baseline model M1 and alternative specifications {M2, M3, M4, M5} as described in Table A1. Vertical shaded bars correspond to NBER recession dates. $\rho_{i,j}$ for $i, j = 1, ..., 5$ and $i \neq j$ denotes the (unconditional) correlation between uncertainty measures obtained from models M_i and M_j.

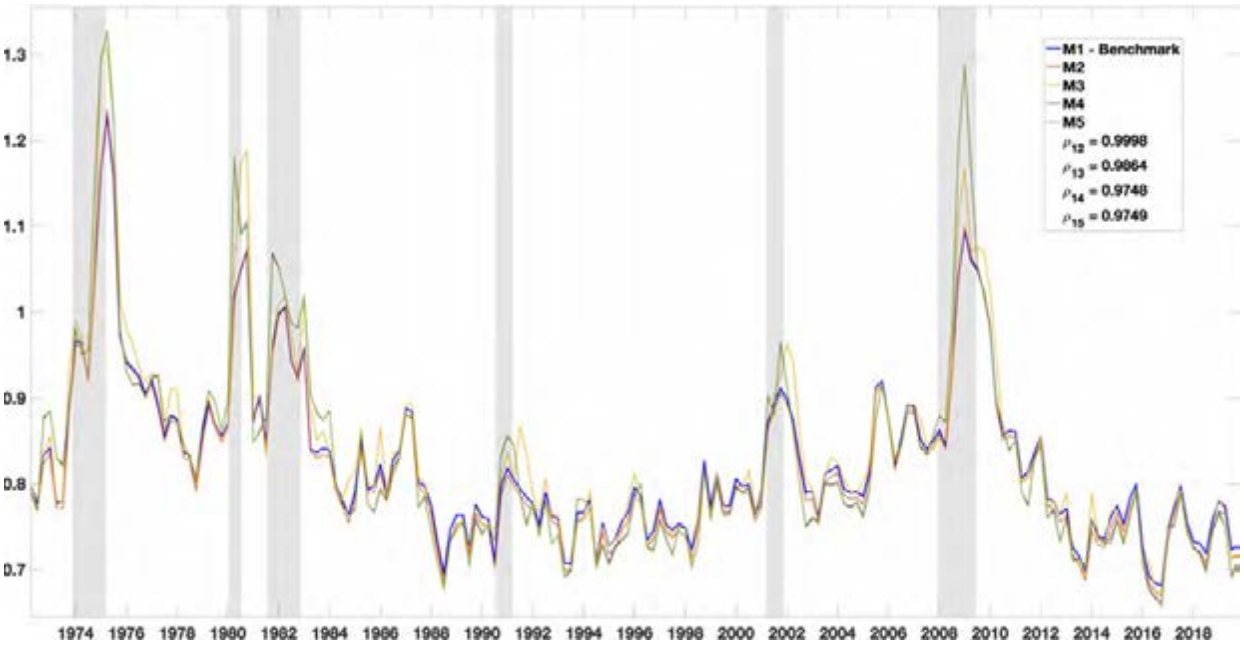
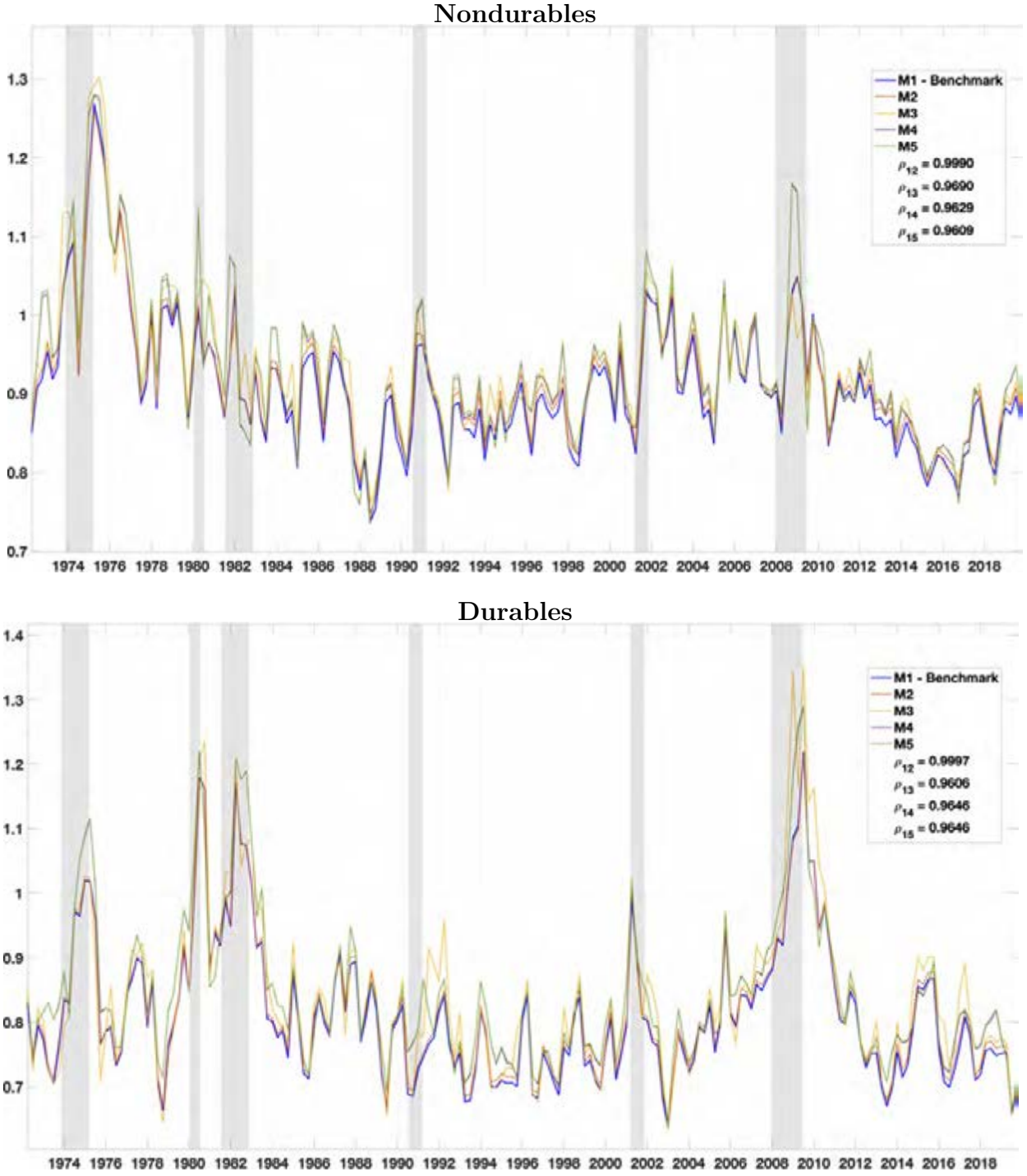


Figure A3: **Sectoral Uncertainty (Alternative Models).** Measures of durables and nondurables uncertainty obtained from disaggregated Industrial Production data. Estimates denote series estimated with the baseline model M1 and alternative specifications {M2, M3, M4, M5} as described in Table A1. Vertical shaded bars correspond to NBER recession dates. $\rho_{i,j}$ for $i, j = 1, ..., 5$ and $i \neq j$ denotes the (unconditional) correlation between uncertainty measures obtained from models M_i and M_j.



Model Comparison

Formal model comparison via Bayes-factor computation is challenging for models, such as ours, that allow for stochastic volatility. Therefore, we adopt a simpler approach to conduct model selection by computing the DIC associated with each model in Table A1. Following Berg et al. (2004) we calculate this statistic as follows:

$$\text{DIC} = -4 \sum_{l=1}^{L} \log p(\mathbf{y}|\mathcal{Z}^l_{\text{mode}}\boldsymbol{\theta}^l) + 2 \log p(\mathbf{y}|\mathcal{Z}_{\text{mode}}, \boldsymbol{\theta}_{\text{mode}}),$$

where $\mathcal{Z}_{\text{mode}}$, $\boldsymbol{\theta}_{\text{mode}}$ denotes the maximum a posteriori estimates that maximize the likelihood function, $p(\mathbf{y}|\mathcal{Z}_{\text{mode}}, \boldsymbol{\theta}_{\text{mode}})$. To compute the expression above we use $L = 50000$ post-burnin MCMC draws. Table A2 reports the DIC for each model.

Table A2: DIC Estimates

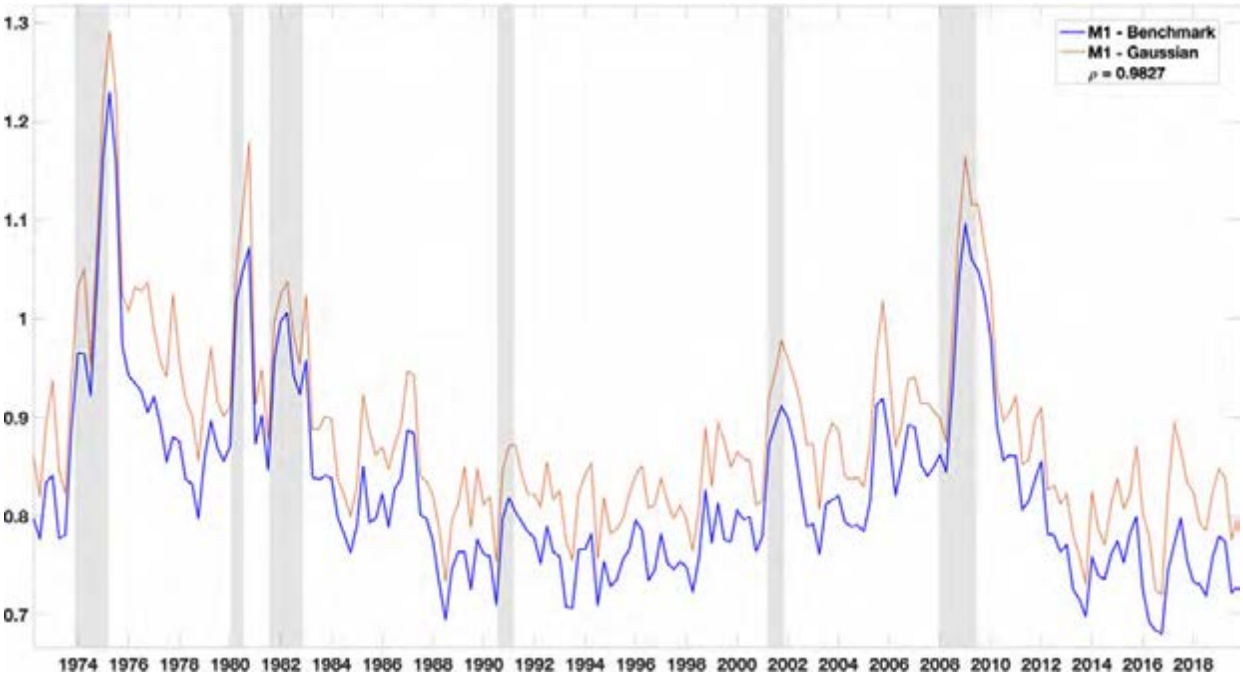
M1	M2	M3	M4	M5
$\mathbf{-9.07 \times 10^6}$	-9.06×10^6	-8.04×10^6	-8.28×10^6	-8.28×10^6

Notes: Number in bold denotes the best in-sample fit model.

A2.2 A "Strictly" Gaussian Representation

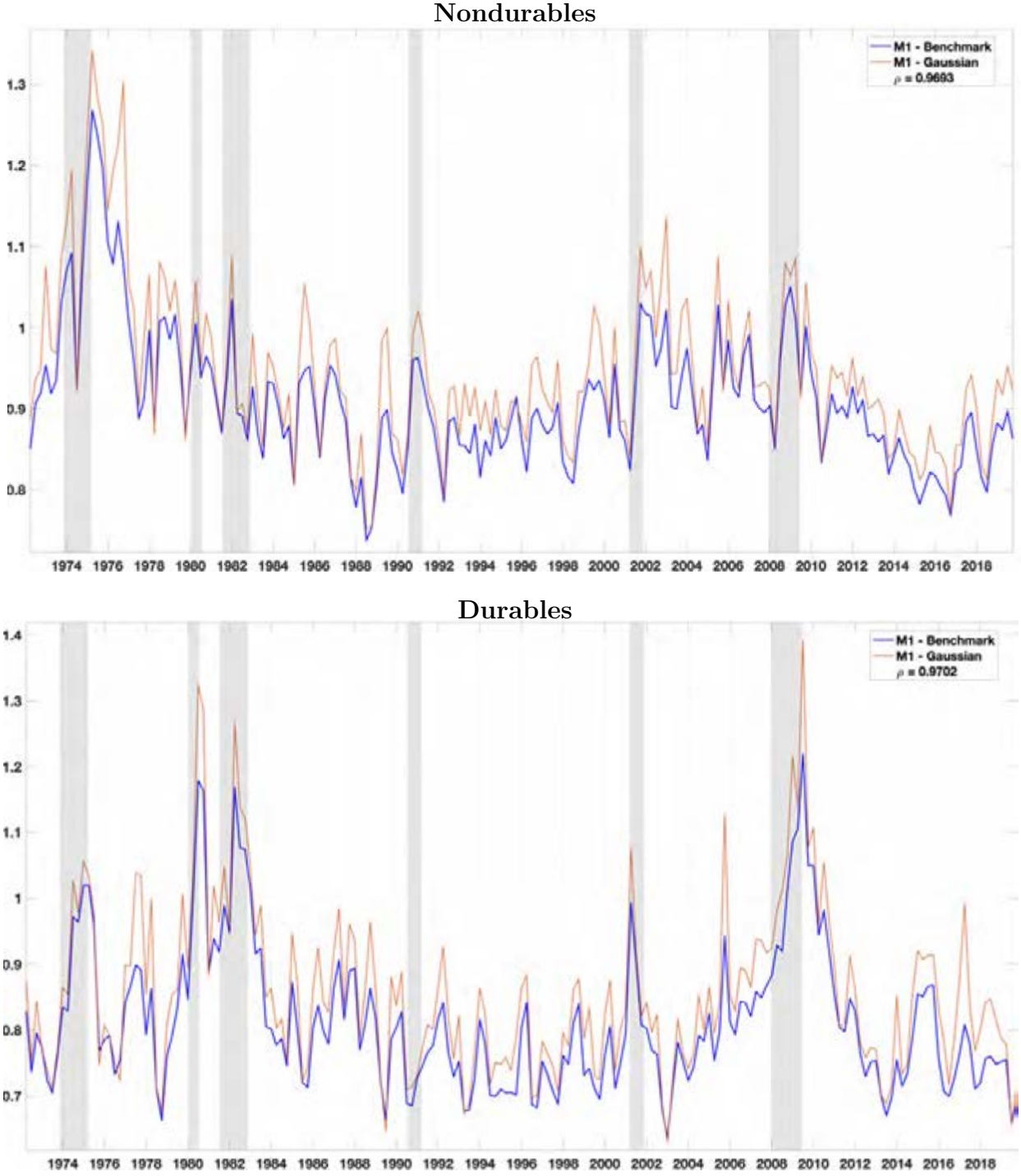
Figures A4 and A5 report our estimates for common and sectoral uncertainty based on a representation of our baseline model (M1) that excludes the scaling matrix $\mathbf{L}_\Psi$ in (A1). We refer to such a representation as the Gaussian variant of our baseline model since, as discussed in Section 2.3, the composite error term $\mathbf{L}_\Psi \mathbf{L}_h \mathbf{e}$ marginalized over $\mathbf{L}_\Psi$, follows a Student-t distribution. Uncertainty measures derived from this "strictly" Gaussian specification are in line with what we expected. In other words, we note a modest upward level shift in all uncertainty series. This follows from the fact that large volatility shifts, albeit idiosyncratic, have a greater impact on our uncertainty measures in the absence of an industry-specific scale parameter.[4]

Figure A4: **Common Uncertainty (Student-t versus Gaussian).** Measure of uncertainty common across 185 IP sectors estimated with the benchmark model M1 and its Gaussian version as an alternative modeling approach. Vertical shaded bars correspond to NBER recession dates. ρ denotes the (unconditional) correlation between the series obtained from each specification.



[4]We have computed the DIC for the Gaussian and Student-t versions of our baseline model, with the latter receiving considerably stronger support from the data.

Figure A5: **Sectoral Uncertainty (Student-t versus Gaussian).** Measure of nondurables and durables uncertainties estimated with the benchmark model M1 and its Gaussian version as an alternative modeling approach. Vertical shaded bars correspond to NBER recession dates. ρ denotes the (unconditional) correlation between the series obtained from each specification.



A2.3 Alternative VAR Specifications

As discussed in Section 4.5, we also examine whether our impulse response results would hold under different VAR specifications. To this end, we recompute impulse responses from VAR models with seven macroeconomic variables (similar to Bloom (2009)) and a smaller scale VAR with five macroeconomic variables (similar to Leduc and Liu (2016) and Alessandri and Mumtaz (2019)). Figures A6 and A7 present the results based on these different specifications. Notably, our key finding on the expansionary (contractionary) effect of nondurables (durables) uncertainty on economic activity carries over to all these alternative VAR specifications.

Figure A6: **Responses of Real Activity to Uncertainty Shocks (a VAR with 7 macroeconomic variables).** Impulse responses denote a one standard deviation shock to the common, nondurables, and durables uncertainty measures. In addition to the uncertainty measures, the VAR model contains $[IP, EMP, PCE, WAGE, HOURS, FFR, SP500]$ as the macroeconomic variables. Solid lines denote posterior medians, and the shaded areas represent the 67% equal-tailed posterior credible interval.

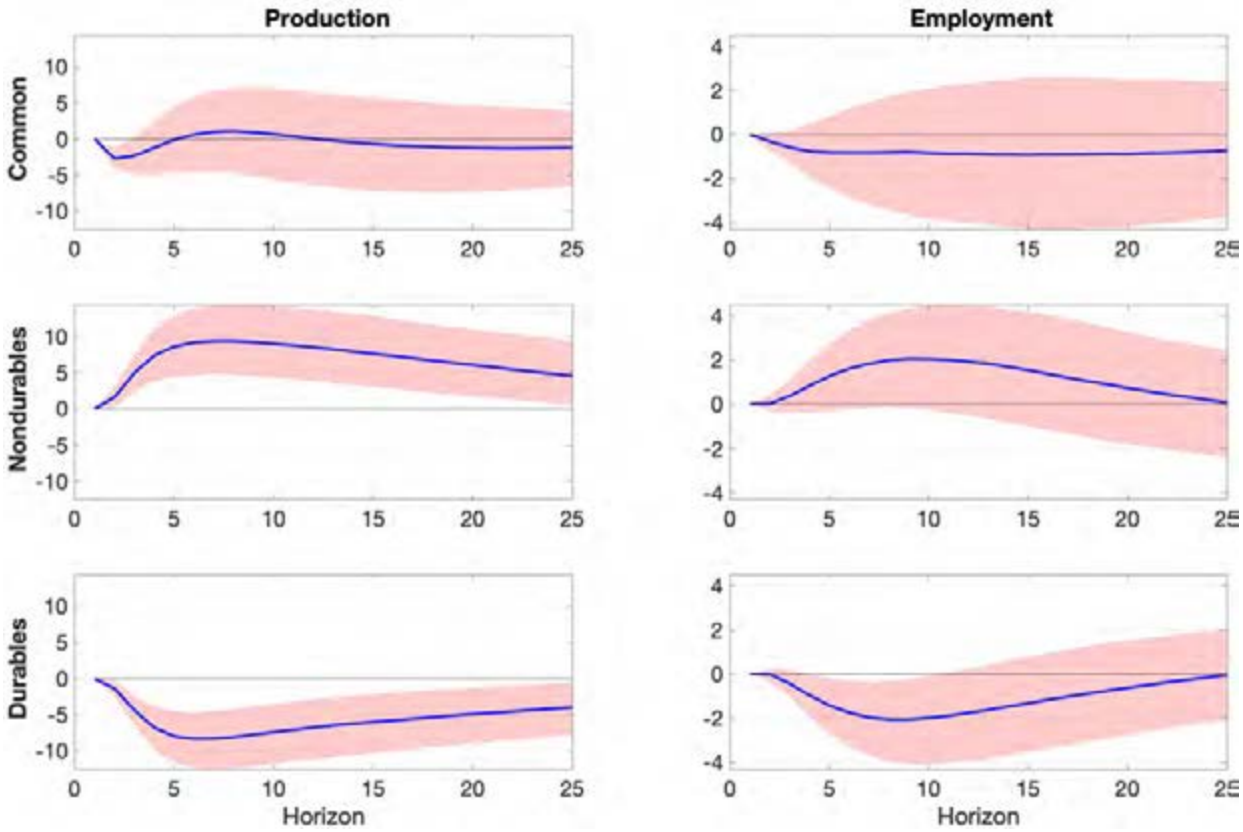
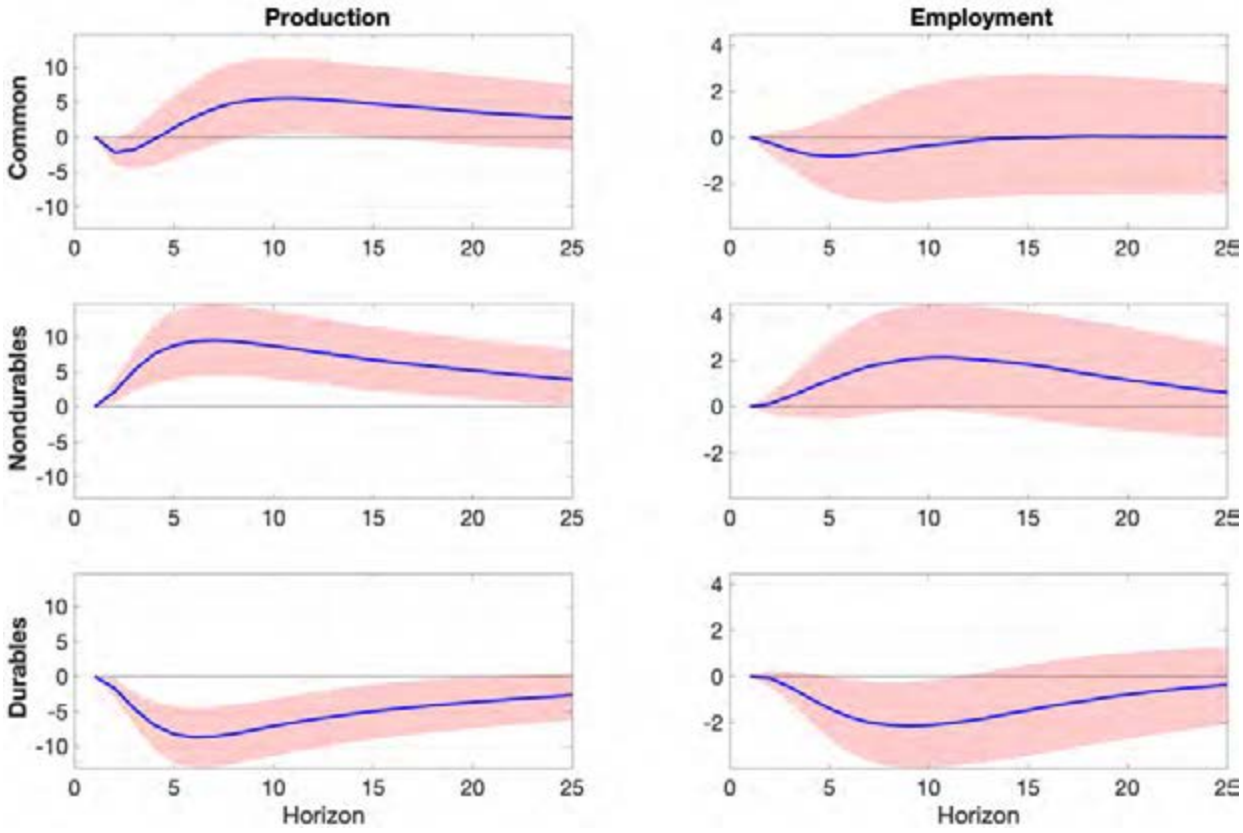


Figure A7: **Responses of Real Activity to Uncertainty Shocks (a VAR with 5 macroeconomic variables).** Impulse responses denote a one standard deviation shock to the common, nondurables, and durables uncertainty measures. In addition to the uncertainty measures, the VAR model contains $[IP, EMP, PCE, FFR]$ as the macroeconomic variables. Solid lines denote posterior medians, and the shaded areas represent the 67% equal-tailed posterior credible interval.



A3 Dataset Description

Table A3 provides additional details on all the industries listed in the disaggregated U.S. industrial production dataset adopted in our study. Sectors (i.e., durables and nondurables) are defined in accordance with the North American Industry Classification System (NAICS) and are thus given by the first two NAICS digits. There are 80 series in the nondurables sector, covering industries mainly in the food, textile, paper and printing, and petrochemical subsectors. The durables sector accounts for 105 series, covering industries mainly in the metal and nonmetal manufacturing, machinery and electronics, and transportation equipment subsectors.

Table A3: The table presents the U.S. Industrial Production Series with their associated sector names and NAICS codes.

Sector	**Industry**	**NAICS Code**
Manufacturing - Nondurables	Food	311
Manufacturing - Nondurables	Animal food	3111
Manufacturing - Nondurables	Grain and oilseed milling	3112
Manufacturing - Nondurables	Sugar and confectionery product	3113
Manufacturing - Nondurables	Fruit and vegetable preserving and specialty food	3114
Manufacturing - Nondurables	Dairy product	3115
Manufacturing - Nondurables	Dairy product (except frozen)	31151
Manufacturing - Nondurables	Fluid milk	311511
Manufacturing - Nondurables	Creamery butter	311512
Manufacturing - Nondurables	Cheese	311513
Manufacturing - Nondurables	Dry, condensed, and evaporated dairy product	311514
Manufacturing - Nondurables	Ice cream and frozen dessert	31152
Manufacturing - Nondurables	Animal slaughtering and processing	3116
Manufacturing - Nondurables	Animal (except poultry) slaughtering and meat processing	311611-3
Manufacturing - Nondurables	Beef	311611-3pt.
Manufacturing - Nondurables	Pork	311611-3pt.

Sector	Industry	NAICS Code
Manufacturing - Nondurables	Miscellaneous meats	311611-3pt.
Manufacturing - Nondurables	Poultry processing	311615
Manufacturing - Nondurables	Bakeries and tortilla	3118
Manufacturing - Nondurables	Other food	3119
Manufacturing - Nondurables	Coffee and tea	31192
Manufacturing - Nondurables	Beverage and tobacco product	312
Manufacturing - Nondurables	Beverage	3121
Manufacturing - Nondurables	Soft drink and ice	31211
Manufacturing - Nondurables	Breweries	31212
Manufacturing - Nondurables	Tobacco	3122
Manufacturing - Nondurables	Textile mills	313
Manufacturing - Nondurables	Fiber, yarn, and thread mills	3131
Manufacturing - Nondurables	Fabric mills	3132
Manufacturing - Nondurables	Textile and fabric finishing and fabric coating mills	3133
Manufacturing - Nondurables	Textile product mills	314
Manufacturing - Nondurables	Textile furnishings mills	3141
Manufacturing - Nondurables	Carpet and rug mills	31411
Manufacturing - Nondurables	Other textile product mills	3149
Manufacturing - Nondurables	Apparel	315
Manufacturing - Nondurables	Leather and allied products	316
Manufacturing - Nondurables	Paper	322
Manufacturing - Nondurables	Pulp, paper, and paperboard mills	3221
Manufacturing - Nondurables	Pulp mills	32211
Manufacturing - Nondurables	Paper mills	32212
Manufacturing - Nondurables	Paper (except newsprint) mills	322121
Manufacturing - Nondurables	Paperboard mills	32213
Manufacturing - Nondurables	Converted paper product	3222
Manufacturing - Nondurables	Paperboard container	32221
Manufacturing - Nondurables	Paper bag and coated and treated paper	32222

Sector	Industry	NAICS Code
Manufacturing - Nondurables	Other converted paper products	32223,9
Manufacturing - Nondurables	Printing and related support activities	323
Manufacturing - Nondurables	Petroleum and coal products	324
Manufacturing - Nondurables	Petroleum refineries	32411
Manufacturing - Nondurables	Aviation fuel and kerosene	32411pt.
Manufacturing - Nondurables	Distillate fuel oil	32411pt.
Manufacturing - Nondurables	Automotive gasoline	32411pt.
Manufacturing - Nondurables	Residual fuel oil	32411pt.
Manufacturing - Nondurables	Other refinery output	32411pt.
Manufacturing - Nondurables	Paving, roofing, and other petroleum and coal products	32412,9
Manufacturing - Nondurables	Chemicals	325
Manufacturing - Nondurables	Basic chemical	3251
Manufacturing - Nondurables	Organic chemicals	32511,9
Manufacturing - Nondurables	Basic inorganic chemicals	32512-8
Manufacturing - Nondurables	Industrial gas	32512
Manufacturing - Nondurables	Synthetic dye and pigment	32513
Manufacturing - Nondurables	Other basic inorganic chemical	32518
Manufacturing - Nondurables	Alkalies and chlorine	32518pt.
Manufacturing - Nondurables	Resin, synthetic rubber, and artificial and synthetic fibers and filaments	3252
Manufacturing - Nondurables	Resin and synthetic rubber	32521
Manufacturing - Nondurables	Plastics material and resin	325211
Manufacturing - Nondurables	Synthetic rubber	325212
Manufacturing - Nondurables	Artificial and synthetic fibers and filaments	32522
Manufacturing - Nondurables	Pesticide, fertilizer, and other agricultural chemical	3253
Manufacturing - Nondurables	Pharmaceutical and medicine	3254

Sector	Industry	NAICS Code
Manufacturing - Nondurables	Paints, soaps and toiletries, and other chemical products	3255-9
Manufacturing - Nondurables	Paints and other chemical products	3255,9
Manufacturing - Nondurables	Paint, coating, and adhesive	3255
Manufacturing - Nondurables	Paint and coating	32551
Manufacturing - Nondurables	Soap, cleaning compound, and toilet preparation	3256
Manufacturing - Nondurables	Plastics and rubber products	326
Manufacturing - Nondurables	Plastics products	3261
Manufacturing - Nondurables	Rubber products	3262
Manufacturing - Nondurables	Tire	32621
Manufacturing - Nondurables	Rubber products excl tires	32622,9
Manufacturing - Durables	Wood products	321
Manufacturing - Durables	Sawmills and wood preservation	3211
Manufacturing - Durables	Plywood and misc. wood products	3212,9
Manufacturing - Durables	Veneer, plywood, & engineered wood product	3212
Manufacturing - Durables	Reconstituted wood product	321219
Manufacturing - Durables	Other wood products	3219
Manufacturing - Durables	Millwork	32191
Manufacturing - Durables	Wood container and pallet	32192
Manufacturing - Durables	All other wood products	32199
Manufacturing - Durables	Manufactured home (mobile home)	321991
Manufacturing - Durables	Nonmetallic mineral products	327
Manufacturing - Durables	Clay, lime, gypsum, & miscellaneous nonmetallic mineral products	3271,4,9
Manufacturing - Durables	Clay and miscellaneous nonmetallic mineral products	3271,9
Manufacturing - Durables	Clay product and refractory	3271
Manufacturing - Durables	Pottery, ceramics, and plumbing fixture	32711
Manufacturing - Durables	Clay building material and refractories	32712

Sector	Industry	NAICS Code
Manufacturing - Durables	Other nonmetallic mineral product	3279
Manufacturing - Durables	Lime and gypsum product	3274
Manufacturing - Durables	Glass and glass product	3272
Manufacturing - Durables	Glass container	327213
Manufacturing - Durables	Cement and concrete product	3273
Manufacturing - Durables	Cement	32731
Manufacturing - Durables	Concrete and product	32732-9
Manufacturing - Durables	Primary metals	331
Manufacturing - Durables	Iron and steel products	3311,2
Manufacturing - Durables	Pig iron	3311,2pt.
Manufacturing - Durables	Raw steel	3311,2pt.
Manufacturing - Durables	Coke and products	3311,2pt.
Manufacturing - Durables	Construction steel	3311,2pt.
Manufacturing - Durables	Consumer durable steel	3311,2pt.
Manufacturing - Durables	Can and closure steel	3311,2pt.
Manufacturing - Durables	Equipment steel	3311,2pt.
Manufacturing - Durables	Miscellaneous steel	3311,2pt.
Manufacturing - Durables	Nonferrous metals	3313,4
Manufacturing - Durables	Alumina & aluminum production & processing	3313
Manufacturing - Durables	Primary aluminum production	331313pt.
Manufacturing - Durables	Secondary smelting and alloying of aluminum	331314
Manufacturing - Durables	Misc aluminum materials	331315,8pt.
Manufacturing - Durables	Aluminum extruded product	331318pt.
Manufacturing - Durables	Nonferrous metal (except aluminum) production and processing	3314
Manufacturing - Durables	Nonferrous metal (except aluminum) smelting and refining	33141
Manufacturing - Durables	Primary smelting & refining of copper	33141pt.

Sector	Industry	NAICS Code
Manufacturing - Durables	Primary smelting & refining of nonferrous metal (except copper and aluminum)	33141pt.
Manufacturing - Durables	Foundries	3315
Manufacturing - Durables	Fabricated metal products	332
Manufacturing - Durables	Forging and stamping	3321
Manufacturing - Durables	Cutlery and handtool	3322
Manufacturing - Durables	Architectural and structural metals	3323
Manufacturing - Durables	Hardware	3325
Manufacturing - Durables	Spring and wire product	3326
Manufacturing - Durables	Machine shops; turned product; and screw, nut, and bolt	3327
Manufacturing - Durables	Coating, engraving, heat treating, and allied activities	3328
Manufacturing - Durables	Other fabricated metal products	3329
Manufacturing - Durables	Ball and roller bearing	332991
Manufacturing - Durables	Machinery	333
Manufacturing - Durables	Agriculture, construction, and mining machinery	3331
Manufacturing - Durables	Agricultural implement	33311
Manufacturing - Durables	Farm machinery and equipment	333111
Manufacturing - Durables	Construction machinery	33312
Manufacturing - Durables	Mining and oil and gas field machinery	33313
Manufacturing - Durables	Industrial machinery	3332
Manufacturing - Durables	Commercial and service industry machinery and other general purpose machinery	3333,9
Manufacturing - Durables	HVAC, metalworking, and power transmission machinery	3334-6
Manufacturing - Durables	Ventilation, heating, air-con and commercial refrigeration	3334
Manufacturing - Durables	Metalworking machinery	3335

Sector	Industry	NAICS Code
Manufacturing - Durables	Engine, turbine, and power transmission equipment	3336
Manufacturing - Durables	Computer and electronic products	334
Manufacturing - Durables	Computer and peripheral equipment	3341
Manufacturing - Durables	Communications equipment	3342
Manufacturing - Durables	Audio and video equipment	3343
Manufacturing - Durables	Semiconductors & related electronic components	3344
Manufacturing - Durables	Navigational, measuring, electromedical, and control instruments	3345
Manufacturing - Durables	Electrical equipment, appliance & component	335
Manufacturing - Durables	Household appliance	3352
Manufacturing - Durables	Small electrical appliance	33521
Manufacturing - Durables	Major appliance	33522
Manufacturing - Durables	Electrical equipment except appliances	335XA
Manufacturing - Durables	Electric lighting equipment	3351
Manufacturing - Durables	Electrical equipment	3353
Manufacturing - Durables	Other electrical equipment & component	3359
Manufacturing - Durables	Battery	33591
Manufacturing - Durables	Communication & energy wire & cable	33592
Manufacturing - Durables	Other electrical equipment	33593,9
Manufacturing - Durables	Transportation equipment	336
Manufacturing - Durables	Motor vehicles	3361
Manufacturing - Durables	Automobile and light duty motor vehicle	33611
Manufacturing - Durables	Automobile	336111
Manufacturing - Durables	Light truck and utility vehicle	336112
Manufacturing - Durables	Heavy duty truck	33612
Manufacturing - Durables	Motor vehicle body and trailer	3362
Manufacturing - Durables	Truck trailer	336212
Manufacturing - Durables	Motor home	336213

Sector	Industry	NAICS Code
Manufacturing - Durables	Travel trailer and camper	336214
Manufacturing - Durables	Motor vehicle parts	3363
Manufacturing - Durables	Aerospace products and parts	3364
Manufacturing - Durables	Aircraft and parts	336411-3
Manufacturing - Durables	Railroad eqpt, ships and boats, and other transportation equipment	3365-9
Manufacturing - Durables	Railroad rolling stock	3365
Manufacturing - Durables	Ship and boat building	3366
Manufacturing - Durables	Other transportation equipment	3369
Manufacturing - Durables	Furniture and related products	337
Manufacturing - Durables	Household and institutional furniture and kitchen cabinet	3371
Manufacturing - Durables	Office and other furniture	3372,9
Manufacturing - Durables	Miscellaneous	339
Manufacturing - Durables	Medical equipment and supplies	3391

References

Alessandri, Piergiorgio and Haroon Mumtaz, "Financial regimes and uncertainty shocks," *Journal of Monetary Economics*, 2019, *101*, 31–46.

Berg, Andreas, Renate Meyer, and Jun Yu, "DIC as a model comparison criterion for stochastic volatility models," *Journal of Business and Economic Statistics*, 2004, *22*, 107–120.

Bloom, Nicholas, "The impact of uncertainty shocks," *Econometrica*, 2009, *77* (3), 623–685.

Chan, J.C.C. and I. Jeliazkov, "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 2009, *1(1-2)*, 101–120.

Chib, Siddhartha, "Markov chain Monte Carlo methods: Computation and inference," *Handbook of Econometrics*, 2001, *5*, 3569–3649.

Kroese, Dirk P, Thomas Taimre, and Zdravko I Botev, *Handbook of Monte Carlo methods*, Vol. 706, John Wiley & Sons, 2013.

Leduc, Sylvain and Zheng Liu, "Uncertainty shocks are aggregate demand shocks," *Journal of Monetary Economics*, 2016, *82(C)*, 20–35.

McCracken, Michael W. and Serena Ng, "FRED-MD: A monthly database for macroeconomic research," *Journal of Business and Economic Statistics*, 2016, *34(4)*, 574–589.

Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima, "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 2007, *140(2)*, 425–449.

Sectoral Uncertainty

Efrem Castelnuovo[a] Kerem Tuzcuoglu[b] Luis Uzeda[b]

[a]University of Padova

[b]Bank of Canada

BoC-IFC Satellite Seminar
July 15, 2023

The views expressed in this presentation are ours and do not necessarily reflect the position of the Bank of Canada.

MOTIVATION

- There are countless studies on uncertainty since the Great Recession
  Bloom [2009], Jurado, Ludvigson, and Ng [2015], Ludvigson, Ma, and Ng [2021], Baker, Bloom, and Davis [2016], Carriero, Clark, and Marcellino [2016, 2018], Jo and Sekkel [2019], Mumtaz and Musso [2019]
  - For comprehensive literature reviews see:
    Bloom [2014], Cascaldi-Garcia et al. [2020], Castelnuovo [2022].
- The literature has focused predominantly on **aggregate** measures of uncertainty
- Sectors usually face **additional sector-specific** uncertainty shocks on top of the economy-wide uncertainty,
  - E.g. transportation, retail stores, health care system during Covid-19.
- We know that aggregate uncertainty is countercyclical and contractionary

▶ But what do we know about sectoral uncertainty?
And, how different is it from aggregate uncertainty?

MOTIVATION

- There are countless studies on uncertainty since the Great Recession
  Bloom [2009], Jurado, Ludvigson, and Ng [2015], Ludvigson, Ma, and Ng [2021], Baker, Bloom, and Davis [2016], Carriero, Clark, and Marcellino [2016, 2018], Jo and Sekkel [2019], Mumtaz and Musso [2019]
  - For comprehensive literature reviews see:
    Bloom [2014], Cascaldi-Garcia et al. [2020], Castelnuovo [2022].
- The literature has focused predominantly on **aggregate** measures of uncertainty
- Sectors usually face **additional sector-specific** uncertainty shocks on top of the economy-wide uncertainty,
  - E.g. transportation, retail stores, health care system during Covid-19.
- We know that aggregate uncertainty is countercyclical and contractionary

▶ But what do we know about sectoral uncertainty?
And, how different is it from aggregate uncertainty?

The Model in 1 Slide

A Hierarchical Common Stochastic Volatility Model:

$$y_t = X_t\beta + \overbrace{\Sigma_t^{1/2}\Psi_t^{1/2}e_t}^{u_t}, \qquad e_t \sim \mathcal{N}(0,\ I_N)$$

- y_t is $(N \times 1)$ containing $\mathcal{S}$ sectors/groups.
- **predictable component:** X_t may contain observables and/or latent dynamic factors (possibly with SV).
- **measures of uncertainty:** collected in the stochastic volatility of the unpredictable component $\Sigma_t^{1/2} = \texttt{diag}\left(\exp\left(\frac{h_{1,t}}{2}\right), \cdots, \exp\left(\frac{h_{N,t}}{2}\right)\right)$ and $h_t = \Lambda_c v_t^{common} + \Lambda_s v_t^{sectoral}$ with block exogeneity in Λ_s.
- **capturing the outliers:** $\Psi_t = \texttt{diag}(\Psi_{t,1}, \ldots, \Psi_{t,N})$ is a scaling matrix that collects idiosyncratic volatility and turns u_t into Student-t

Why Sectoral Uncertainty?

- Are sectoral dynamics important? Yes!
  - Great Moderation reflected a decline in the importance of aggregate shocks relative to sector-specific shocks.
    Foerster et al. [2011], Atalay [2017], Garin et al. [2018]
  - Macro tail-risk during recessions are due to sectoral-dominance
    Acemoglu et al. [2012], Acemoglu et al. [2017]
  - Microeconomic shocks may not 'average out' due to, e.g., the nonlinear nature of the disaggregated production structure.
    Baqaee and Farhi [2019]

- What about sectoral uncertainty? Empirical evidence is scant.
  - The effect of sector-level uncertainty on unemployment is more persistent than the aggregate uncertainty (Choi and Loungani [2015])
  - Consumption and investment TFP volatility shocks have different implications for macroeconomic and financial variables. (Segal [2019])
  - Aggregate and sectoral uncertainty based on median forecast error of earnings-per-share ratios. (Ma and Samaniego [2019])

Why Sectoral Uncertainty?

- Are sectoral dynamics important? Yes!
  - Great Moderation reflected a decline in the importance of aggregate shocks relative to sector-specific shocks.
    Foerster et al. [2011], Atalay [2017], Garin et al. [2018]
  - Macro tail-risk during recessions are due to sectoral-dominance
    Acemoglu et al. [2012], Acemoglu et al. [2017]
  - Microeconomic shocks may not 'average out' due to, e.g., the nonlinear nature of the disaggregated production structure.
    Baqaee and Farhi [2019]

- What about sectoral uncertainty? Empirical evidence is scant.
  - The effect of sector-level uncertainty on unemployment is more persistent than the aggregate uncertainty (Choi and Loungani [2015])
  - Consumption and investment TFP volatility shocks have different implications for macroeconomic and financial variables. (Segal [2019])
  - Aggregate and sectoral uncertainty based on median forecast error of earnings-per-share ratios. (Ma and Samaniego [2019])

This Paper

- New empirical framework to estimate sectoral uncertainty
  - Allows for *joint* estimation of common and sector-specific uncertainty
    - Extends work on common-stochastic volatility models
  - Incorporates outlier adjustment

- Measurement
  - Apply our framework to a large dataset of disaggregated industrial production series for the US that covers nearly 200 industries in the **durables** and **non-durables** sectors.

- Sectoral uncertainty and macroeconomic dynamics
  - heterogeneity of common & sectoral uncertainty in terms of evolution and role as drivers of the business cycle
    - common uncertainty: contractionary ('drop-rebound-overshoot')
    - durables uncertainty: contractionary
    - nondurables uncertainty: expansionary!

Data

- U.S. Industrial Production data between 1972Q1 – 2019Q4
- Similar dataset to Foerster, Sarte, and Watson [2011] and Garin, Pries, and Sims [2018]
- 185 Industries grouped into 2 Sectors (Durables and Nondurables) based on the NAICS

Table 1: The IP dataset

Sectors	**# Industries**
Manufacturing - Nondurables	80
Manufacturing - Durables	105

Each series is transformed into standardized growth rates.

$$y_t = \begin{bmatrix} \text{frozen food} \\ \text{bakeries and tortilla} \\ \text{textile furnishings} \\ \text{paper mills} \\ \text{organic chemicals} \\ \vdots \\ \text{wood products} \\ \text{iron products} \\ \text{semiconductors} \\ \text{motor vehicles} \\ \vdots \end{bmatrix}$$

- We will estimate 1 common and 2 sector-specific volatility factors

Decomposing Uncertainty

$$y_t = X_t\beta + \Sigma_t^{1/2}\Psi_t^{1/2}e_t, \qquad \Sigma_t^{1/2} = \texttt{diag}\left(\exp\left(\frac{h_{1,t}}{2}\right), \cdots, \exp\left(\frac{h_{N,t}}{2}\right)\right)$$

- Overall uncertainty is captured by h_t:

$$\begin{bmatrix} h_{1,t} \\ \vdots \\ h_{N,t} \end{bmatrix} = \Lambda_c v_{c,t} + \Lambda_s v_{s,t} \quad \text{and} \quad \begin{bmatrix} v_{c,t} \\ v_{s,t} \end{bmatrix} = \Phi_v \begin{bmatrix} v_{c,t-1} \\ v_{s,t-1} \end{bmatrix} + \eta_{v,t}$$

- The uncertainty in an industry ($h_{i,t}$) is driven by a common ($v_{c,t}$) and a sector-specific ($v_{s,t}$) component.

- $v_{c,t}$ is a factor common to all industries while $v_{s,t}$ collects factors that are common only to industries within the same sector s

- The VAR model for ($v_{c,t}$, $v_{s,t}$) allows for uncertainty spillovers

Volatility Loadings

- Recall: $h_t = \Lambda_c v_{c,t} + \Lambda_s v_{s,t}$.
- Let's write the loadings explicitly for durables & nondurables sectors

$$\Lambda_c = \begin{bmatrix} \lambda_1^c \\ \lambda_2^c \\ \vdots \\ \vdots \\ \lambda_N^c \end{bmatrix} \quad \text{and} \quad \Lambda_s = \begin{bmatrix} \lambda_1^d & 0 \\ \vdots & \vdots \\ \lambda_{N_d}^d & 0 \\ 0 & \lambda_{N_d+1}^{nd} \\ \vdots & \vdots \\ 0 & \lambda_N^{nd} \end{bmatrix}$$

- Log-volatility of an industry:
  $h_{i,t} = \lambda_i^c v_{c,t} + (\mathbf{1}_{\{i \in d\}} \times \lambda_i^d) v_{d,t} + (\mathbf{1}_{\{i \in nd\}} \times \lambda_i^{nd}) v_{nd,t}$
- Identification requires $1 = \lambda_1^c = \lambda_1^d = \lambda_{N_d+1}^{nd}$
- Hierarchical factor structure similar to Kose et al. [2003], Del Negro and Otrok [2007], Moench et al. [2013], and Mumtaz and Musso [2019]
- CCM[2016]: single common factor
  CCM[2018]: macro & financial factors but no common (aggregate) one.

Idiosyncratic Volatility

- Recall: $y_t = X_t\beta + \overbrace{\Sigma_t^{1/2}\Psi_t^{1/2}e_t}^{u_t}$.

$$\Psi_t^{\frac{1}{2}} = \texttt{diag}\left(\psi_{1,t}^{\frac{1}{2}}, \cdots, \psi_{N,t}^{\frac{1}{2}}\right) \; s.t.$$

$$\psi_{i,t} \overset{i.i.d.}{\sim} \mathcal{IG}\left(\frac{\nu_\psi}{2}, \frac{\nu_\psi}{2}\right) \text{ for } i = 1, \cdots, N \text{ and } t = 1, \cdots, T,$$

- $\psi_{i,t}$ is a latent scaling factor that makes the composite error term u_t have fat-tails
- Useful for absorbing industry-specific outliers.
- Parametrization similar to Jacquier et al. [2004] and Chib et al. [2006]

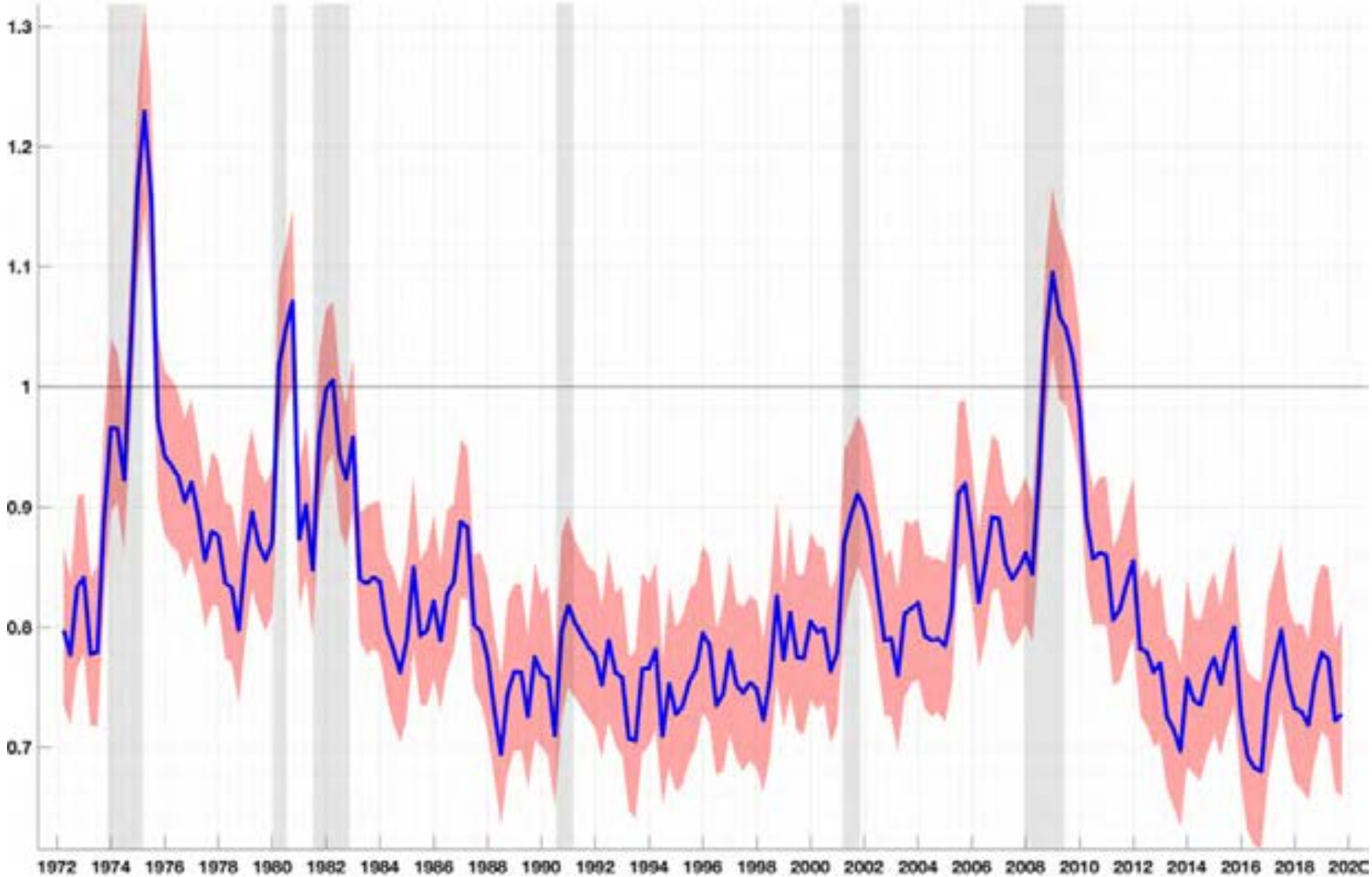
Estimation

- Bayesian methods (15-step Gibbs sampler)
- Standard class of priors
- Level-factors estimated using precision sampling techniques as in Chan and Jeliazkov [2009]
- Volatility-factors estimated combining precision sampling with the auxiliary mixture sampling method from Omori, Chib, Shephard, and Nakajima [2007]
- Algorithm is efficient and reasonably fast (running 10K iterations takes around 30 minutes)
- Mixing properties are quite good with 50K draws + 10K burn-in

Results

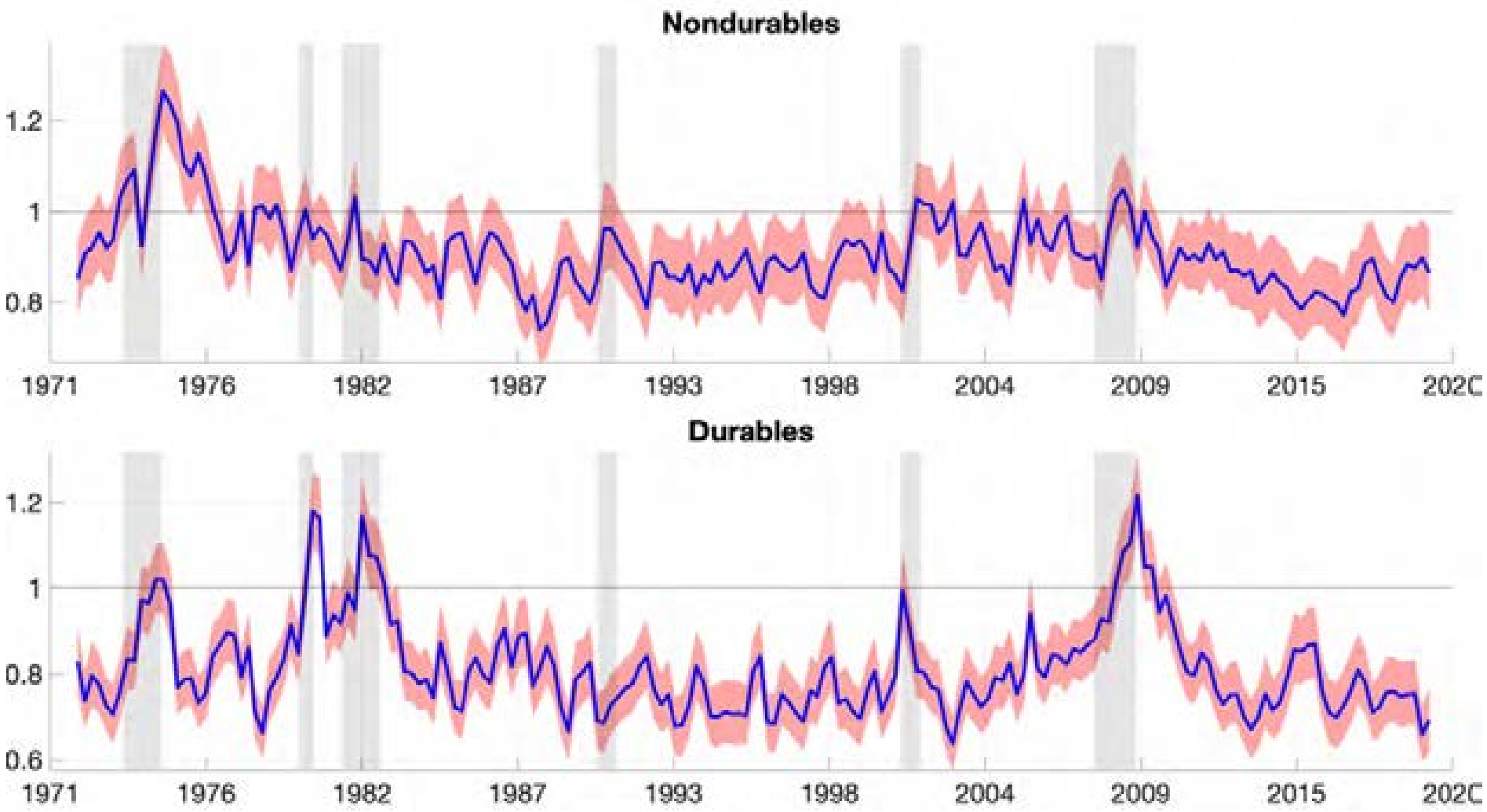
Common Uncertainty

Figure 1: Common uncertainty



Notes: This figure plots the common volatility factor. The solid line denotes the posterior median for $\exp(v_t^c/2)$ together with its equitailed 68% posterior credible interval. Vertical shaded bars correspond to NBER recession dates.

Sectoral Uncertainty



Notes: This figure plots the sector-specific volatility. The solid line denotes the posterior median for $\exp(v_t^s/2)$ together with its equitailed 68% posterior credible interval. Shaded vertical bars correspond to NBER recession dates.

Correlations

Variable	Common	Nondurables	Durables
Real GDP	-0.23	-0.12	-0.19
Industrial Production	-0.30	-0.14	-0.26
Investment	-0.24	-0.12	-0.19
Employment	-0.37	-0.15	-0.38
Real PCE	0.33	0.29	0.20
CPI-Inflation	0.23	0.23	0.12
NBER Recession Indicator	0.55	0.28	0.51
JLN Macroeconomic Uncertainty	0.63	0.32	0.60
LMN Real Uncertainty	0.65	0.39	0.53
LMN Financial Uncertainty	0.36	0.24	0.29
Economic Policy Uncertainty	0.14	0.09	0.11
Monetary Policy Uncertainty	0.07	0.09	0.02
Fiscal Policy Uncertainty	0.13	0.08	0.10
VIX	0.44	0.28	0.35
Common	1		
Nondurables	0.36	1	
Durables	0.43	0.20	1

Effects on Real Economy: Revisiting Jurado, Ludvigson, Ng (2015)

- 12-variable VAR where uncertainty measures are placed last (recursive identification)
- We repeat the exercise twice by replacing **Sectoral Uncertainty** with nondurables and durables uncertainty measures.
- We are particularly interested in the responses of Production and Employment

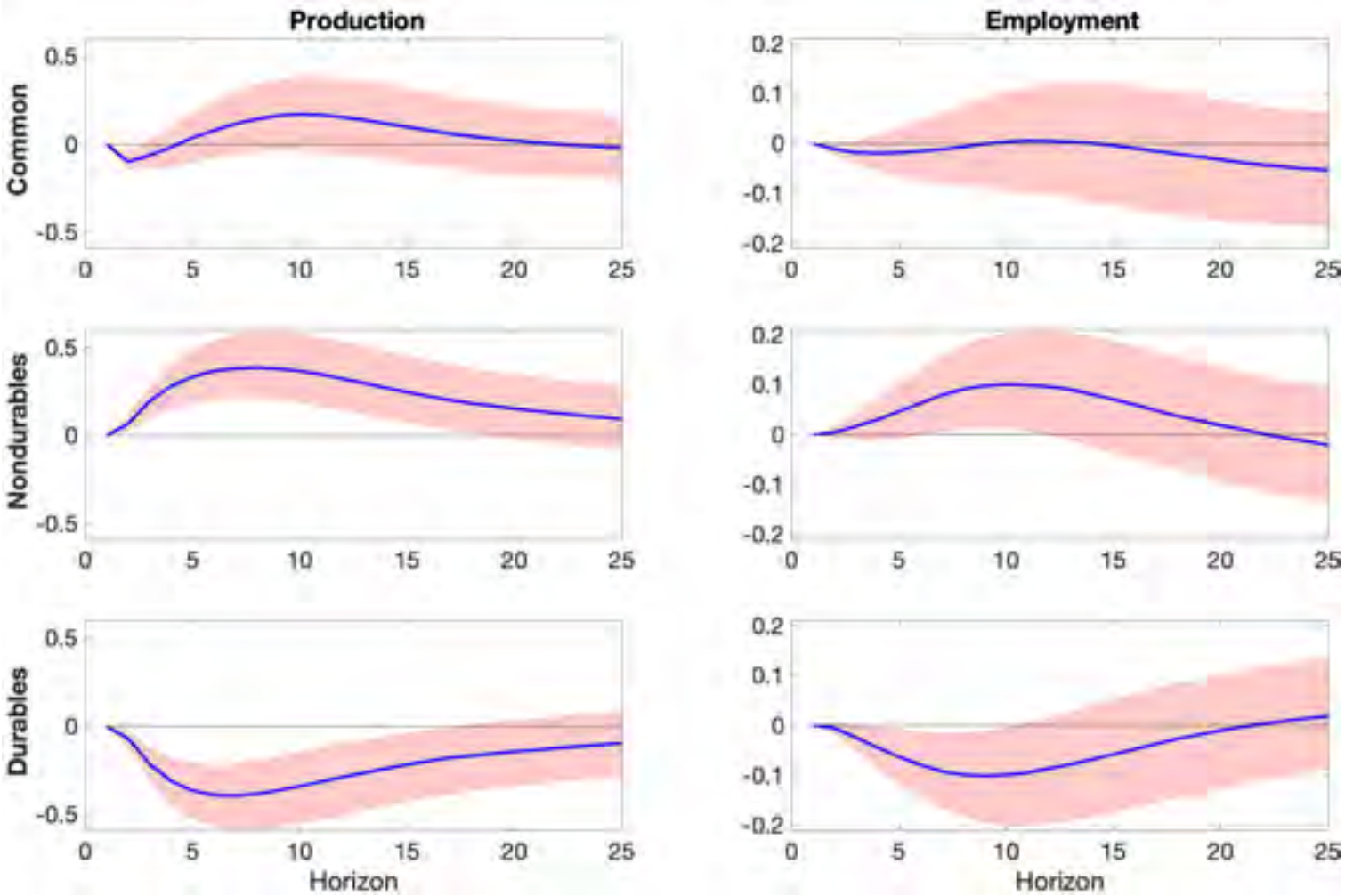
$$\begin{bmatrix} \text{log(real IPI)} \\ \text{log(employment)} \\ \text{log(real consumption)} \\ \text{log(PCE deflator)} \\ \text{log(real new orders)} \\ \text{log(real wage)} \\ \text{hours} \\ \text{federal funds rate} \\ \text{log(S\&P 500 Index)} \\ \text{growth rate of M2} \\ \textbf{Common Uncertainty} \\ \textbf{Sectoral Uncertainty} \end{bmatrix}$$

Takeaway: Not all uncertainty measures are alike!

- Sector-specific uncertainty is not always contractionary.

Responses of Production and Employment

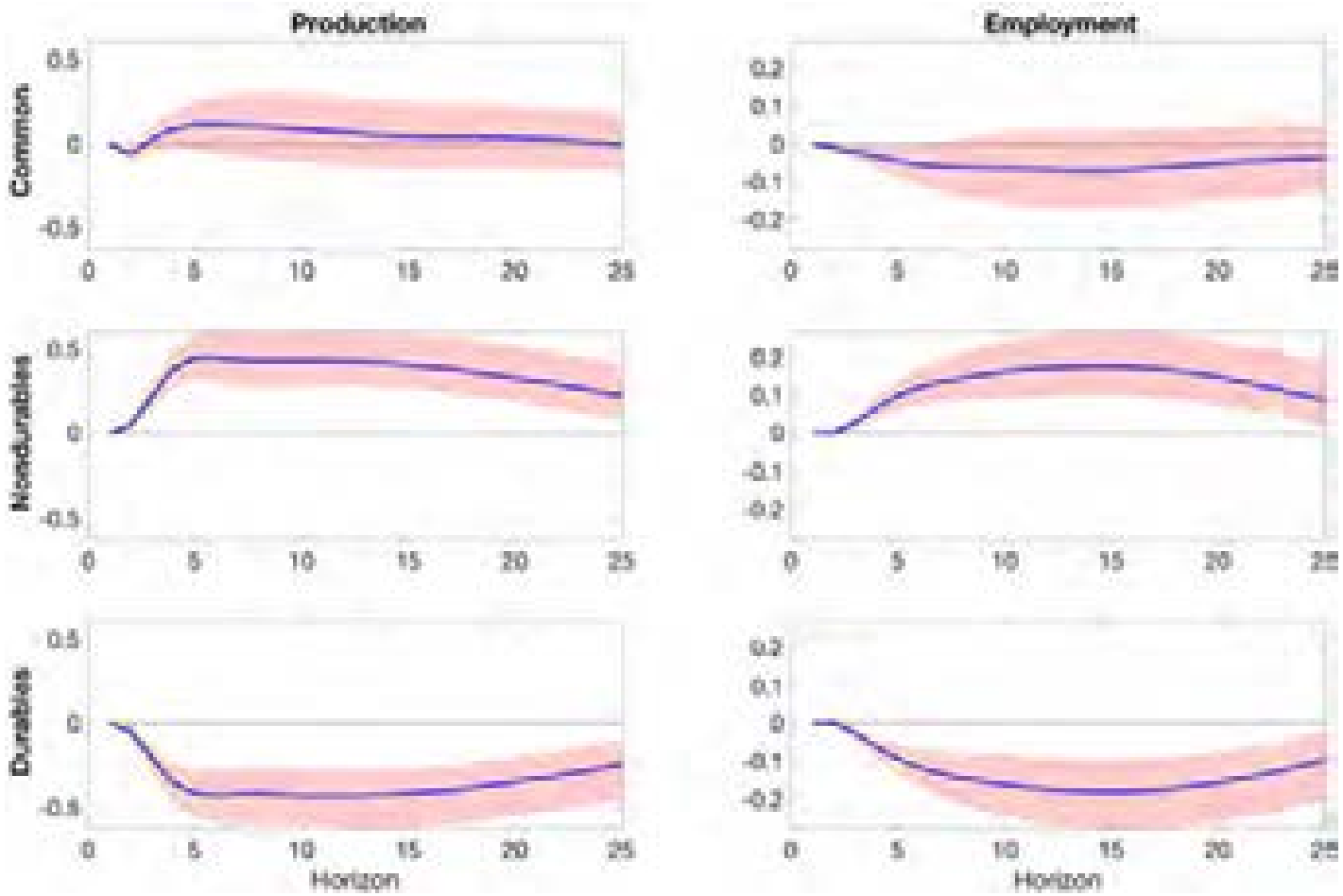
Figure 2: Impulse Responses of Production and Employment to Uncertainty Shocks



Notes: Impulse responses denote percentage point responses to a one standard deviation surprise movement in the uncertainty measures. Solid lines denote posterior medians and the shaded areas represent the 68% equal-tailed posterior credible interval.

Uncertainty During the Great Moderation Sample: 1984Q1–2019Q4

Figure 3: Impulse Responses of Production and Employment to Uncertainty Shocks



Notes: Impulse responses denote percentage point responses to a one standard deviation surprise movement in the uncertainty measures. Solid lines denote posterior medians and the shaded areas represent the 68% equal-tailed posterior credible interval.

FEVD Comparison

- Foerster et al. [2011], Atalay [2017], Garin et al. [2018]: after the Great Moderation, first-moment sectoral shocks become more important as drivers of the business cycle. What about the second-moment sectoral shocks?

Table 2: Forecast Error Variance Decomposition (FEVD) for Uncertainty Shocks to Industrial Production and Employment.

	72Q2–19Q4	**84Q1–19Q4**
Ind. Production: % explained by		
Common	1.4%	1.8%
Nondurables	4.0%	11.5%
Durables	4.1%	11.9%
Employment: % explained by		
Common	1.1%	2.3%
Nondurables	1.4%	6.8%
Durables	1.5%	7.4%

Notes: FEVD results are constructed over a 24-quarter horizon.

Robustness

- Five different structures for X_t
  - All specifications yields very similar uncertainty factors
  - The correlations between the benchmark and different versions ≥ 0.96
  - Smaller set of control variables (i.e. only z_t or only f_t), generally, results in larger uncertainty factors
  - Including SV in level factors do not change (almost) anything for our uncertainty measures

- Normally distributed residuals $u_t = \Sigma_t^{1/2} e_t$ (instead of Student-t)
  - There is now more variation left to be captured by $\Sigma_t^{1/2}$
  - As expected, "new" uncertainty series pick up some of the idiosyncratic volatility that used to be captured by ψ_{it}
  - In general, $v_t^{Gaussian} \geq v_t^{Student\text{-}t}$. But the correlation is around .97

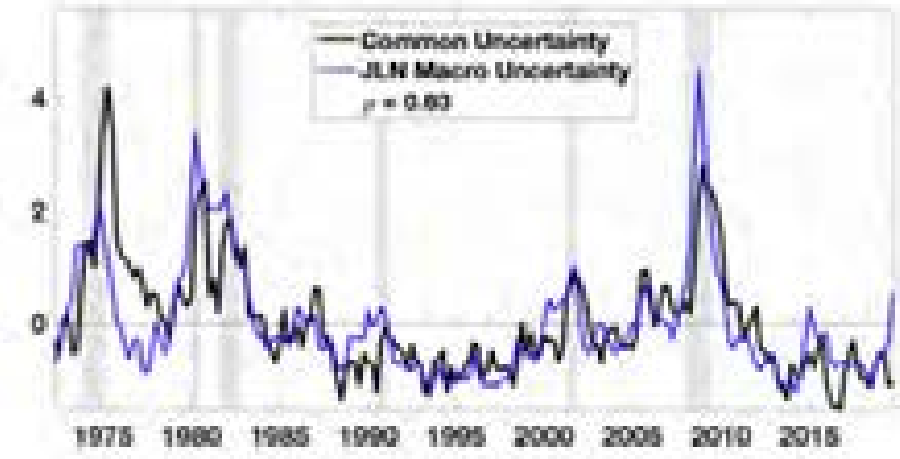
- Different VAR specifications (with 5 and 7 variables)
  - very similar IRFs and FEVDs

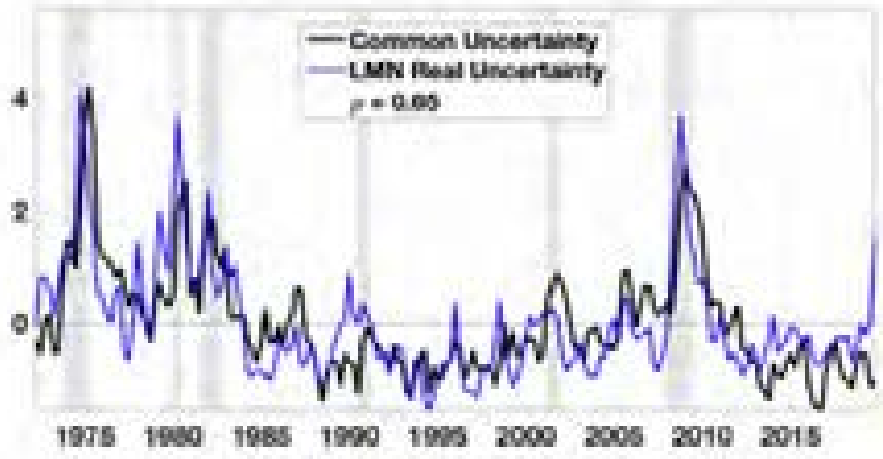
Conclusion

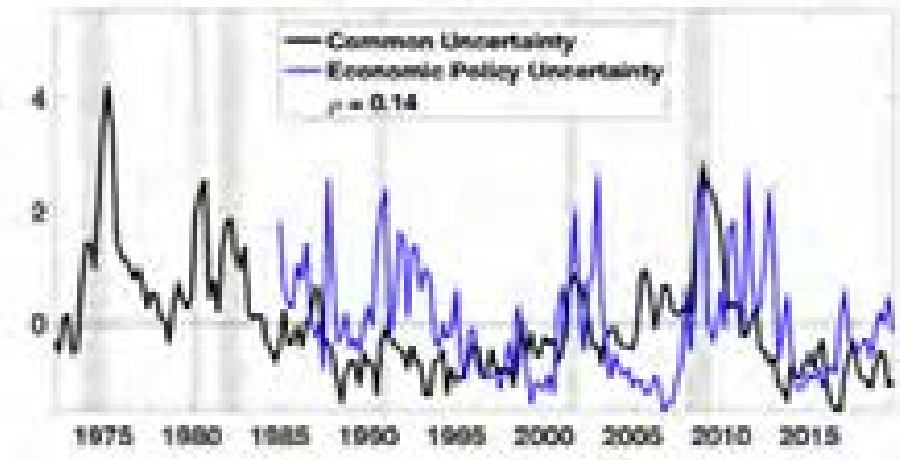
- In this paper, we propose a new framework for jointly extracting common and sector-specific uncertainty.
- We showed that there are significant heterogeneities among different uncertainty measures (both in terms of evolution and effects)
- Business cycle effects typically attributed to aggregate uncertainty may in part be driven by durable goods uncertainty

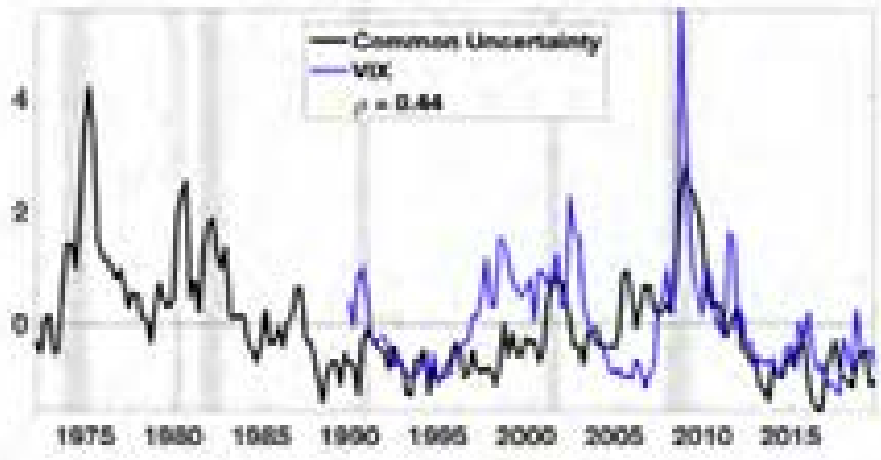
APPENDIX

Common vs Other Measures of Uncertainty









Notes: JLN Macro denotes uncertainty as estimated by Jurado, Ludvigson, and Ng [2015]. LMN Real denotes uncertainty as in Ludvigson, Ma, and Ng [2021]. Economic Policy denotes uncertainty as as in Baker, Bloom, and Davis [2016]. VIX is the Financial Volatility Index produced by the Chicago Board Options Exchange. Measures are normalized for ease of comparison. ρ stands for the unconditional correlation coefficient between two measures of uncertainty. The vertical shaded bars correspond to NBER recession dates.

Specifications For Predictable Component

$$y_t = X_t\beta + u_t$$

Model Identifier	Control Variables in X_t
M1	7 factors obtained from the FRED-MD database 4 IP-specific factors
M2	7 factors obtained from the FRED-MD database 4 IP-specific factors featuring stochastic volatility
M3	7 factors obtained from the FRED-MD database
M4	4 IP-specific factors
M5	4 IP-specific factors featuring stochastic volatility

$$\texttt{Controls for M1:} \begin{cases} X_t = [I_N \otimes z_t' \quad I_N \otimes f_t'], \\ z_t = (z_{1,t}, \cdots, z_{7,t})' & \text{– McCracken and Ng factors,} \\ f_t = (f_{1,t}, \cdots, f_{4,t})' & \text{– IP-specific factors,} \\ f_t = \Phi_f f_{t-1} + \eta_{f,t} & \eta_{f,t} \sim \mathcal{N}\left(0,\ \texttt{diag}\left(\sigma^2_{f_1}, \cdots, \sigma^2_{f_4}\right)\right) \end{cases}$$

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Territorial activity in Chile, by macrozone and economic activity[1]

Pablo García Silva,

Vice-Governor, Central Bank of Chile,
and Chair of the IFC

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

High Level Panel: Insights on the way forward
Granular data: new horizons and challenges for central banks

Pablo García Silva
Deputy Governor, Central Bank of Chile
Chair, Irving Fisher Committee on Central Bank Statistics

July 15th 2023

Introduction

Good evening. I am glad to share with you some remarks on granular data, emphasizing their importance for policy making as well as the challenges it poses.

First, I would like to express my congratulations to all the presenters of today for their insightful contributions. The presentations have provided valuable perspectives on the transformative power of granular data for central banks.

As I mentioned in the morning, there is an immense value on the usage of granular data in our work. They allow us to expand the scope of our analysis and to react promptly making decisions in the context of uncertainty and complexity we are facing today. Now, I would like to share with the audience an example of granular data analysis. In our experience, this has allowed a greater and timely focus on regional linkages of production and demand, a long sought aim for policy makers at the national and regional level.

From a perspective of policy and research, we have sufficient evidence of the value of granular data, justifies the costs/risks of granular data collection. Granular data or high frequency data been helpful during the pandemic or in preparedness for future shocks or in other policy and research work.

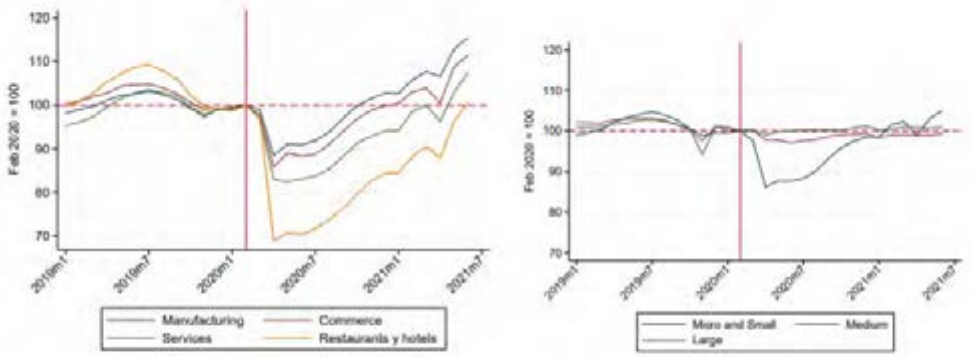
In terms of the value of granular data, I believe that today's conference sessions have confirmed the immense knowledge and insights that we can get from granular data. Moreover, as the horizon of financial technology continues to expand, granular data plays a pivotal role in helping us navigate and regulate this rapidly evolving frontier. This knowledge helps in turn central banks to formulate more effective policies as they can be better targeted to specific sectors or groups of firms, as was

evident during the recent Covid-19 pandemic. Let me share with you some empirical examples that demonstrate how granular data have helped us during the pandemic[1][2].

The Covid-19 crisis had severe affected firms' performance, including firms' sales, investment plans, connections with suppliers, and financial decisions, among other aspects. Many companies were even forced to close temporarily or permanently. These impacts translated into a strong shock to the employment and income of their workers. Therefore, firms are key agents in understanding the adjustment of the economy to the Covid-19 crisis. Because of their relevance in the transmission and amplification of the shock, the mitigation and support policies deployed since the beginning of the crisis were mainly implemented through firms.

Some of the key messages that Covid-19 crisis left us were that it severely affected firms' performance. However, a substantial heterogeneity was found in this adjustment, with more adverse effects centered on micro and small firms, and on industries that reliant more on social contact, such as services, and restaurants and hotels.

Figure 4: Number of firms by sector and size



(a) By sector (b) By size

Note: Number of firms with positive sales every period, normalized to 100 in February, 2020. Seasonally adjusted. In panel (a), the services sector includes personal services, and business services. In panel (b), the classification of firms by size considers annual real sales, following the National Accounts guidelines. The thresholds are based on yearly sales expressed in a unit of account indexed to inflation, called *unidad de fomento* (UF). Micro firms: less than 2,400 UF. Small firms: 2,400-25,000 UF. Medium firms: 25,000-100,000 UF. Large firms: more than 100,000 UF. Based on inflation and market exchange rates at the time of writing, 25,000 UF are approximately equivalent to USD 950,000. Source: Monthly tax form F29 and authors' own calculations.

Source: Albagli et. al. (June 2023). Page 12.

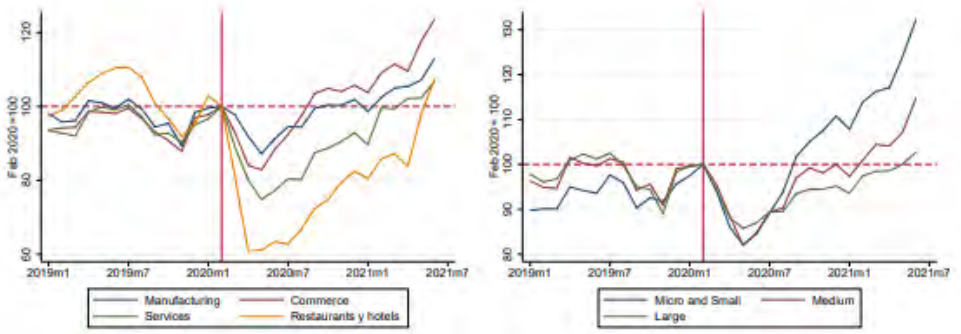
For instance, in Chile, even though the number of firms declined in all industries, the number of restaurants and hotels firms declined of nearly 40%. By contrast, in services industry, the number of firms declined nearly 20%, whereas in manufacturing and commerce, the decline was about 10%. The latter two industries have led the recovery in the number of firms, while restaurants and hotels, and services, remained below their pre-pandemic level until 2021. By firms' size, the decline in the

[1] Albagli, E., Fernández, A., Guerra-Salas, J., Huneeus, F. & Muñoz, P. (June 2023): "Anatomy of Firms' Margins of Adjustment: Evidence from the COVID Pandemic". Central Bank of Chile. DTBC N°981.

[2] *Las Empresas Chilenas Durante la Crisis del Covid-19* (March 2023). *Serie Económica y Financiera BCCh*.

number of firms was completely driven by micro and small firms (firms with sales of less than approximately USD 950,000 in the previous year), with the number of medium and large firms hovering slightly above the pre-pandemic level throughout the sample. The evolution of the number of micro and small, medium, and large firms during the pandemic shock differs from the episode of social unrest that Chile experienced in October 2019, when violent protests broke across the country. In that episode, the number of firms of all sizes declined. At the same time, we observed a large decline and a swift recovery in the number of firms reporting sales, as well as in sales themselves.

Figure 6: Sales: Heterogeneity by sector and size



(a) Sector (b) Size

Note: Seasonally adjusted real sales, normalized at 100 in February 2020. The service sector includes personal services and business services. For details on the classification of firms by size, see the note to figure 4. Sources: Electronic invoice; and authors' own calculations.

Source: Albagli et. al. (June 2023). Page 13.

The evolution of total sales masks substantial heterogeneity at the firm level. The behavior of sales is qualitatively similar to that of the number of firms (panel a in figure 4): sales in restaurants and hotels declined dramatically and remain below their pre-pandemic level up to mid 2021. Sales in services also declined substantially and remained below their pre-pandemic level throughout 2020. Sales in manufacturing and commerce, on the other hand, display swift and strong recoveries. By firms' size micro and small, and medium-sized firms were the hardest hit by the COVID shock, with sales dropping about 20%. However, micro and small, and medium-sized firms also experienced the strongest recovery, with sales reaching pre-pandemic levels in the second half of 2020. Large firms display less fluctuation, with sales declining and expanding less than micro and small, and medium-sized firms in the contractionary and recovery phases. It is important to note that results on firm size are not an artifact of industry effects, e.g., that most restaurants and hotels are small firms and, since this industry was badly hit, it drives the decline we see for small firms. It has been verified that micro and small firms are the most affected within each of the industries.

Employment also experienced a large decline, and its recovery seemed to be lagging that of sales, following a U shape rather than a V shape. The employment protection program, a furlough scheme funded by the government, it helped mitigate this effect.

The decline in employment was larger and more persistent than that of sales, so that by the end of 2020, it remained below its pre-pandemic level. The results in terms of heterogeneity within firm's industry and size was analogous to those previously explored for the case of sales.

In April 2020 the Employment Protection Law (LPE for its acronym in Spanish) was approved by Congress and became another landmark program in the set of policies aimed at supporting firms and, in particular, the relationships they have formed over time with their workers. LPE gave firms important room to maneuver and scale back production without incurring in costly layoffs, while giving them also the chance to scale up quickly and without incurring in hiring costs once the economy recovered. Throughout the program, nearly 120,000 firms enrolled at least one worker. This number represents about 45% of firms that had at least one worker in February 2020. Importantly, LPE contributed to mitigating the decline in aggregate employment due to the COVID shock. In fact, in its absence, employment would have dropped more than twice as it did relative to the pre-pandemic level.

Firm access to domestic bank credit increased during the COVID pandemic. This countercyclicality of credit marks a change with respect to previous crises in Chile, when credit contracted.

On March 16, 2020, two weeks after the first COVID case was identified in Chile, the Central Bank of Chile announced it was lowering its monetary policy rate to 50 bp, which it considered its effective lower bound. Importantly, a separate set of unconventional policies to counteract the economic effects of COVID were also announced. At the core of these measures was a new credit facility (FCIC by its acronym in Spanish) aimed at providing medium-term liquidity to commercial banks at very low rates for up to 4 years, conditional on banks providing loans to small and medium-sized firms. Throughout the COVID pandemic, FCIC provided nearly 40 billion U.S. dollars in loans to banks. A few weeks after this, in April 2020, a complementary policy to FCIC was launched by the Chilean government through FOGAPE-COVID, a state-backed fund that would provide sovereign guarantees of up to 85% of commercial bank loans to firms. The combination of these two credit support policies provided resources and incentives for banks to lend to firms affected by the COVID shock.

There was widespread access to this program. By the end of 2020, nearly 250.000 firms (40% of firms reporting positive sales in February 2020) obtained at least one loan through this program. Importantly, access was equally strong among firms that were performing relatively well and those that were being highly impacted by the crisis. Loans were given relatively more to firms in commerce and manufacturing, and mostly to micro and small, incumbent firms.

The countercyclicality of bank credit in the COVID crisis marks a stark contrast with two previous crises—the global financial crisis of the late 2000s and the Asian financial crisis of the late 1990s, when domestic bank credit decelerated. Crucially, the countercyclical response of bank credit was driven by these credit support policies.

Conclusions and challenges

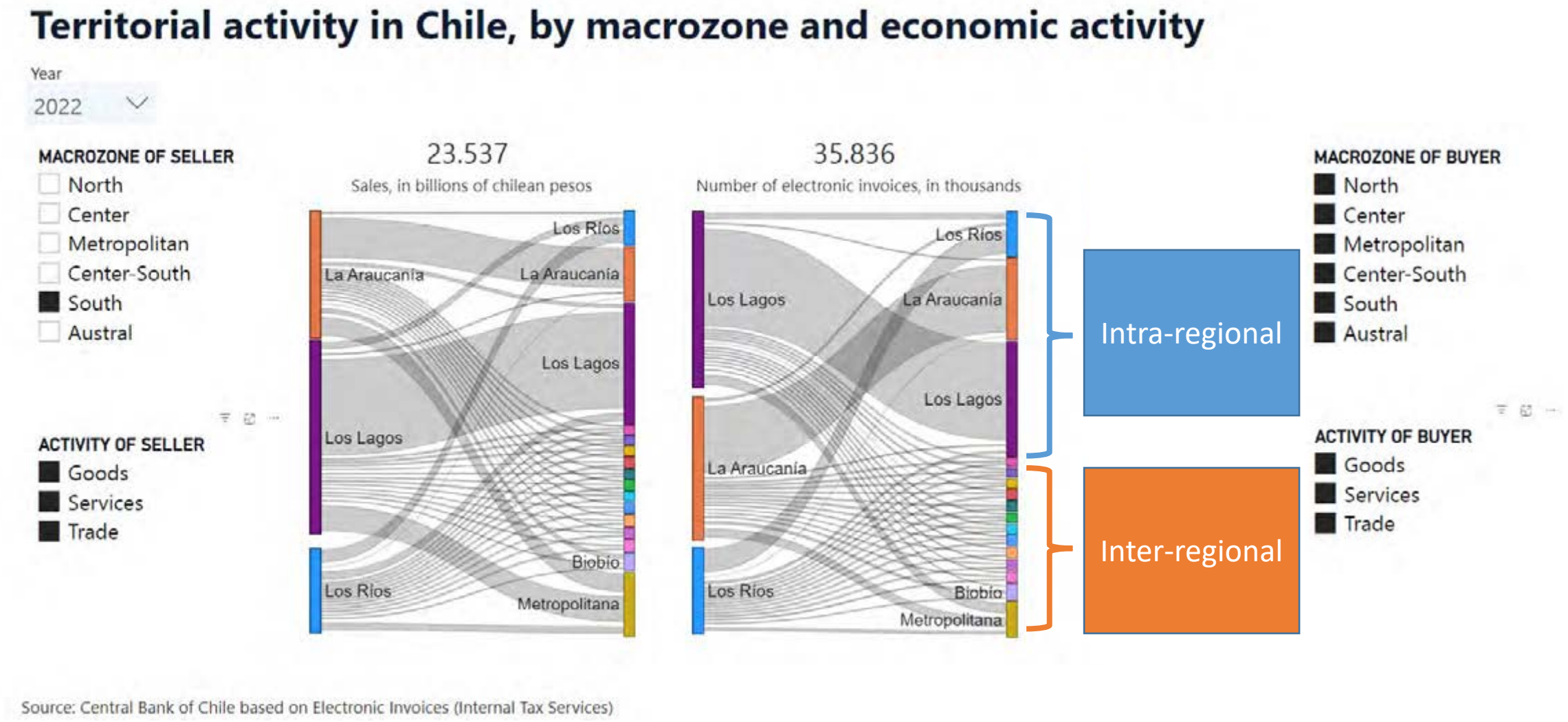
The analysis of the experience of firms in Chile during the Covid-19 crisis was made possible in large part by access to data from administrative records on their performance, in dimensions such as sales, investment, employment, relationship with the banking system, foreign trade, among others. The use of this anonymized microdata for analysis purposes is part of agreements that the Bank has signed with institutions such as the tax authority, the Financial Regulatory Commission and the pension regulator, among others. Promoting this type of agreements is quite important in order to be able to use microdata in the analysis. A robust data governance strategy, which guarantees a secure data handling, storage, and processing, in conjunction with cybersecurity and data privacy policies, becomes a key objective.

From my perspective, the advantages clearly outweigh the drawbacks. I don't imagine a world where policymaker prefer to have less information while taking decisions. Rather than viewing this as a novel challenge, I see it as a golden opportunity to set an institutional standard for excellence in this field.

Looking towards the future, I think that the first key area of focus in the short term is the strengthening of data privacy and security, so we can be sure that when we handle the extensive data at our disposal, we do it correctly and responsibly. However, achieving this is not simply a matter of implementing more stringent protocols or upgrading our technological infrastructure, the heart of the issue lies in cultivating a robust data compromise culture within our people in our central banks. It is about reshaping perceptions, so our team members view this challenge not as an exercise of bureaucracy, but as a mission-critical commitment to our institution's integrity and to the individuals we serve.

A second immediate focus is to use this data in a smart way. In our modern era, where we are bombarded with relentless stimulus from countless sources, often spending entire days behind screens, and navigating through a sea of disinformation and fake news. It's then imperative that we exercise discernment in the information we generate and disseminate to our authorities, and to the public. In the face of this informational overload, our responsibility is to generate the most meaningful and accurate data, providing our audience with reliable, insightful content that truly adds value and upholds the credibility of our institution.

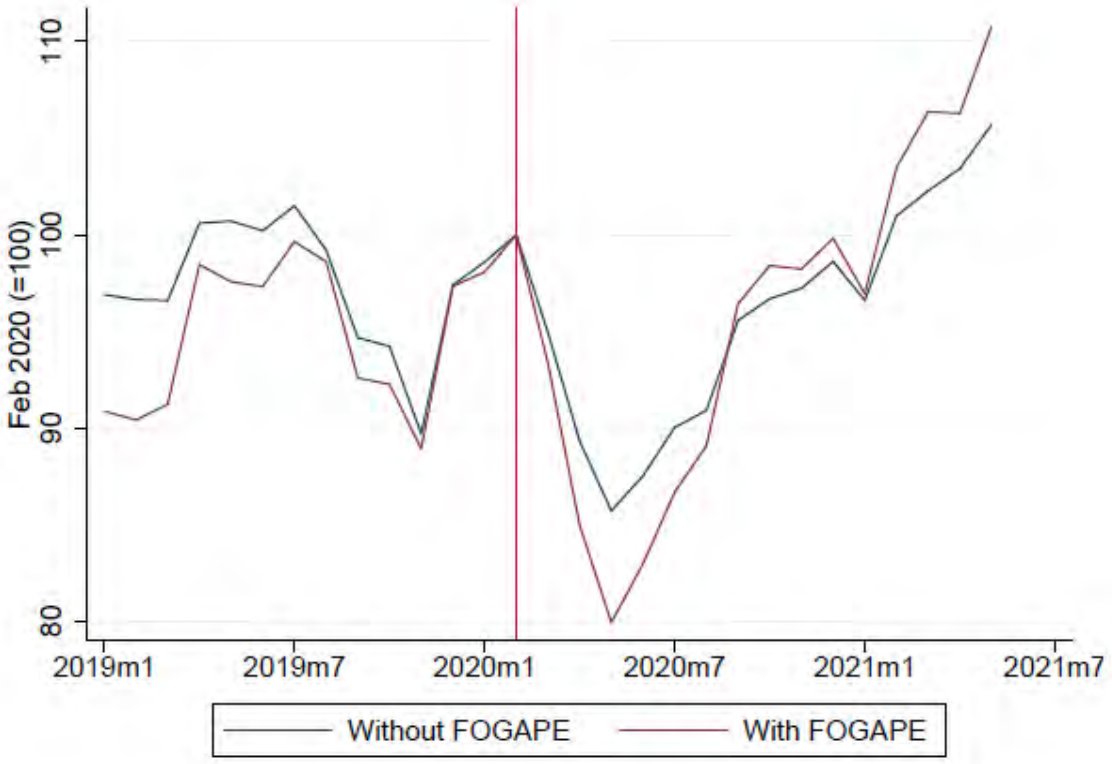
Electronic invoicing allows timely construction of regional/provincial GDP, and economic interlinkages at the sectoral and regional level.



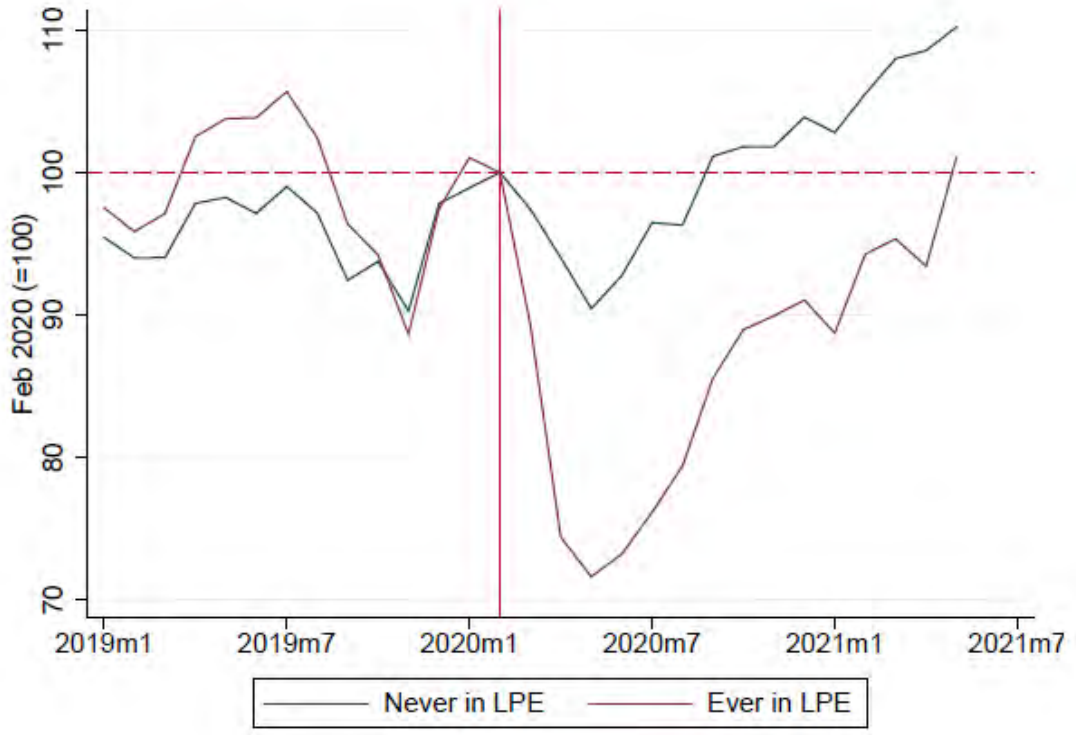
Source: Central Bank of Chile based on Electronic Invoices (Internal Tax Services)

Merged granular data on sales, employment and finance allowed identification of policy impacts during COVID19.

Loan Guarantees (**FOGAPE**) / Employment Protection (**LPE**)



(a) FOGAPE-COVID and sales



(a) LPE and sales

Source: Albagli et. al. (June 2023).

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Classification of participants in the foreign exchange market [1]

Amir Khatib,
Bank of Israel

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Classification of Participants in the Foreign Exchange Market

Amir Khatib[1]

Abstract

This Article presents an analysis of granular data in the Bank of Israel's Big Data database of foreign exchange market transactions. The aim of this study is to expand the existing information on this market by predicting a sector and subsector of new unidentified entities.

Supervised classification algorithms, such as Support Vector Machines (SVM), Logistic regression, and Random Forest, are used to identify real sector market participants as exporters and importers, as well as to allocate unknown foreign market participants to either the financial or nonfinancial sector.

The results of this research demonstrate the benefit of analyzing granular data and of the ability of the chosen algorithms to accurately predict sectors and subsectors of unidentified new entities. The findings of this study allow the Bank of Israel to expand information about entities in the foreign exchange market in order to monitor market activity.

[1] Bank of Israel, amir.khatib@boi.org.il

INTRODUCTION

Similar to other countries, the foreign exchange (forex) market in Israel is the largest financial market in terms of trading turnover. In Israel, the daily average turnover in this market is about US$11 billion. This is significantly greater than turnover in the equities market (US$0.4 billion) or in the bond market (US$0.9 billion). Moreover, the offshore shekel foreign market is almost twice the shekel turnover in Israel.

Similar to many other small open economies, the forex market in Israel plays a major role in the complex effects of monetary policy decisions. Its significance is due to two main factors: first, the relatively high pass-through level (0.25) from the exchange rate to the Consumer Price Index; and second, the deep effect of changes in the exchange rate on the tradable segment of the real economy. Fluctuations in the exchange rate can have significant implications for export-oriented industries, importers, and sectors that depend on the trade of others. Understanding the complex interrelationships between the exchange rate and the tradable segment is essential for policy makers wishing to cultivate stability and sustainable economic growth. (Abir, 2020)

The main participants in the Israeli foreign exchange market are the business sector (exporters and importers), institutional investors (pension funds, mutual funds, and insurance companies), and foreign institutions. Understanding the dynamics of these sectors is essential in understanding the implications of monetary policy decisions. Real-time information on the microstructure of the market and the flow of transactions provides input that is important for policymakers' considerations.

Recognizing the importance of this information, the Bank of Israel initiated reporting requirements for transactions in the forex and shekel interest rate markets in June 2016. The financial intermediaries are required to provide daily reports at the transaction level, which enable the central bank to monitor and analyze market participants' behavior by various sectors and subsectors, and even by individual participants. An in-depth understanding of market flows provides an infrastructure for decision-making with regard to monetary policy and improving macroprudential tools.

Each sector has its own unique motives and different sensitivities to changes in the exchange rate. Therefore, by harnessing the power of advanced machine learning and supervised algorithms on the reporting database, we can better understand the foreign exchange market's dynamics by expanding the sector identification of the existing variety of participants in the market.

This study identifies the sector of unidentified market participants through a methodical analysis of the database of market participants identified by sector. In particular, we expand the identification of participants from the business sector into importers and exporters, and foreign institution participants into financial and nonfinancial sectors.

Recognizing the fact that each sector operates with unique objectives and demonstrates varying sensitivities to changes in the exchange rate, the classification of participants becomes a central factor in understanding the complex dynamics of this market. Researchers and market analysts can gain deeper insights regarding behavior, risk, and the market's potential effects related to various sectors.

The findings of this study will provide an improved basis for market analysis, which will enable policy makers to make more informed decisions and to response efficiently to changes in the activity of the market's segments and the dynamics between them.

Description of the database

The database managed by the Bank of Israel's Information and Statistics Department contains data obtained pursuant to the reporting order implemented in June 2016. Within this framework, roughly 50 financial intermediaries from Israel and abroad report daily data on foreign exchange transactions involving the shekel. The reports include detailed information on the features of the transaction, such as time of transaction (down to the second), instrument, value amount, transaction rate, and details of the parties to the transaction. The information is organized and processed into trading volumes, conversions, and foreign exchange exposures by various sectors, which allows for a comprehensive analysis of developments in the sectors of the foreign exchange market.

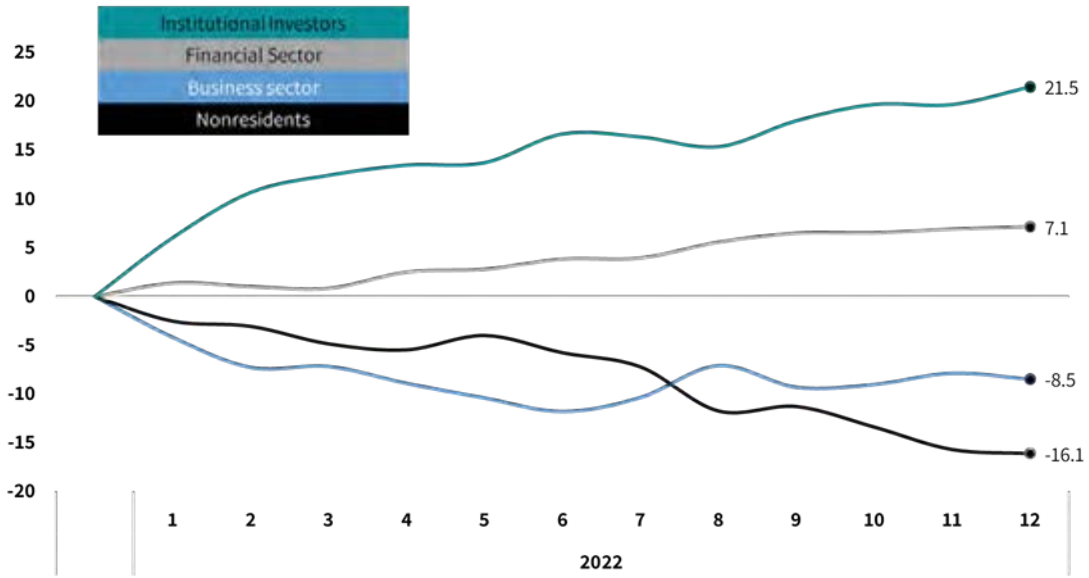
The Bank of Israel's Information and Statistics Department uses a meticulous data processing methodology in order to ensure the reliability and consistency of the information. The data undergo thorough controls, including logical and statistical tests, in order to verify their precision. The data serve to create various reports and products that offer important insights regarding the foreign exchange market. These reports are used as essential tools for decision makers at the central bank to monitor the market.

The data on transactions come from three main sources. Domestic banking corporations began reporting on their transactions in 2008, and foreign financial institutions and domestic financial intermediaries began reporting on their transactions in 2017. The data are obtained on a daily basis, and provide details of all transactions made during the previous day.

The database is comprised of about 50 reporting entities and tens of thousands of active participants in the OTC market. With millions of records received each year, the database contains a significant volume of trading activity. In order to ensure consistency and international standardization, the reporting format was adapted as much as possible to the guidelines set by the ISDA.

The Bank of Israel uses the forex database to produce various products, such as net cumulative foreign exchange purchases by main sector (Figure 1), daily trading volume by type of instrument, implied standard deviations in shekel/dollar options, and more. These products offer important insights regarding market trends, and help decision makers formulate efficient policy.

Figure 1: net cumulative foreign exchange purchases by main sector



Methodology

Data preparation

The data in the Bank of Israel's database undergo built-in logical and statistical controls in order to ensure precision and reliability. These controls are intended to identify any error or inconsistency, and to provide feedback to reporting entities.

Therefore, for the purpose of the analysis in this paper, the required data preparation is minimal. Thanks to the meticulous data preparation tool used by the Bank of Israel, the database for this analysis is already reliable and precise. This allows us to focus mainly on implementing controlled algorithms and content engineering techniques in order to extract valuable insights and efficiently classify forex market participants.

This section outlines the steps taken to prepare the database for analysis.

Selecting relevant features

We extract a group of key variables from the forex database, which will provide us with important insights regarding participants' behavior. These characteristics mainly include:

- Type of instrument: specific type of financial instrument involved in each transaction.
- Currency: The pair of currencies involved in the transaction.
- Financial intermediary: The reporting entity involved in the transaction.
- Transaction period: The duration during which each transaction took place.
- Value amount: Presents the value amount of the transaction.
- Purchase/Sale indicator: If the transaction includes the purchase or sale of shekels.
- Transaction rate: Reflects the exchange rate at which the transaction was made.
- The prevailing exchange rate at the time of the transaction.

One-hot encoding of categorical features

In order to efficiently integrate categorical features such as type of product, currency, and financial intermediary, our analysis implements one-hot encoding. This process transforms categorical variables into binary vectors, and enables the machine learning algorithms to precisely interpret them. A binary variable is assigned to each category in a feature, and a separate column is created for each unique category.

Calculating final features and grouping by market participant level

Following the selection of the relevant features and the start of one-hot encoding, we gather the data at the transaction level to the market participant level. This includes a calculation of various indices and grouping the data based on the unique identifiers of each participant. The compilation methods may include total transaction amounts, calculation of the average exchange rates, or deriving participant-specific indicators.

Transformation and Standardization

As part of the data preparation process, logarithmic transformation and standardization of features are employed to ensure optimal performance of the models. Transformation involves modifying the distribution or representation of the features to meet certain assumptions or improve interpretability. standardization, on the other hand, is applied to ensure that all features are on a similar scale and have comparable ranges. This is particularly important for models that are sensitive to the scale of features, such as SVM and KNN. Transformation and scaling techniques

can enhance model performance, improve convergence, and prevent certain features from dominating the modelling process.

Feature engineering

Combining our in-depth knowledge of the forex market with insights from discussions with brokers and market participants over the years, we built additional features in order to further enrich the data array, help distinguish between the various sectors, and improve the models' proper forecasting.

For each of the following features, we first present the insight that we know from the market and the need to build it:

1. Median by net forex transaction days

Importers generally engage in net forex purchases, while exporters tend to sell.

In order to capture this behavior, we calculate the net foreign exchange transactions for each market participant over a defined period. By accumulating net daily buy/sell positions, we calculate a variable that is the median representing the main tendency of these transactions. This variable provides important insights regarding the overall trading activity of importers and exporters, and sheds light on their behavior and preferences.

2. Index of activity in anomalous depreciation

Exporters frequently exploit periods in which the shekel is depreciated in order to sell large volumes of foreign exchange. In contrast, importers tend to make currency transactions based on demand, irrespective of the exchange rate.

In order to capture this behavior, we calculate an index of activity under anomalous depreciation. This index is calculated by comparing the current shekel exchange rate with the moving monthly average of the exchange rate.

On days when there is an anomalous gap between the exchange rate level at the time of the transaction and the moving average in the previous month (significant depreciation of the shekel), we calculate the net total transactions (purchases/sales) of each entity, and divide that by the overall trading volume. This variable provides an insight regard the extent to which entities exploit the depreciation in order to make larger net transactions relative to their overall trading activity.

By integrating this variable, we can identify market participants that show a greater tendency to exploit the depreciation and make significant net transactions during such periods.

3. Spot foreign exchange purchase ratio

Importers have a greater tendency to purchase foreign exchange through spot transactions for current trade activity.

This variable presents the ratio between foreign exchange spot purchases and the total number of spot transactions for each market participant. Using this ratio, we can quantify the share of spot transactions that included foreign exchange purchases. Higher values of this ratio indicate a greater tendency to purchase foreign exchange via spot transactions.

These features give us a more comprehensive understanding of participants' behavior and enable us to more precisely classify entities in the forex market. By capturing the unique transaction patterns of exporters and importers, these features contribute to a deeper understanding of participants' strategies, approaches to risk management, and their potential effects on market dynamics.

4. Activity ratio for market participation

Financial participants are active in the market on more trading days

This variable measures the share of days on which the participant was active, relative to the total number of business days. It provides valuable information regarding the participant's consistent involvement and presence in the market over time.

In order to calculate this activity ratio, we first determine the total number of business days for each participant. This is calculated by taking the participant's first transaction recorded in the database and counting the number of business days from that point onward.

We then calculate the number of days on which the entity was active, by dividing the number of days on which the entity made transactions by the total number of business days. This ratio represents active participation in the market by the entity's rate within the given timeframe.

Additional Features

5. Average transaction period

This presents the time period between the transaction time and the agreed payment time. By averaging these repayment periods over all of the participant's transactions in the market, we derive that participant's average transaction period.

6. Forward foreign exchange purchase ratio

This variable presents the ratio of forward forex purchases to the overall number of forward transactions for each entity. This ratio is calculated by dividing the number of forward forex purchases by the total number of forward transactions.

Feature selection

In order to select the most relevant features from the first set of 31 features, we used a recursive feature elimination (RFE) used by Zhang, C., Li, Y., Yu, Z., & Tian, F. (2016) on the training data. The feature selection technique removes features in an iterative manner by building a model on the other features and evaluating their importance based on the model's precision.

For the purpose of our evaluation, we used a random forest model that is known for its ability to efficiently deal with categorical and numeric features. In order to ensure robustness, we carried out a 10-fold cross-validation repeated 5 times to evaluate the model's precision. This enabled us to obtain reliability evaluations of the model's performance.

```
Recursive feature selection

Outer resampling method: Cross-Validated (10 fold, repeated 5 times)

Resampling performance over subset size:

 Variables Accuracy  Kappa AccuracySD KappaSD Selected
         1   0.8976 0.7844    0.06068  0.1285
         2   0.8863 0.7612    0.06040  0.1261
         3   0.9045 0.7995    0.06122  0.1288
         4   0.9037 0.7980    0.06121  0.1285
         5   0.9107 0.8129    0.06783  0.1429
         6   0.9168 0.8259    0.06496  0.1360
         7   0.9194 0.8311    0.06448  0.1363
         8   0.9220 0.8364    0.06205  0.1309        *
         9   0.9185 0.8293    0.06429  0.1350
        10   0.9220 0.8364    0.06267  0.1322
        11   0.9186 0.8288    0.06305  0.1334
        31   0.9194 0.8308    0.06325  0.1331
```

In addition to the RFE, we used two filtering criteria to sharpen the feature selection process. First, we filtered features based on variance inflation factor (VIF) values, eliminating each feature with a VIF of greater than 5. This stage helps handle problems of multi-colinearity that may be due to highly-correlated predictions.

Second, we weighted the Pearson correlation between predictions and removed feature pairs with a correlation coefficient of greater than 0.7. This further ensures that the selected features are not too excessive or highly correlated with each other.

In conclusion, by using RFE on classification of participants from the business sector into importers and exporters, we identified a subgroup of 8 features that generated the highest average accuracy on the validations folds data through cross-validation within the training set and that met the selection criteria. However, one feature was not included due to a VIF value greater than 5. We thus obtained a final selection of 7 features.

Same procedure applied on classification of Nonresidents to either financial or Non-financial sector, we obtained a final selection of 12 features.

Training, parameter tuning, and model selection

At this stage, we trained and tuned various supervised classification algorithms using the training set, which constituted 70 percent of the available data. The main objective was to identify the best hyperparameters for each model. We obtained this by using 5-folds cross-validation (CV).

The following algorithms were selected for training and tuning: random forest, support vector machines (SVM), logistic regression, K-nearest neighbors (KNN), and linear discriminant analysis (LDA). Using the CV, we methodically evaluated various combinations of hyperparameters for each algorithm in order to determine the optimal form that would generate the highest performances.

After obtaining the best hyperparameters for each model, we continued to evaluate their performances on a separate validation set that included 20 percent of the data. This validation set enabled us to choose the model that demonstrated the most promising results in terms of precision. The aim was to choose the model that showed the best performance on unobserved data.

In order to ensure the stability and inclusion of the chosen model, we subjected it further to the array of tests that included the remaining 10 percent of the data. By evaluating the model's performances in this testing array, we can evaluate its ability to make precise predictions on completely unobserved data. This step provided a decisive index of the model's stability and implementation potential in the real world.

The combination of CV and validation set evaluation ensured that we chose a model that not only demonstrated excellent performances during the testing, but also presented good performances on unobserved data, thereby providing us with certainty of its classification abilities.

Results

The analysis on importers and exporters led to the selection of K-nearest neighbors (KNN) model as the most appropriate for classification within the business sector. The model was chosen because it had the highest precision in the validation set and achieved 0.938 accuracy. In order to validate the model's performance and avoid over-fitting, the accuracy was also on the training set, which produced a similar precision score of 0.935.

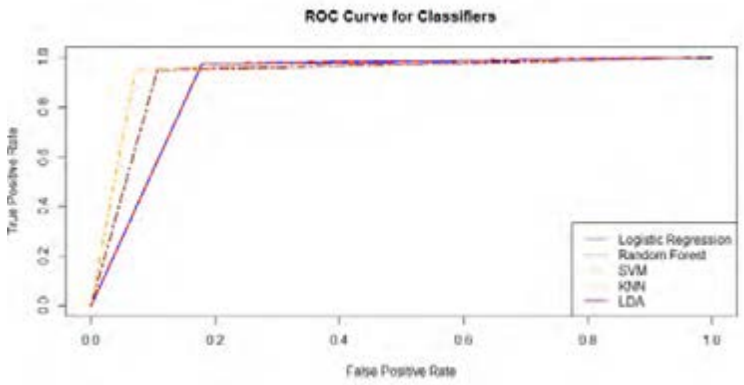
In order to further evaluate model's generalizability, it was tested on the holdout array or on a holdout dataset or test set (10%). The KNN model demonstrated a high level of precision on the test array as well, achieving a precision score of 0.939. and random forest This indicates that the model is stable and has the ability to precisely classify importers and exporters in the business sector in the forex market.

The optimal hyperparameter chosen for the KNN model in the training environment was k=5. This indicates that the model takes into account the five nearest neighbors when making classification decisions. In total, the KNN model proved itself as the most precise and reliable algorithm for classifying forex market participants, while demonstrating its matching to this specific analysis.

Moving on to the classification of nonresidents into financial and non-financial sectors, the Random Forest model was found to be the optimal choice for this classification task. The Random Forest model proved to be the most suitable algorithm for classifying nonresidents into the financial and non-financial sectors within the forex market. With its accuracy of 0.789 on the validation dataset, 0.787 mean accuracy through cross-validation, and 0.785 accuracy on the holdout test set, the model demonstrated its ability to effectively differentiate between the two sectors.

Importers/Exporters Classification model selection results

model	Accuracy in validation dataset	Mean Accuracy in train set
KNN	**0.938**	**0.935**
SVM	0.92	
LDA	0.92	
Random Forest	0.91	
Logistic Regression	0.91	

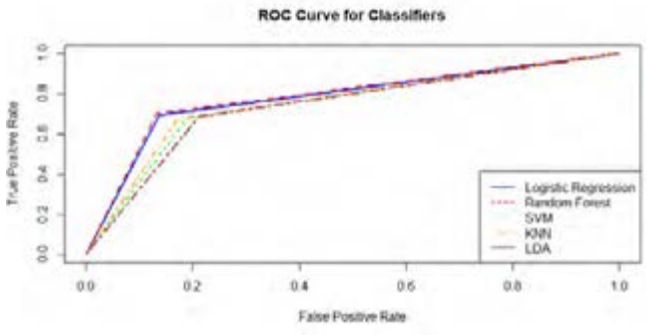


Selected Model

model	Validation dataset (20%)	Holdout dataset (10%)	Hyper parameters
KNN	**0.938**	**0.939**	**k=5**

Nonresident – Financials/non-financials Classification model selection results

model	Accuracy in validation dataset	Mean Accuracy in train set (cv)
Random Forest	**0.789**	**0.787**
Logistic Regression	0.770	
KNN	0.754	
SVM	0.748	
LDA	0.739	



Selected Model

model	Validation dataset (20%)	Holdout dataset (10%)	Hyper parameters
Random Forest	**0.789**	**0.785**	**Mtry=3,ntrees=500**

Conclusion and future work

In conclusion, the analysis conducted on both importers and exporters, as well as nonresidents classified into financial and non-financial sectors, yielded promising results using the K-nearest neighbors (KNN) model and Random Forest model, respectively. These models demonstrated high accuracy in classifying market participants within their respective subsectors in the forex market.

The utilization of these models for forecasting purposes provides valuable insights and enhances the central bank's ability to monitor and analyze different market segments. By accurately identifying importers, exporters, and distinguishing between financial and non-financial participants among nonresidents, a more comprehensive and detailed picture of the subsector-level market dynamics is obtained.

This deeper understanding of market flows serves as a crucial input for informing monetary policy decisions and developing effective macroprudential tools. The precise classification of market participants aids in assessing the overall health and stability of the forex market, enabling policymakers to make informed choices regarding interest rates, exchange rate management, and other regulatory measures. Additionally, it assists in identifying potential risks and vulnerabilities within specific subsectors, allowing for the implementation of targeted policies to mitigate these risks and ensure financial stability.

Future work in this area could involve expanding the analysis to include other subsectors or market segments, such as specific industries or types of financial institutions. By incorporating a wider range of variables and labeled data, the models' accuracy and predictive power could be further enhanced.

REFERENCES

1. Abir, A. (2020). "Developments in Israel’s foreign exchange market and the implications for the conduct of monetary policy and financial stability". *BIS Papers*, *113*, 163-171.
2. Schreiber, B. Z. (2014). Identifying speculators in the FX market: A microstructure approach. *Journal of Economics and Business*, *73*, 97-119.
3. Kumari, R., & Srivastava, S. K. (2017). Machine learning: A review on binary classification. *International Journal of Computer Applications*, *160*(7).
4. Zhang, C., Li, Y., Yu, Z., & Tian, F. (2016). "Feature selection of power system transient stability assessment based on random forest and recursive feature elimination". In *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)* (pp. 1264-1268). IEEE.



Classification of Participants in the Foreign Exchange Market

Amir Khatib

BoC-IFC seminar

15 July 2023

11

Billion $ per day

Foreign Exchange Market Turnover

The foreign exchange market in Israel plays a crucial role in the transmission of monetary policy, similar to other small open economies:

1. High pass-through of the exchange rate into inflation
2. Impact on tradable sector - Imports and Exports

Market participants, broken down by sector

Israeli residents


Nonresidents


Banking system


Financial sector


Households


Institutional investors


Business sector

Financial

Non-Financial

Exporters

Importers

Description of the database



Foreign exchange transactions– FXOption, Forward, Spot, FXSwap



50 financial intermediaries in Israel and abroad



Daily reports

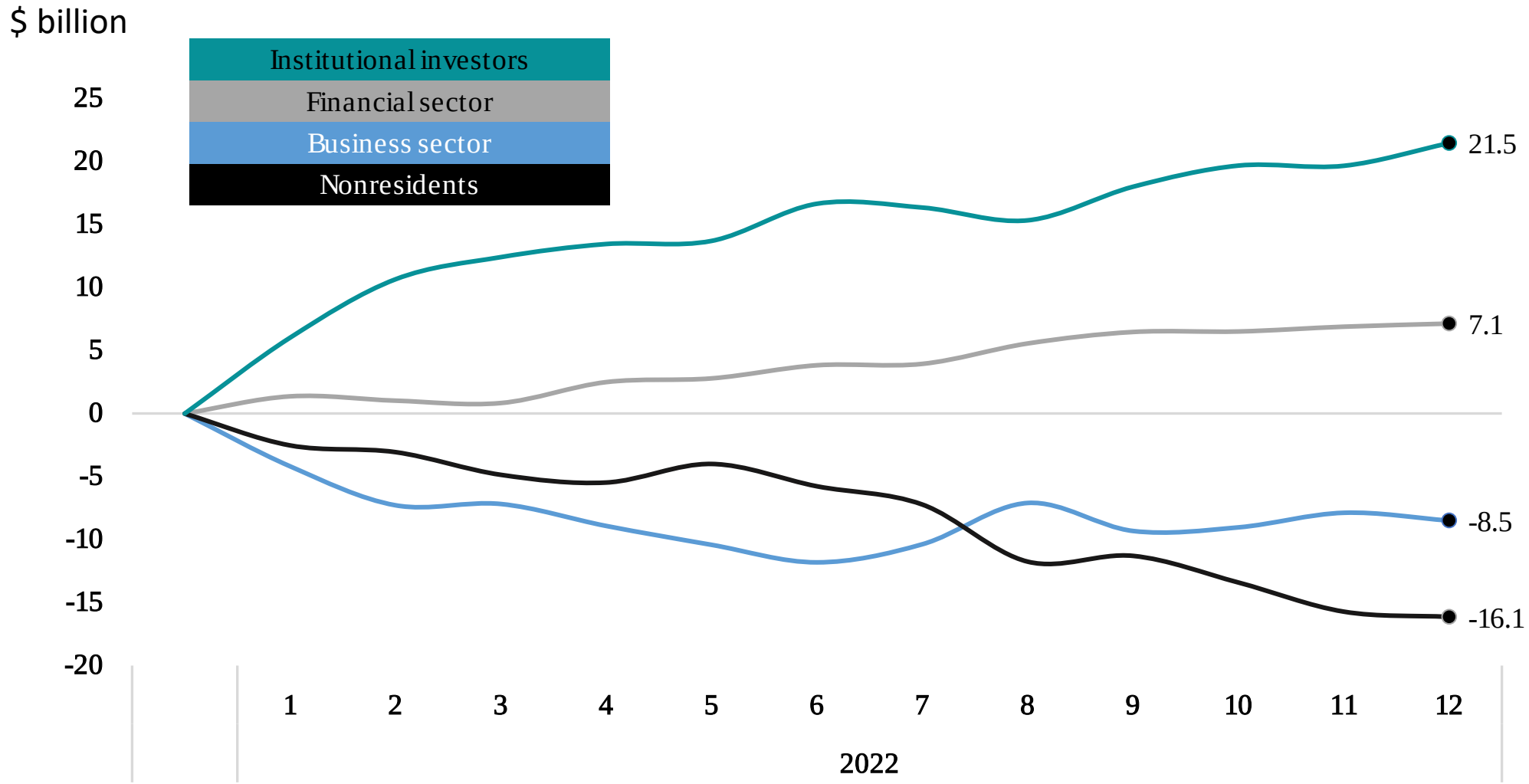


Detailed reporting at the transaction level

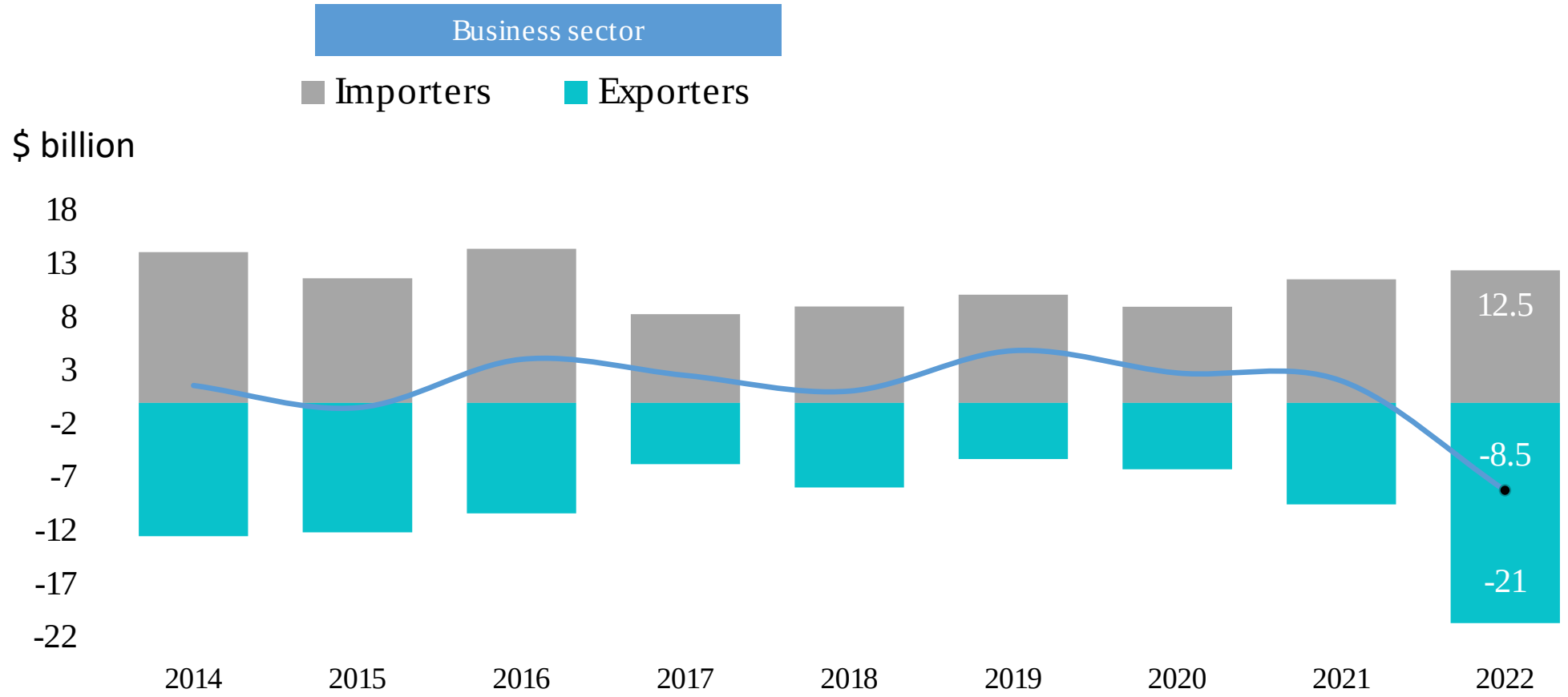


4 million transactions per year

Net cumulative foreign exchange purchases, by the main sectors

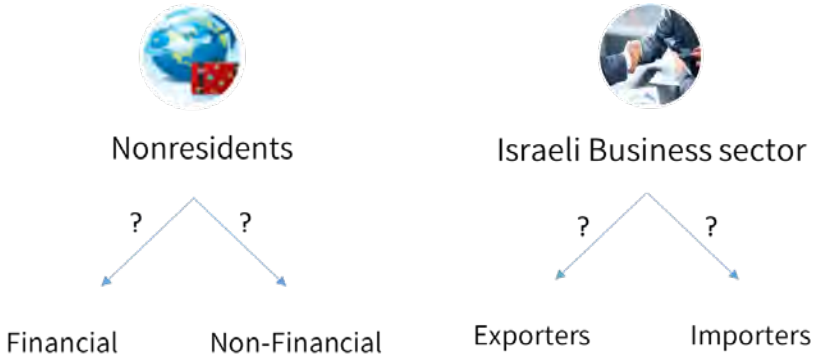


Net foreign exchange purchases



Problem statement

In certain cases, a market participant is not assigned to a subsector



Transaction date	Participant ID	Product	Currency basis	Notional amount	Buy/Sell NIS	Sector	Subsector
01/04/2018	52000117	Spot	USD/ILS	50,000,000	Buy	Business sector	N/A
05/04/2018	ADASFGFD	Forward	USD/ILS	30,000,000	Sell	Nonresident	N/A
07/04/2018	BANKESD12	Forward	EUR/ILS	1,000,000	Buy	Nonresident	Financial

The solution: Use of a **Supervised Classification** algorithm in order to classify participants to a subsector

Data preparation

- Relevant features selected from the database

  Product type, currency, financial intermediary, transaction period, sum, buy/sell indicator, and transaction rate.

- One-Hot Encoding of categorical features

cur_othr	cur_sek	cur_chf	cur_jpy	cur_gbp	cur_eur	cur_usd
0	0	0	0	0	1	0
0	0	0	0	0	1	0
0	0	0	0	0	1	0
0	0	0	0	0	1	0

- Group by Market Participant level, transformation and scaling of features
- Feature Engineering
- Partition data into Train, Validation, Test (70%,20%,10%)

Feature engineering



Importers/Exporters

(1) Importers generally engage in net forex purchases, while exporters tend to sell

Variable	Explanation
MED_Net	Median (by days) net forex purchases

(2) Exporters utilize periods in which the shekel is depreciated in order to sell foreign exchange at large scales, as opposed to importers, who buy in response to demand, without regard to the exchange rate

Variable	Explanation
LN_NETMA1	On days in which there is an anomalous depreciation in the shekel (relative to a 1-month moving average), net total forex purchases of the participant divided by total trading volume

The main variables that we added based on our familiarity with the foreign exchange market and conversations with the intermediaries and market participants over the years

Feature engineering



Nonresidents – Financial/Non Financial

(3) Financial participants are active in the market on more trading days

Variable	Explanation
Activity ratio	The share of days on which the participant was active, relative to the total number of business days

(4) Financial participants trading volume is higher than Non-Financial participants trading volume

Variable	Explanation
LN_VOL	The daily average trading volume

The main variables that we added based on our familiarity with the foreign exchange market and conversations with the intermediaries and market participants over the years

Feature selection - Recursive Feature Elimination (RFE)

```
Recursive feature selection

Outer resampling method: Cross-Validated (10 fold, repeated 5 times)

Resampling performance over subset size:

 Variables Accuracy  Kappa AccuracySD KappaSD Selected
         1   0.8976 0.7844    0.06068  0.1285
         2   0.8863 0.7612    0.06040  0.1261
         3   0.9045 0.7995    0.06122  0.1288
         4   0.9037 0.7980    0.06121  0.1285
         5   0.9107 0.8129    0.06783  0.1429
         6   0.9168 0.8259    0.06496  0.1360
         7   0.9194 0.8311    0.06448  0.1363
         8   0.9220 0.8364    0.06205  0.1309        *
         9   0.9185 0.8293    0.06429  0.1350
        10   0.9220 0.8364    0.06267  0.1322
        11   0.9186 0.8288    0.06305  0.1334
        31   0.9194 0.8308    0.06325  0.1331
```



Importers/Exporters

In addition,
VIF <5 , pearson correlation between predictors <0.7

7 features selected

Train various Supervised Classification algorithms on train set (70%) and tune parameters (5-fold CV)

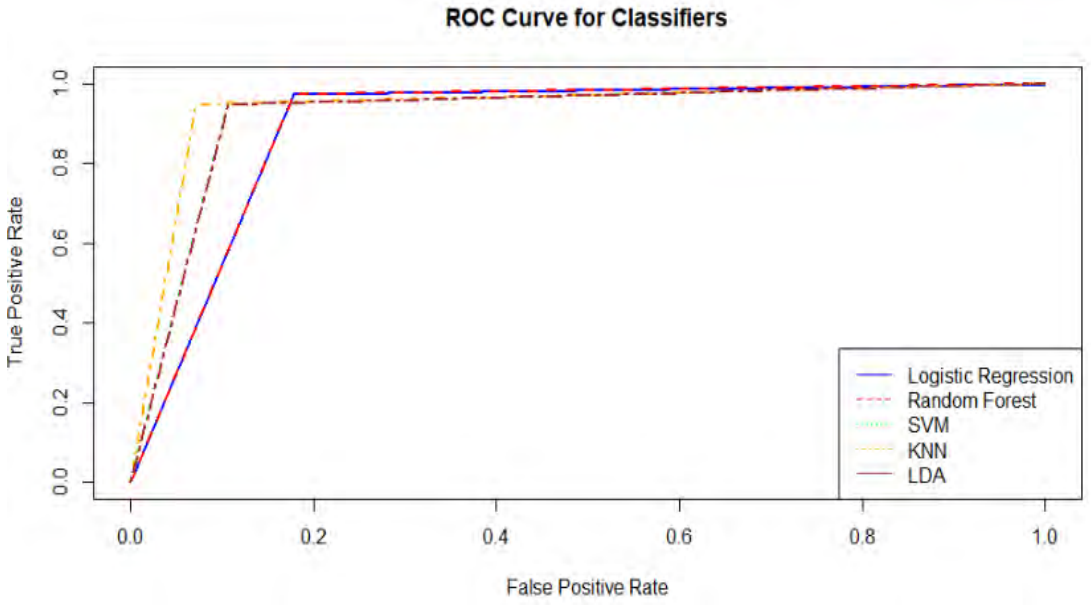
1. Random Forest
2. Support Vector Machines (SVM)
3. Logistic regression
4. K-Nearest Neighbors (KNN)
5. Linear Discriminant Analysis (LDA)

Model selection – on Validation dataset



Importers/Exporters

model	Accuracy in validation dataset	Accuracy in train set
KNN	ě┼Ň38	ě┼Ň39
SVM	ě┼Ň2	
LDA	ě┼Ň2	
Random Forest	0.91	
Logistic Regression	0.91	

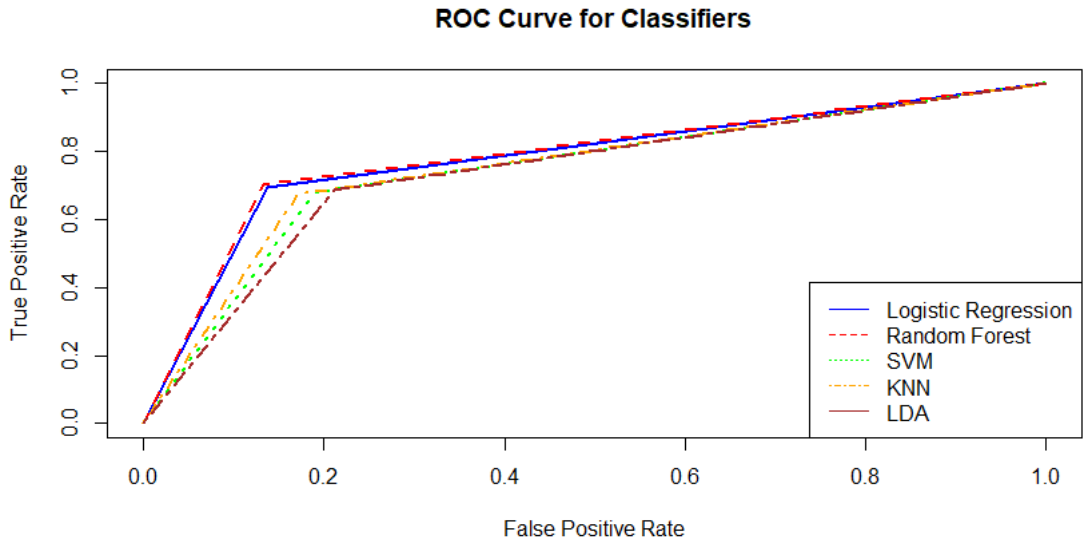


Model selection – on Validation dataset



Nonresidents – Financial/Non Financial

model	Accuracy in validation dataset	Mean Accuracy in train set (cv)
Random Forest	**0.789**	ě┬ŃńŃ
Logistic Regression	0.770	
KNN	ě┬ŃĽl’	
SVM	ě┬Ń’ń	
LDA	ě┬ŃĽŇ	



Selected models



Importers/Exporters

model	Validation dataset (20%)	Holdout dataset (10%)	Hyper parameters
KNN	0.938	0.939	k=5



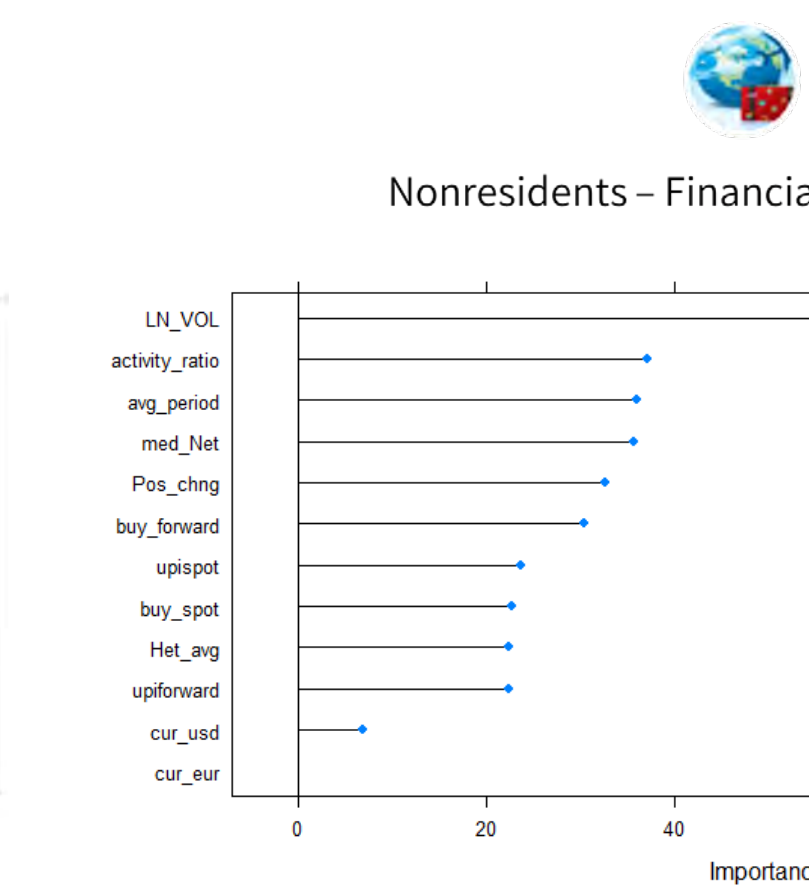
Nonresidents – Financial/Non Financial

model	Validation dataset (20%)	Holdout dataset (10%)	Hyper parameters
Random Forest	ě┼ŃňŇ	0.785	Mtry=3,ntrees=500

Data interpretation



Importers/Exporters



Nonresidents – Financial/Non Financial

Summary



Israeli forex market significance



Granular data reporting from market participants



Accurate subsector predictions using supervised classification models



Enhanced central bank market analysis capabilities



Thank you for your attention.

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Access to CSI for research purposes[1]

Matthias Rumpf,
European Central Bank

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

EUROPEAN CENTRAL BANK

EUROSYSTEM

Access to CSI for research purposes

BoC–IFC satellite seminar on granular data

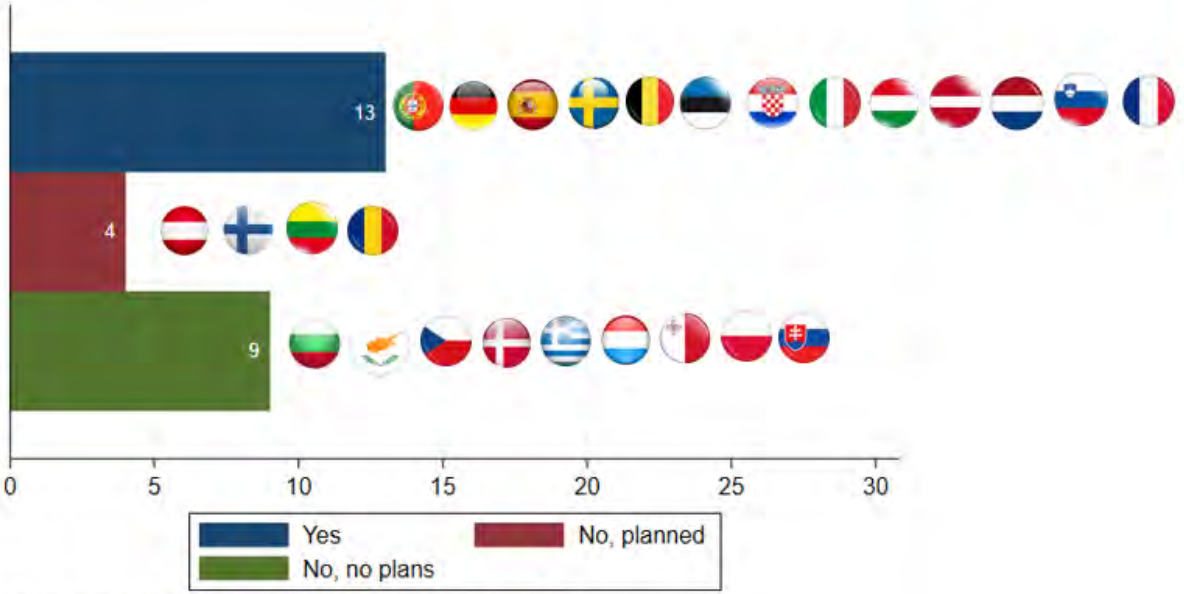


15/07/2023

Matthias Rumpf
European Central Bank

Most NCBs provide access to CSI microdata or plan to do so

Do you share or plan to share CSI?



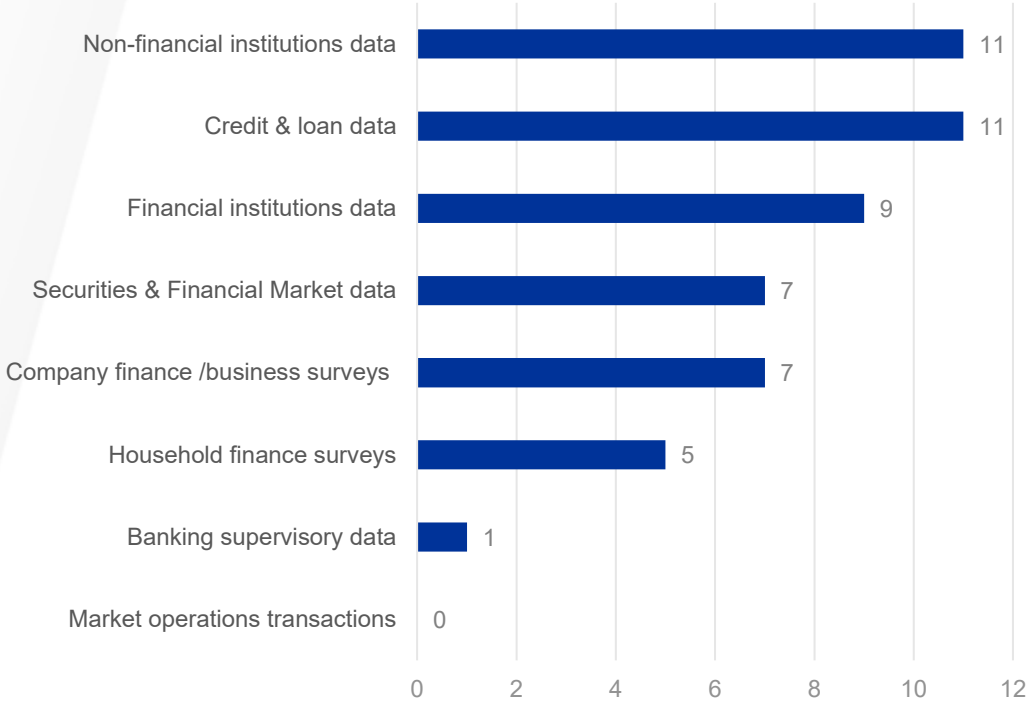
Graph based on answers from 26 surveys, one answer per survey.
Source: STC WG SC - Survey on microdata access for external researcher.

www.ecb.europa.eu ©

Available datasets varies across NCBs

For which datasets do you/will you grant access?

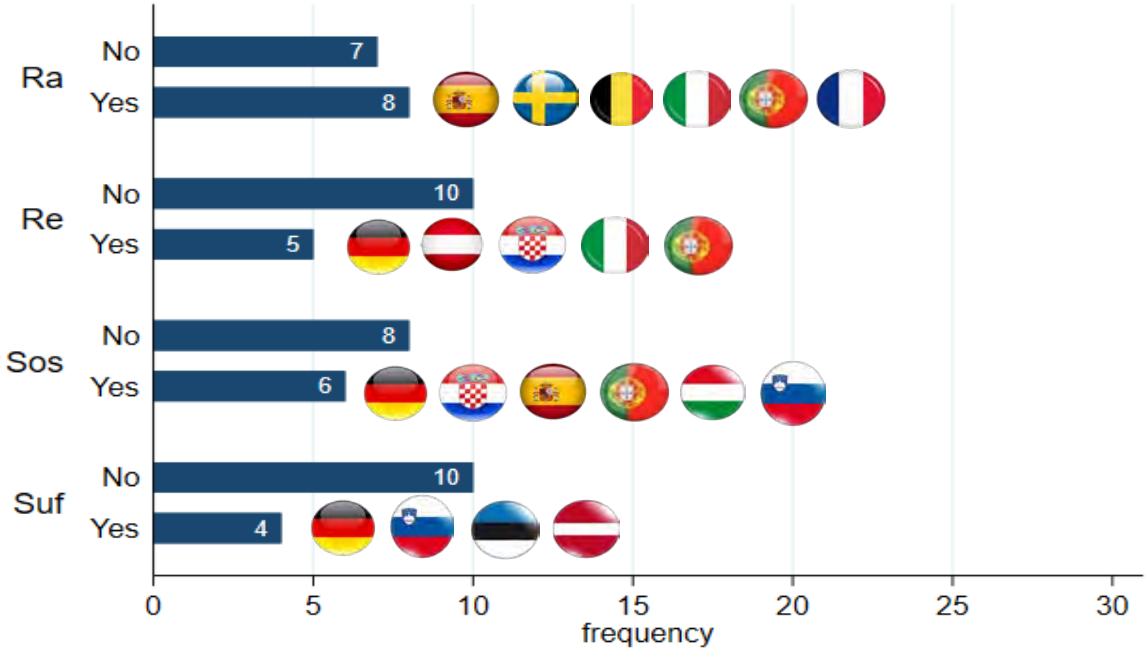
	Number of datasets
1 NCB	7
4 NCBs	5
3 NCBs	4
3 NCBs	3
1 NCB	2
1 NCB	1



www.ecb.europa.eu ©

Access modes applicable to CSI

Access granted to CSI by access mode



Graph based on answers from 14 (SOS,SUF) or 15 (RE, RA) surveys, one answer per survey.
Source: STC WG SC - Survey on microdata access for external researcher.

www.ecb.europa.eu ©

Access to CSI based on EU Council Regulation 2533/98

Article 8.1 c

[The ESCB may grant] scientific research bodies access to confidential statistical information which does not allow direct identification, and with the prior explicit consent of the authority which provided the information;

Article 8.6

Confidential statistical information may be exchanged within the ESCB to grant scientific research bodies access to such information, in accordance with paragraphs 1(c) and 2.

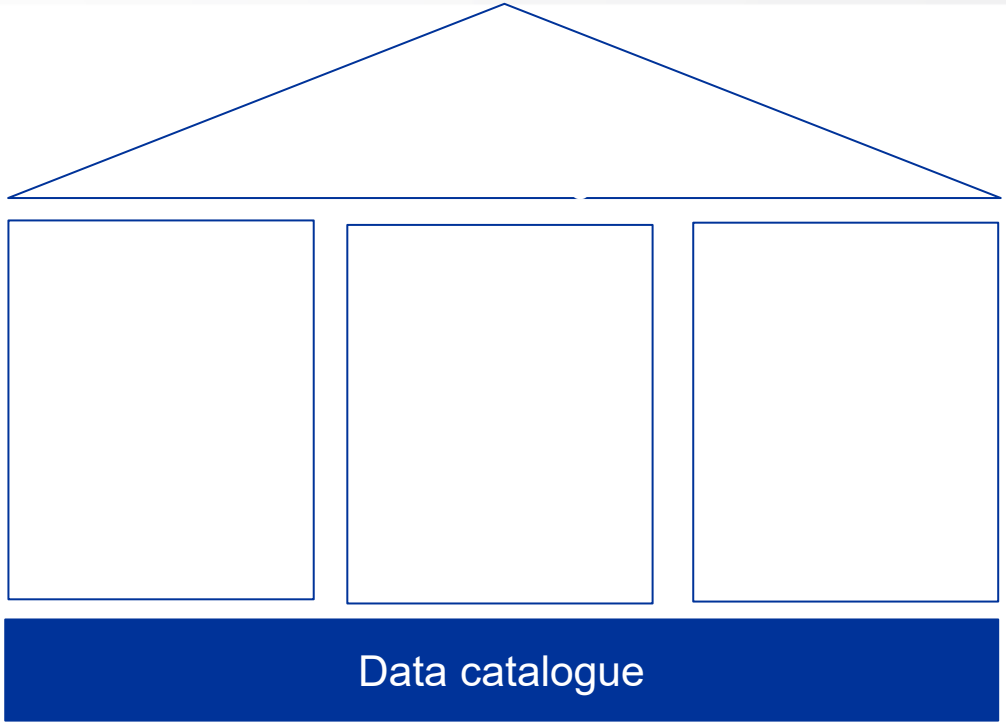
www.ecb.europa.eu ©

Challenges



www.ecb.europa.eu ©

The building blocks



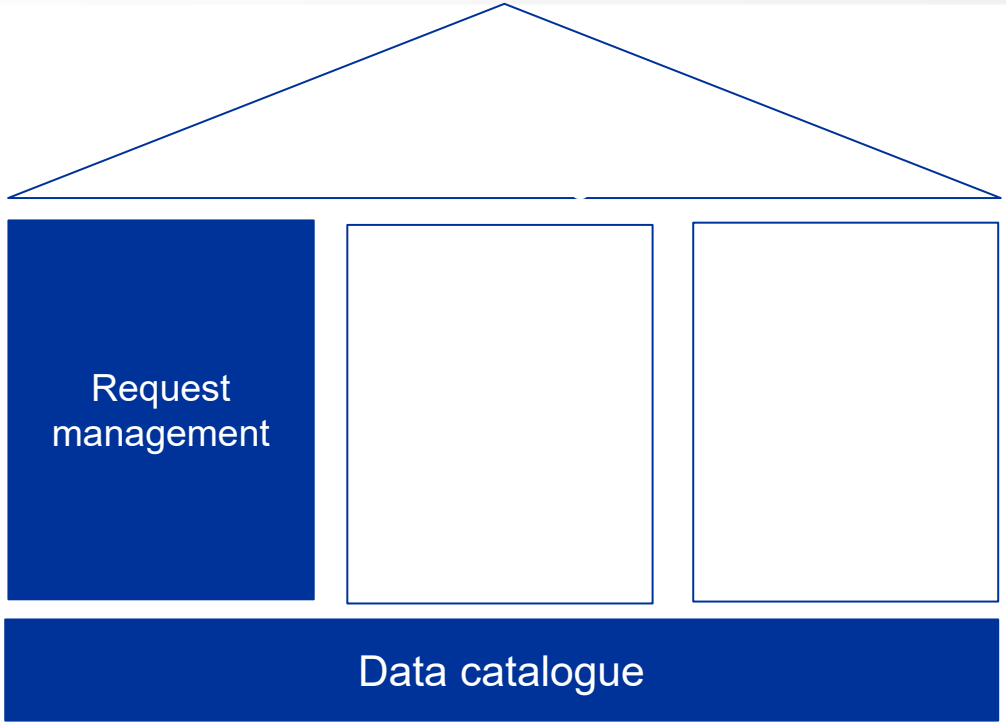
Data catalogue in three steps

- List of access points
- Standardized metadata
- Machine readable service

➔ In line with catalogue for statistical data and Common Data Management (CDM)

www.ecb.europa.eu ©

The building blocks



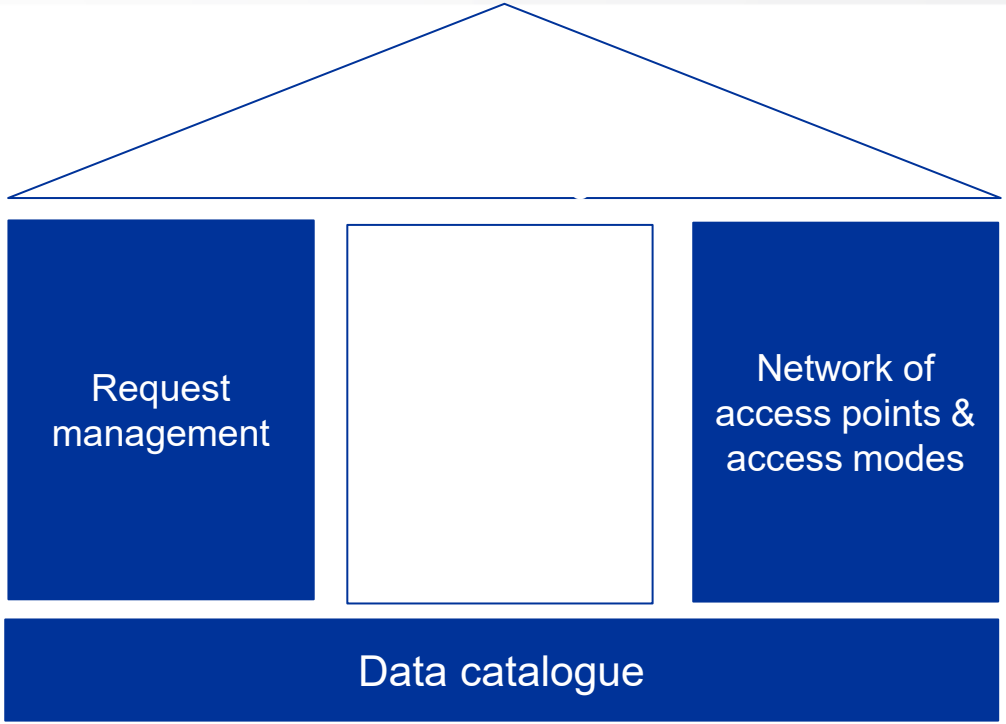
Common request management

- Improve user experience, even while underlying process are not harmonized
- Gather data on user interests

➔ Synergies with management of user requests for statistical data

www.ecb.europa.eu ©

The building blocks



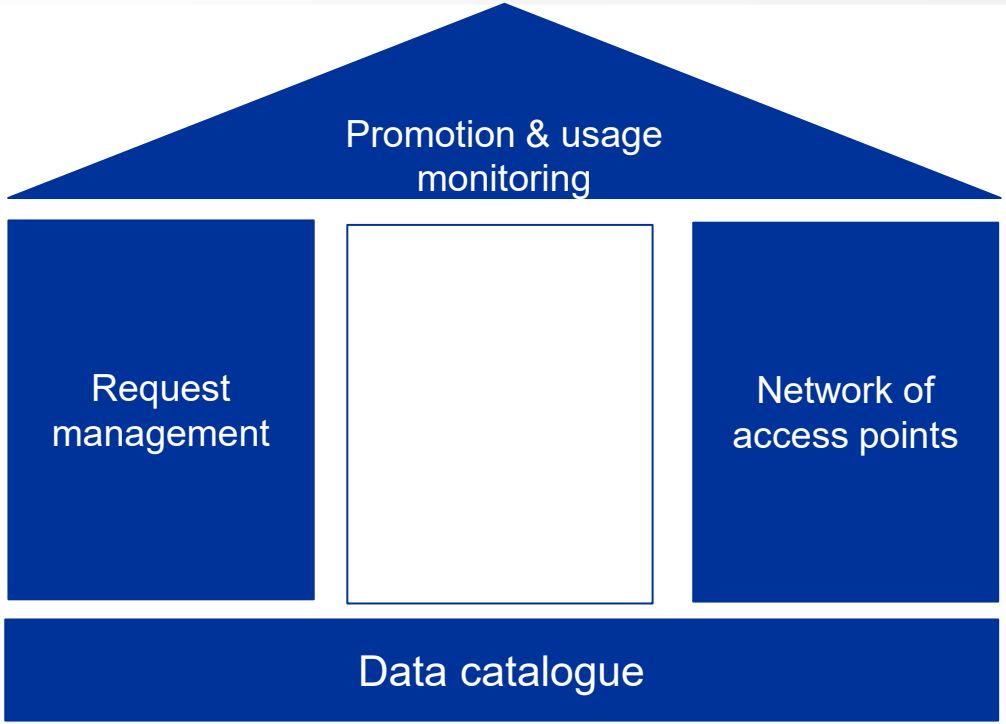
Network of access points

- Enable data sharing for multiple countries
- Enable merging with external data sources

➔ Build on experience from ESS and IDAN network

www.ecb.europa.eu ©

The building blocks



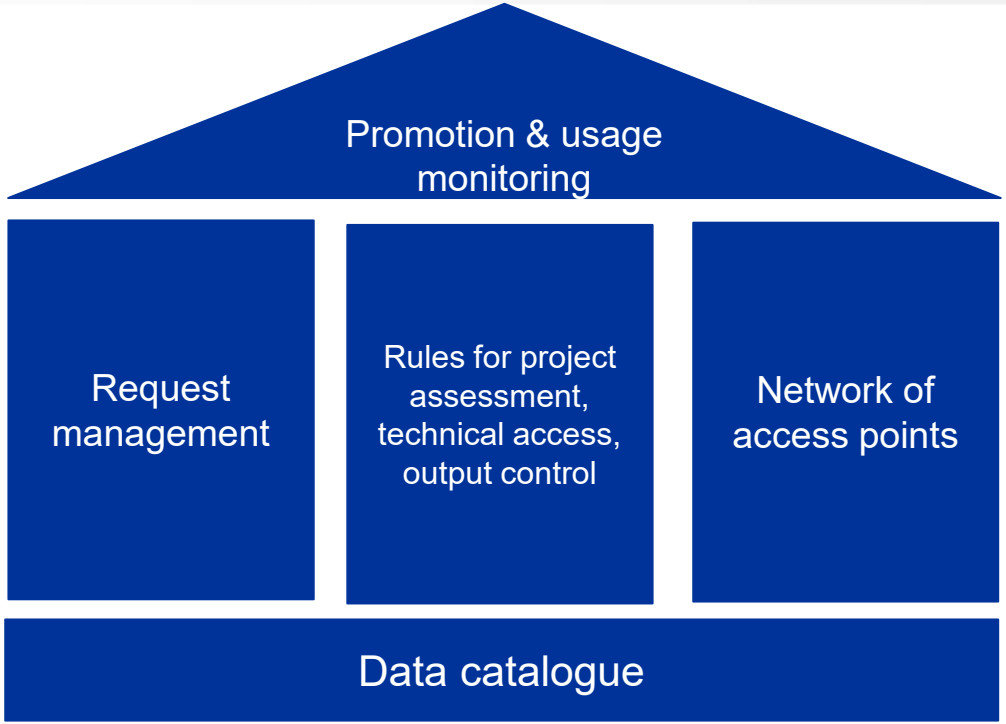
Promotion and usage monitoring

- Outreach to research bodies
- Alignment with ESCB research agendas
- Monitor usage and research output

➔ Collect feedback & adjust service to demand

www.ecb.europa.eu ©

The building blocks



Common rules

- ECB guidelines based on 2533/98
- Common standards for entire process
- Decentralized execution

➔ Fully harmonized rules for granting access to ESCB datasets

www.ecb.europa.eu ©

Thank you!

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Connecting the dots of the international debate on the standardization and granularity of regulatory data[1]

Massimo Casa,
Bank of Italy

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

CONNECTING THE DOTS OF THE INTERNATIONAL DEBATE ON THE STANDARDIZATION AND GRANULARITY OF REGULATORY DATA

by Massimo Casa[1]

Abstract

The European banking reporting 'ecosystem' for statistical, prudential and resolution purposes has become increasingly complex and costly for both authorities and financial institutions. We are confronted with a significant number of surveys nowadays, combining aggregated, item-level and transactional data. The underlying business phenomena are often overlapping and not always defined in a consistent, integrated and reconcilable manner. Against this background, this paper tries to (i) connect the dots of the intense debate that has taken place in various European and international fora over the last five years, focusing on how to rationalize the overall reporting framework, and (ii) map out a possible way forward. Two main streams of innovative initiatives, high-level reports and studies are reviewed. The first aims to define a mandatory, common granular layer of multipurpose (integrated) data, located in banks' IT systems, which would provide programmable code to reference these data in order to generate the reporting of regulatory metrics. The second stream aims to standardize banks' operational systems by generating granular contractual information in near real-time. This paper argues that, despite the attempts to demonstrate the potential benefits of these innovative approaches, their adoption to replace the existing setup still appears premature even in the medium term. While exploring new ways to benefit from 'on demand' access to timely and integrated sources of data, the authorities in Europe have already launched strategic initiatives for the rationalization, standardization and integration of the existing reporting framework. In this respect, the ESCB's IReF and the wider integration initiative at European level, following the EBA's feasibility study, are important and tangible actions to significantly reduce the reporting burden on all stakeholders in the medium term. The cooperation between the European authorities and the banking industry will be key in achieving this general goal; this paper argues that the ESCB's BIRD would provide an important contribution.

JEL classification: C81, G21, M15.
Keywords: standardization, granularity, integrated reporting, machine-executable reporting, smart contracts.

Contents

Introduction 5

1. Criticalities in the current regulatory reporting framework 6

2. Ongoing innovative initiatives focused on standardization and granularity 7

3. A more feasible and likely scenario for the future 16

Conclusions 20

References 21

[1] Bank of Italy, Statistical Data Collection and Processing Directorate.

Introduction[1]

European banks are required to report a wide range of data to central banks and other competent authorities on a regular basis to enable them to carry out their institutional functions (e.g. monetary policy, financial stability, banking supervision and resolution).

These reports have undergone profound changes over the last two decades[2], which has led to faster response times and increased volume,[3] granularity, and a growing complexity in the regulatory requirements. As concluded by the European Commission (2021), on the one hand, stricter reporting obligations have 'helped address information gaps identified during the global financial crisis and reflect the increasing complexity and interconnectedness of the financial system'; on the other, they have 'placed an increasing burden on both the reporting entities and the supervisory authorities receiving and analysing the data'.

In 2019, as part of the comprehensive fitness check of European Union (EU) supervisory reporting requirements in EU financial services (European Commission, 2019), the European Commission (EC), while concluding that the current reporting requirements are necessary and effective for supervisory purposes, identified significant inefficiencies in the way they are defined and data are collected, specifically: (i) the reporting requirements are not sufficiently harmonized; (ii) data are not efficiently shared and reused among authorities; (iii) the reporting requirements are not always consistent across reporting frameworks; (iv) the reporting requirements are often developed by different authorities in silos, without coordinating and considering what is already in place; and (v) the current design of the reporting requirements does not allow for automated processing, and the lack of standardization hinders the application of modern technologies. Such inefficiencies not only negatively impacted the quality and usability of the reported data, but also generated additional costs for the reporting entities. Furthermore, it was noted that the reporting frameworks do not make full use of modern IT tools.

One of the most significant changes has been the growing demand for increasingly granular data. Important collections of granular data have been launched in recent years. In September 2018, the Eurosystem central banks launched AnaCredit, which contains harmonized, detailed information on individual bank loans (with a minimum threshold of €25,000) granted in the euro area to counterparties identified as legal entities. At the same time, all banking groups under the direct supervision of the ECB started reporting granular information on global securities holdings (SHSG data). Other initiatives included: (a) money market statistical reporting (MMSR), launched in 2016, which collects transaction-by-transaction data on a daily basis for more than 50 large euro-area banks in four different segments of the euro money market; (b) European Markets Infrastructure Regulation (EMIR) reporting on derivatives transactions and (c) securities financing transactions reporting (SFTR).

---

[1] The views expressed are solely those of the author and do not necessarily reflect the opinion of Banca d'Italia. A preliminary version of the work was presented at the satellite seminar on 'Granular data: new horizons and challenges for central banks' co-organized by the Irving Fisher Committee on Central Bank Statistics and the Bank of Canada and held in Ottawa on 15 July 2023 in conjunction with the 64th World Statistics Congress of the International Statistical Institute (https://www.bis.org/ifc/events/ifc_230715_agenda_final.pdf).

[2] The importance of European harmonized information needs 'grew over time together with the progress of the integration process that resulted in the launch of the single monetary policy and, later on, in the reform of the European architecture of the micro and macro prudential supervision and the establishment of the SSM and SRM. With the data gaps that emerged after the global financial crises at the end of the first decade of the current century and the initiatives of the G20, Financial Stability Board (FSB) and International Monetary Fund (IMF), new requirements for credit and financial data, and the related information projects, have transcended the European dimension to become global' (Casa et al., 2022b).

[3] 'In aggregate, financial institutions in the EU are subject to more than 500 reporting obligations comprising more than 1,000 tables with more than 70,000 individual data points' (European Commission, 2019, pp. 52-55).

1. Criticalities in the current regulatory reporting framework

Reporting agents find it increasingly difficult to respond, in a timely and accurate manner, to the numerous new requests for information received from multiple authorities and to the frequent changes to pre-existing reports. In turn, such difficulties might also impact on data quality to the extent that the authorities cannot guarantee that the various surveys are collected on the basis of consistent (regulatory) interpretations across reporting entities.

Table 1 – The current framework of regulatory reporting for European banks

Survey	Regulating Authority	Purpose	Frequency	Granularity	Business scope	Banks required
BSI – Balance Sheet Items	ECB	statistical	monthly	aggregated	assets and liabilities	all banks (some derogations)
MIR – MFI Interest Rate statistics	ECB	statistical	monthly	aggregated	interest rates	sample of banks
AnaCredit	ECB	multipurpose	monthly quarterly	item-level data	loans	all banks (some derogations)
SHSG - Securities Holdings Statistics Group	ECB	multipurpose	quarterly	item-level data	holdings of securities	all banking groups
EBA-ITS on Supervisory Reporting	EC + EBA	supervisory and resolution	monthly quarterly semiannual yearly	mostly aggregated	accounting, risk and resolution data	all banks
SRB Resolution Reporting	SRB	resolution	yearly	aggregated	resolution data	significant institutions (banks under SSM direct supervision)
MMSR - Money Market Statistical Reporting	ECB	statistical	daily	quasi-transactional	money market	sample of banks
EMIR reporting	EC + ESMA	multipurpose	daily	quasi-transactional	derivatives	all banks (market operators)
SFT reporting	EC + ESMA	multipurpose	daily	quasi-transactional	securities financing transactions	all banks (market operators)
Locational banking statistics	BIS	multipurpose	quarterly	aggregated	assets and liabilities	internationally active banks
Consolidated banking statistics	BIS	multipurpose	quarterly	aggregated	consolidated positions	internationally active banking groups
Survey of FX and OTC derivatives markets	BIS	multipurpose	triennial	aggregated	OTC derivatives	sample of institutions
Local reporting	local authority				in many countries central credit registers	
Ad hoc data collections	supervisory competent authority	supervisory				

As shown in Table 1, the current regulatory reporting framework is composed of a set of national and European statistical, prudential and resolution requirements, as well as 'ad hoc' data collections. The proliferation of surveys and the lack of an effective *ex ante* coordination among the various authorities

(which adopt different 'languages' and collect partially redundant data) contributed significantly to the sharp increase in the reporting costs. According to the available official estimates for harmonized supervisory reporting, these costs have almost doubled over the past decade for reporting agents[4] .

The process described above is continuously evolving. The implementation of the upcoming CRR III banking regulation package, the new information requirements in relation to the environmental, social and governance (ESG) factors and risks and the further shift towards more granular data for statistical reporting, as part of the Eurosystem's Integrated Reporting Framework (IReF) due to be completed by 2027, will introduce further innovations in the coming years. As a consequence, financial institutions will be called once more to 'translate' new reporting requirements into their high granular terms and into changes to their internal data models. Since the regulations cannot define the connection to that granular level, each institution will individually interpret, translate, implement and manage the high amounts of necessary rules and data elements that allow to map their internal data to the information requirements defined by the authorities.

In the described framework, the level of cooperation between the authorities could be strengthened. Synergies that could arise through the exchange of data between different authorities are not leveraged because processing is organized in administrative silos and the legal basis for data exchange, that would prevent overlaps among the different regulatory frameworks, needs to be made more effective and efficient[5]. Moreover, the reporting system does not make use of modern technological capabilities as it would be required.

Furthermore, the current regulatory reporting is not aligned to the changes in the financial landscape. A transformational shift in the volume, speed and variety of data is driving the innovative use of financial technology (the so-called FinTech), leading to rapid changes in the financial environment.

Supervisors, for example, are faced with the challenge of having to assess rapidly evolving risks to business models and technology-driven changes that may affect financial stability, with regulatory data that are infrequent, backward-looking and collected according to legacy frameworks and not support new tools or applications that could enable real-time or early warning analytics. Technology solutions could enable supervision to be more forward-looking, insights-based and data-driven. In addition to regulatory data, in fact, supervisors use other sources of information to help inform their prudential assessments, unstructured data not organized in a specific or unified format, unlike regulatory reporting which is based on data that are structured. Gathering intelligence from these unstructured data sources in addition to regulatory reporting requires a fundamental shift away from current processes to ones that are enabled by technology.

2. Ongoing innovative initiatives focused on standardization and granularity

Since 2016 there has been a proliferation of studies trying to find innovative solutions to the criticalities described in the previous section, leveraging standardization and granularity on one side

[4] As far as supervisory reporting regulated by EBA is concerned, it has been estimated for the period 2018-2020 that European banks overall bear an annual cost, including ongoing costs and implementation costs, of about 5.5 billion euros (European Banking Authority, 2021a). Based on the results of a questionnaire addressed to institutions, the most prominent areas of concerns are complexity, the amount of information to be reported, internal data extraction and calculations and the stability of the EBA supervisory reporting framework.

[5] This aspect has been highlighted also by the industry: 'Share Information: amongst the authorities instead of asking several times for data that has been already provided' (European Banking Federation, 2021).

and technology on the other side; such studies are classified under the general heading of 'Regulatory Technology', better known as RegTech.

While there is a consensus that future improvements to the reporting framework should be the result of an intense and effective cooperation between the various authorities and the industry, the solutions proposed so far are quite different in terms of standardization and granularity of the data.

In particular, they differ with respect to the layer of the data they suggest to standardize in the information system of the financial institutions: in the more 'visionary' proposals, their operational system, impacting directly on the way transactions are recorded; other solutions their data warehouse/data lake, which hosts the so-called input layer; in the more conservative ones, the reports that they have to submit to the authorities (i.e. the output layer).

Also in terms of level of granularity there is a great difference between the various proposals which are debated, in descending order:

i. granularity at the level of operational data (smart contracts in pure business language);
ii. granularity at the level of common input data (business-oriented in its design but also contaminated by regulatory definitions);
iii. item-level reporting (e.g. loan-by-loan or security-by-security, as it is the case, respectively, for AnaCredit or SHSG in the euro area) combined with multidimensional structures for more aggregated data.

Below the focus is on the first and second types of solutions, starting with those based on common input data, the so-called 'machine-executable reporting solutions'.

The most interesting aspect of these solutions is that a reporting regulation would be drafted in a form that allows for its automatic implementation by the financial institutions. This would contribute not only to facilitating compliance with regulatory requests but also to reducing the times necessary to obtain new structured and periodic information and to transform a new information requirement into a reporting obligation (a process which might take a relatively long period of time).

Two main features characterize these new regulatory reporting solutions: (a) a very high level of granularity of the information requests from the authorities; (b) the regulatory authorities bear the responsibility to define the transformations of primary data into the aggregated information necessary for their analyses.

In general, these solutions rely on a single data dictionary, which allows data standardization and must be shared and applied by all the parties involved. A first possibility is that the authorities use the dictionary to regulate in detail granular input data and their transformations to calculate aggregated information. In this framework instructions are defined 'as a code' and the authorities describe the transformations as a machine-executable code that reporting entities can perform directly to calculate the required reports from input data. Alternatively, in the case of the so-called 'data-pull models', authorities shall describe in the dictionary the data that reporters must deliver in a staging database, i.e. an information storage area that the authorities will access in order to carry out the necessary processing for their own purposes; hence, the responsibility for processing granular data to obtain regulatory aggregates is left to the authorities. In addition, the concept of predefined reporting frequency is overcome, as an authority can potentially pull available data when needed.

The BIS Innovation Hub Project Ellipse[6] is a prototype that aims at offering to the community of authorities an opportunity to reflect, explore and collaborate on common innovative granular solutions to streamline data collections and enhance the analytical capabilities of supervisors. The primary objective of the project was to perform a proof-of-concept (PoC) of an integrated regulatory data and analytics platform – known as Ellipse - that combines structured granular data and unstructured sources of data relevant to analyze specific, new events in real-time; in particular, it allows to apply advanced data analysis techniques to generate early warning indicators, analytics and prudential metrics (Bank for International Settlements Innovation Hub and Monetary Authority of Singapore, 2022). So far, Phase 1 of the project - the collection of information - and Phase 2 - the advanced use of data – have been completed. For the purposes of this paper, the Phase 1 is particularly relevant to the extent that it investigated how data-driven supervision could be enabled by machine-executable digital reporting, using a cross-border common data model. It was shown that regulatory reporting requirements can be expressed in unambiguous machine-readable logical reporting instructions underpinned by a consistent data model[7]. A coded version of the calculation steps for generating regulatory reports can also be published alongside regulations to ensure a clear understanding of the expected data at the most granular level.

The Bank of England, in partnership with the Financial Conduct Authority (FCA) and a small group of financial institutions, has carried out a pilot of technology (Digital Regulatory Reporting - DRR[8]) to explore the feasibility for regulators to publish a code version of their instructions – so called 'machine-executable regulation', which financial institutions could automatically convert into code that runs directly in their internal systems. The pilot tested the concept using synthetic mortgage data supplied by the institutions in a standardized format, running on a dummy system. Authorities could then run reports according to their needs or pull data directly from institutions via an application programming interface (API). In a second phase of DRR, a collaboration with the International Swaps and Derivatives Association (ISDA) Common Domain Model (CDM) project for derivatives was also carried out. This collaboration investigated how to build derivative reports and digital regulation using the technology that ISDA was developing, with the intention to extend their tools to meet DRR needs. Four data points for two separate transactional derivative reports were described using the CDM's version of digital regulation. A single data model covering the production of multiple regulatory reports suggested that the DRR approach could be scalable across reports, and therefore more convenient from an economic point of view. The digital version of the reporting instructions was linked to the CDM and the CDM to real test files of operational data provided by institutions. To the extent that they adopt the CDM, this could significantly reduce the cost of implementing a DRR approach for derivatives reporting. This work also proved how regulators could modify reports by adopting digital regulation, defining a single common data model for a product type and reports generated from this single model.

The German financial supervisory authority (BaFin) - in partnership with Deutsche Bundesbank, credit institutions, service providers, industry associations and a consultancy firm – carried out a feasibility study for 'A reporting system for the future' (BaFin, 2022) with the aim to develop options for redesigning the reporting system in the direction of i) reducing the burden on institutions (and supervisors), ii) improving supervisors' analytical capabilities, and iii) ensuring the interoperability

[6] See https://docs.bis.org/ellipse/ .

[7] The PoC's mortgage attributes have been modelled using the Common Domain Model (CDM) developed by the International Swaps and Derivatives Association (ISDA) for OTC derivatives, cash securities, securities financing and commodities. Testing the CDM's use and extensibility for other product domains such as mortgage loans was key to the PoC, as globally applicable common data models that can be used across products could reduce the number of data models in use by financial institutions and ease the mapping burden. In addition, as the CDM is open-sourced, this allows the model to be more widely accessible for testing within existing environments.

[8] See https://www.fca.org.uk/innovation/regtech/digital-regulatory-reporting .

of new approaches with the European legal framework. The study examined whether the institutions' heterogeneous source and result data could be mapped centrally on the basis of a common granular data model. The model should be[9]: 1) 'comprehensive': it should cover the reporting requirements of banking statistics, and the supervisory (CoRep, FinRep, etc.) and resolution regimes; 2) 'standardized' and 'proportionate': it should be commonly applicable to all institutions; 3) 'sustainable': it should be continuously extensible because the model grows with its requirements; 4) based on the 'report once' principle: data that can be expressed on granular level should only be contained and queried in the model once; 5) 'mixed granular': the model should be capable of mapping both granular and, when necessary, aggregated data. Institutions would have to submit a mixed granular data set for each group entity (solo) and for each consolidated subgroup and group in order to comply with the current reporting requirements. Aggregates that are summable could be calculated based on the mixed granular data provided and the common machine-readable transformation rules that are uniformly defined for all institutions[10].

The European Commission is also exploring new approaches to data collection. To this end, it initiated a project with the aim of assessing how regulatory reporting could be made ready for the digital age (European Commission, 2022). The project, developed in close cooperation with the European Securities and Markets Authority (ESMA) and PwC, had three key objectives: 1) developing, a PoC for a machine-readable and executable representation of reporting requirements on derivatives specified in the European Market Infrastructure Regulation (EMIR, level 1 regulation) and the connected Regulatory Technical Standards and Implementing Technical Standards (RTS and ITS, level 2 regulation)[11]; 2) assessing whether a machine-readable and executable reporting (MRER) system can lead to more effective and efficient development of reporting requirements and reduce the reporting burden for reporting entities, including an assessment of the costs and benefits of such an approach; 3) identifying regulatory obstacles to MRER technology-enabled innovation in the financial sector. One of the key elements for an effective implementation of MRER is the appropriateness of the data model. The project team has recommended the selection of one input data model that is supported by the majority of the industry, as its adoption by the financial institutions would be key in order to decrease the reporting burden. The project team has researched and tested different options but no immediate fit-for-purpose data model has been identified. Both the evaluated data models– the Common Domain Model (CDM) proposed by ISDA, and the ISO20022 Business model –require some kind of extensions in order to be incorporated into the MRER-solution. The project team has identified the necessary requirements and has recommended ESMA to play a leading role in guiding the industry in designing a fit-for-purpose data model for selected reporting frameworks.

Table 2 contains a summary of the pros and cons of the machine-executable reporting solutions, as they were assessed during the projects described above.

[9] The study assessed that, in its November 2021 edition, the 'Banks' Integrated Reporting Dictionary' (BIRD) developed by the European System of Central Banks (ESCB), in cooperation with the banking industry, comes very close to the criteria in terms of its own objectives, governance setup and model structure.

[10] The study has revealed that between 75 and 90 per cent of all data aggregates in the current banking supervisory and statistical reporting system could be generated from granular data.

[11] These regulatory texts specify the technical standards on the minimum details of the data to be reported to trade repositories.

Table 2 - Pros and cons of the machine-executable reporting solutions

PROS	
Consistency across reports	Where multiple reports rely on the same underlying data, this approach could eliminate some duplication by mapping to a consistent, underlying input layer (Bank of England, 2020).
Consistency across institutions	The introduction of standardized, consistent central master data for institutions in addition to reference data for securities were cited as additional potential types of relief (BaFin 2022). This could also lead to reducing possible misinterpretations of certain regulation by different market participants (European Commission, 2022).
Flexibility	A greater cost saving from this approach could come from making new requests cheaper and faster to respond to, since institutions would avoid having to identify and extract data from elsewhere in their systems if it existed in a common input layer (Bank of England, 2020).
Compliance	Various types of relief, such as easier interpretation, improved interaction or a reduction in ad hoc queries (BaFin, 2022).
Transparency	Digitizing regulatory reporting rules may lead to other benefits, such as increased transparency of regulation, which is a priority for institutions (Bank of England, 2020). Compared to the current situation, during the publication stage of a new regulation, the new system would allow for less effort in explaining the new regulation, as this would be unambiguously expressed in the published code (European Commission, 2022).
Regulatory efficiency	If parts of the reporting instructions are published as code the need for the publication of guidelines and Q&As is lower (European Commission, 2022).
Common understanding	'Instruction as a code' has the potential to be both readable for reporting domain experts and unambiguously executable by machines. This potentially limits misunderstandings between business analysts and software engineers, and therefore the number of software defects (European Commission, 2022).
CONS	
Lack of incentives	The initial investment and costs needed to achieve this outcome can steer decisions based on business incentives and risk appetite. While data standards initiatives would ultimately benefit if financial institutions and authorities were to adopt a greenfield approach, and do indeed require them to do so, the perceived scale of such an exercise could also deter stakeholders from embarking on it (Bank for International Settlements Innovation Hub and Monetary Authority of Singapore, 2022).
Reporting entities' implementation costs	The introduction of a new data model together with a comprehensive data set for trade record could impose significant costs on the reporting entities, the actual cost differing per reporting entity depending on their size, the complexity of their internal systems and their level of digital maturity (European Commission, 2022).
Legacy lock-in	Dealing with legacy systems in institutions is too great a challenge and too big a risk for the rewards on offer. Legacy systems are about more than just IT. For some long dated financial products, like residential mortgages or interest rate derivatives, changing the system might mean changing financial payments for the product itself (Bank of England, 2021).
Lack of standardization	The lack of standardization of the definitions and descriptions of data by institutions is a significant barrier to improving institutions processes (Financial Conduct Authority, 2020). Regardless of the level at which to define common data inputs, it will be hard to create them and to reach a fair agreement on what the common data inputs are and what they mean (Bank of England, 2021). For some products, there can be a lot of heterogeneity in the way institutions record data about a product, including whether they even capture certain information. This can affect calculations about the optimal level of granularity to aim for in the common input layer (Bank of England, 2020).

Reporting entities' adaptation costs	Defining and creating a new common input layer clearly involves up-front costs. The return on that investment would depend on institutions being able to meet new requests by reference to data already in the input layer. New requests that required data outside this would still generate additional costs, and require time to source and integrate the extra information (Bank of England, 2020).
Only partial feasibility	These solutions may be more feasible for some types of reporting than others. Where the transformations are deterministic aggregations of data from a clearly defined input layer, it should be easier to agree on the transformations and even to write them as code, facilitating automation. However, where the transformations reference concepts that require expert interpretation – such as dynamic references to accounting principles and regulatory constructs like capital metrics – it may be harder to agree on transformations and those transformations may require expert human inputs, such as judgements on classification or valuation (Bank of England, 2020).
One size does not fit all	The transition costs will vary substantially from one institution to another because they depend on various factors, for example institution size, the number of group companies or applicable accounting standards (BaFin, 2022). Large institutions may face additional complications in aggregating certain data across multiple legal entities, and the need to take account of factors that apply across a group such as master netting agreements, set off rights or portfolio credit mitigation (Bank of England, 2020).
Data quality management costs	Various types of relief, such as easier interpretation, improved interaction or a reduction in ad hoc queries, are offset by considerable challenges, such as more granular data quality assurance or the additional data repository (BaFin, 2022).
Regulatory costs	Under the 'instruction as a code' approach the regulator has to develop and maintain the code. This is connected with higher initial costs for the initial development that might be higher than the costs for drafting a legal text in the current form (European Commission, 2022).
Legal issues	When (parts) of the regulation are published as code, which is deployed by the reporting entities, keeping the responsibility for correct reporting with the reporting entities creates additional challenges[12]. It is therefore beneficial to remove the responsibility for correct implementation of the reporting rules from reporting entities and make them responsible for a correct ingestion of a trade into the common input layer. As this is done for the first time: • it has to be assessed whether legal basis can be provided to reporting requirements that are written as code. The 'instruction as a code' could partly replace certain regulatory text and therefore requires a sound legal framework; • further analysis is needed on whether reporting entities can be held responsible for the mapping of the common input layer to the reporting fields, as the 'instruction as a code' will provide for the mapping from the input data model to the reporting code. (European Commission, 2022)
New skills and competencies	For the development and maintenance of the code, the regulator needs to have developers to implement the reporting rules for each field, resources to maintain the IT architecture and regulatory experts who understand both the regulation and the trade records data model (European Commission, 2022).

[12] The issue of the possible assumption by authorities of the responsibility for direct production of regulatory aggregates from granular data was also discussed during the work that led to the publication of the 'EBA report on a feasibility study of an integrated reporting system under article 430c CRR' (European Banking Authority, 2021b).

The other type of solutions may be denoted as ‘smart contract-based’.

The most promising innovation in FinTech is the introduction of the smart contract. If implemented correctly this innovation can allow significant reductions in operating expenses. There will be some reductions in transaction processing costs. However, more important savings will come from reductions in analytic costs (treasury functions, business planning, risk management, financial statements, regulatory reporting, etc.)[13].

Important gains will come from the use of smart financial contracts. Their contribution is three-fold: 1) they are unchangeable and hence create a right which has all the functionalities of a property right; 2) they are self-executing when it comes to generating payment obligations; 3) the same mechanisms of the smart contract that compute a contract’s payment obligations for transaction processing can also reduce analytic costs.

The adoption of smart financial contracts will improve the efficiency of the operations of financial institutions and increase transparency and security for investors and consumers as well as for regulators.

Under the above mentioned European Commission initiative, the MRER (machine-readable and executable reporting) project team has explored various degrees of adoption of the MRER-system (European Commission, 2022). The ‘end-to-end trading system’ is used for reporting of trades as well as for the execution of the trades themselves. The smart contracts are also used for the execution of trades, based on a shared, common data model and executed on a distributed ledger. Regulatory reporting is one of the features of these smart contracts. Within the ‘end-to-end reporting system’ framework, reporting is an afterthought of trading and it is the responsibility of each reporting entity to implement and maintain an adapter that translates trade records from the domain of trading to the domain of MRER and smart contracts. Since in these options explored by the team authorities and financial institutions are already connected to a distributed ledger, it could be possible to leverage this for trading itself, as opposed to limiting its usage for reporting purposes. An end-to-end trading system is an extension to the end-to-end reporting system, as it adds a set of smart contracts that implement trading across regulated entities, without any need for individual adapters. The MRER-code interfaces with the trading code so that it can directly process trade records and generate regulatory reports according to the regulator’s needs. In such a future, MRER automatically and instantly leads to correct reporting and there is no risk of errors in the process. In case of error-free reporting, the quality of financial institutions' data (consistency, integrity, etc.) would no longer be a crucial issue in exchanging information with authorities.

The study ‘An innovative RegTech approach to financial risk monitoring and supervisory reporting’ was published by several authors (Kavassalis *et al*, 2018) with different professional background (academic, consultancy, authority)[14]. Its purpose was to propose a bearer service, which generates and maintains a ‘digital doppelgänger’ for every financial contract in the form of a dynamic transaction document (DTD). In other words, it is a standardized ‘data facility’ which automatically makes important contract data from the transaction counterparties available to the requesting authorities. This would be achieved by sharing certain elements of the DTD on a bearer service, based on a federation of distribution ledgers; such a quasi-simultaneous sharing of risk data is possible because the DTD maintains a record of state in semi-real time, and this state can be verified, also in semi-real time, by any entity with access to the distribution ledgers. The DTD provides a unique representation of each financial contract to be used at every step in the processing chain, including

[13] See also Crisanto *et al* (2020).
[14] See also Sel *et al* (2017) and Triantafyllou *et al* (2018).

all types of regulatory reporting. In terms of process logic, the reporting takes place only once and the users of the DTD (mainly authorities, but potentially also research institutions and other representatives of 'society at large') are those who define how information will be aggregated. The granular data on transactions are transmitted to the DTD supporting infrastructure where it is automatically aggregated and made available for analytical purposes. According to the authors, 'consistent transaction processing and nearly instantaneous regulatory reporting of financial transactions and their associated risks would become possible'.

In the paper 'Effective measurement of the economy in the emerging digital age' the three authors (Gross *et al*, 2021) claim that the benefit of the IReF and BIRD Eurosystem projects in terms of reducing reporting burden while improving its outcomes will increase if they are integrated into a more comprehensive vision and strategy to accompany ongoing digitalization. In the current situation, for instance, reporting agents still need to bridge the BIRD input layer with their operational systems where information on contracts, counterparties and other facts is represented in heterogeneous ways. Hence, further standardization in the representation of the operations underlying financial markets is needed for a more effective measurement in the digital age. To achieve true digitalization, standardization is needed in the foundational level of identifying and representing financial contracts, counterparties and other facts, i.e. the operational data. This will have to take place involving market participants, and possibly leveraging on-going initiatives. The digitalization of financial services opens new possibilities towards the standardization of the technical representation of financial and other contracts. Many initiatives are on-going in order to develop standards for specific financial instruments; however, such initiatives are mostly local, while an effective solution would need to be far wider, being ideally global and covering all contract types. The risk of continuing with uncoordinated and siloed initiatives is to deliver another wave of heterogeneous legacy solutions which would likely cost more. Coordination across jurisdictions is thus key. The strong opinion of the authors is that such standardization will need to be radical and that compromises will be counterproductive. Voluntary adoption could be too slow as time is a critical factor. Legislating the use of standards and public data infrastructures will likely be required. Public authorities will need to intervene and catalyze change to achieve the common good. Focus should initially be on building global infrastructures for the identification of contracts and counterparties. The authors' conclusion is that the digital transformation 'offers a unique opportunity to authorities to strengthen their catalytic role by leading the development of a comprehensive vision and the design of a conceptual architecture that would provide a frame in which local initiatives could flourish while forming a coherent system, sustainable in the digital age'.

Table 3 shows the pros and cons of the solutions based on smart contracts.

Table 3 - Pros and cons of the solutions based on smart contracts

PROS	
Automation	In the long run, the whole ecosystem is expected to benefit by the comprehensive representation of a trade record, which will be a one-off effort that can be reused in the implementation of future regulations. With the machine-executable reporting approach, a regulation can be updated much faster and with less effort for the reporting entities. Moreover, the reporting of existing trades upon a regulation update could then be integrated as an automatic process, requiring little to no human intervention (European Commission, 2022).
Standardization	The smart financial contracts (highly granular and semi-real time modelling of a financial contract's life-cycle, mirroring specific, real contracts) provide a unique presentation of each financial contract to be used at every step in the processing

	chain, including, but not limited to, all varieties of regulatory reporting (Kavassalis *et al*, 2018). Since everyone is looking only at a single record of financial contracts, 'data standardization' effectively happens at the level of the operational data. A main benefit of this approach is that it would make reconciliation efforts a thing of the past. Distributed ledger technology (DLT) has the potential to fundamentally transform the way data is managed by replacing today's practice of financial institutions keeping different records of the same transaction (Crisanto *et al*, 2020). A key aspect of the concept is that the single record of a financial contract stored on the distributed ledger may replace the separate records of all parties of the financial contract (Kavassalis *et al*, 2018).
Efficiency	Most models foresee that authorized parties, such as supervisors, would be able to connect directly to the ledger. In such a system, the current regulatory reporting process could be simply replaced by sharing relevant data aggregates with the supervisor in the DLT network (Crisanto *et al*, 2020).
Accuracy	At the same time, DLT may prevent unintentional, practical reporting errors and reporting of deliberate, e.g. fraudulent, misinformation (Crisanto *et al*, 2020).
Time-to-market	Essentially, consistent transaction processing and nearly instantaneous regulatory reporting of financial transactions and their associated risks would become possible (Kavassalis *et al*, 2018).
Consistency	Smart contracts would integrate in a fully consistent manner both risk views for monitoring and reporting purposes; as a result, they could help reduce compliance costs and progressively close the gap between the operational and the analytical departments, i.e. integrating front, middle and back office (Kavassalis *et al*, 2018).
Transparency	Overall, transparency of the global financial system will increase – firstly, benefitting supervisory authorities and regulators, secondly, financial research and legislation and, last but not the least, society at large (Kavassalis *et al*, 2018).
Compliance	Using this approach simplifies the compliance landscape that financial institutions are subjected to. Specifically, expressing the behavior of real world contracts through algorithms obsoletes a class of complex challenges related to structuring reporting data that at this point deny high fidelity reporting aggregates for the regulator (Sel *et al*, 2017).
New opportunities	This solution also opens new paths to data analytics. Stress tests could be envisaged as frequent, quasi-automated exercises flexible enough to address many questions and test diverse scenarios (Gross *et al*, 2021).
CONS	
Need for public intervention	Smart contracts cannot be expected to be provided by institutions left to themselves, creating a need for specific regulation that will initiate the new construct (Kavassalis *et al*, 2018).
Need for global vision and strategy	Coordination across jurisdictions is key. Authorities have to strengthen their catalytic role by leading the development of a comprehensive vision and the design of a conceptual architecture (Gross *et al*, 2021).
Lack of economic design of smart contracts as public good	Need of a mechanism design theory for smart contracts, i.e. identify the institutional properties of smart contracts that make them economically efficient and successful (Triantafyllou *et al*, 2018).
Lack of design principles of a network	Need a robust evolvable network supporting smart contracts creation, diffusion and sharing (in conditions of data privacy and non-disclosure of sensitive financial data) based on the combined use of: • blockchain and smart contracts; • financial analytics intelligence; • formal reasoning and methods for precise design. (Triantafyllou *et al*, 2018)

***

Overall, the outcome of these two streams of initiatives is that the potential benefits for financial institutions and authorities to move towards a digital reporting framework based on data standards outweigh, in the long run and with a number of assumptions, the initial and running costs.

However, apart from the pros and cons evaluations, it has to be underlined that the available case studies and trials are not yet convincing. It is quite evident, in fact, the attempt to demonstrate the applicability of these innovative solutions starting from simple business cases. In fact, the proof-of-concepts carried out so far have taken into consideration quite standardized financial products (e.g. derivatives or mortgage loans) and very granular (e.g. EMIR reporting) or non-complex requirements, for which the necessary transformations rules to produce the output from the input are extremely linear. But typically this case tends to be the exception rather than the rule. More frequently, writing a version of the reporting instructions in code can be very complex.

These new fields therefore need to be further explored in order to verify their practical applicability.

Nevertheless, it's worth pointing out that real-time transactional data based on smart contracts appears to be more promising for a number of reasons. First of all, they would require higher standardization of the operational system of the institutions at the global level, hence preventing a wave of heterogeneous solutions, which would make the current statistical ecosystem even more complex and somewhat chaotic. Then, this data would enable forward-looking analysis. Moreover, data would be available almost in real time. Finally, in an extreme case, it could even replace reporting (or at least part of it) since granular data could be interpreted using a standardized cash flow generation process inferred from the contractual obligations.

In parallel to further explorations, it will be crucial to meet certain organizational and strategic preconditions for these regulatory reporting initiatives to succeed, as evidenced in a study of the BIS Financial Stability Institute (Crisanto *et al*, 2020). These include:

a. strong commitment and support from top management at both financial authorities and institutions;
b. alignment of vision by engaging transparently, collaboratively and openly with key stakeholders;
c. a culture of innovation that relies on data-driven decision-making, openness to experimentation and questioning ‘legacy thinking’ within financial authorities;
d. a well-defined centralized data strategy and data governance framework within financial authorities; and
e. effective management of the transition to new regulatory reporting processes, particularly by taking a step-wise approach.

3. A more feasible and likely scenario for the future

While exploring innovative and more technology-based solutions to benefit from ‘on demand’ access to granular, timely and integrated sources of data, European authorities have identified five main areas for improvement to modernize and integrate EU reporting and put in place a system that delivers accurate, consistent, and timely data to authorities at EU and national level, while minimizing the aggregate reporting burden for all relevant parties.

Achieving this goal is feasible through gradual modifications of the current reporting frameworks and methods for collecting and processing financial data. Drawing on the lessons learned from the fitness check and further discussions with stakeholders, the EC, in its 'Strategy on supervisory data in EU financial services' (European Commission, 2021), concluded that such a reporting system should include the following main building blocks[15].

1 - Consistent and standardized data: 'Data standardization and a common understanding of the data collected under different reporting frameworks will make it easier to use digital technologies and simplify the transmission, validation, and analysis of the data. Data specifications should rely on clear and common terminology, as well as on common standards, formats and rules for the use of unique identifiers. This will avoid ambiguity in their interpretation and make compliance easier'. 'The Commission will coordinate the development of a common data dictionary to ensure consistency and standardization across the financial sectors'. 'A key element for ensuring data consistency is the full use of internationally accepted common identifiers, such as the unique product identifier (UPI), the unique transaction identifier (UTI) and the legal entity identifier (LEI)'.

2 - Data sharing and reuse: 'Facilitating the sharing and reuse of reported data among national and EU supervisors [but this holds for any other authority imposing reporting obligations] will reduce the burden on reporting entities by avoiding duplicative data requests. It will also allow supervisors [or other authorities] to overcome legal and technical barriers to using data held by another supervisor [or other authority]. Improved data sharing and reuse will require a dedicated regulatory framework and a secure IT environment to streamline data flows, validation and exchange mechanisms. It will also rely on the common specification and standardization of data sets'. 'Data should be reported only once and then shared and reused as needed by the different authorities in the EU'. 'The Commission will identify other legal obstacles to sharing of supervisory, statistical and resolution data and stands ready to address them where necessary'. 'Together with a common data dictionary, this will lay the foundations for a European data space where data can be accessed, shared and reused by all relevant authorities while protecting data confidentiality and security'.

3 - Improved design of reporting requirements: 'Well-designed and drafted legislation is essential for implementing an effective and efficient reporting system. An improved process for the design of reporting requirements should be based on current best practices in applying EU Better Regulation principles to supervisory [or other] reporting, both in the legislation and in the specification of technical standards. It should also ensure that these principles are applied consistently and systematically across reporting frameworks, from the development of the initial requirements to subsequent reviews'. 'The Commission also calls on the ESAs and other EU and national authorities to fully assess the impact of technical reporting instructions they develop'. 'Finally, the Commission will assess options, including a new legal instrument, for streamlining the current lengthy and often complex process of developing regulatory and implementing technical standards for supervisory [or other] reporting'.

4 - Joint governance: 'Designing, implementing and maintaining a modern and improved supervisory [or other] reporting system will require sufficiently robust governance arrangements. These will improve coordination and foster greater cooperation between different supervisory [or other] authorities and other relevant stakeholders, allowing them to share their expertise and exchange information'. An appropriate governance structure will 'provide technical advice on any legislative

[15] In the 'EBF response to the EBA discussion paper on the feasibility study of an integrated reporting system under article 430C CRR' the industry has represented similar views for the way forward to an integrated reporting system (European Banking Federation, 2021).

or non-legislative actions that may be needed'. 'In addition, any future governance structure should encompass the national competent authorities (NCAs) [or other authorities] and include arrangements to bring in industry expertise as needed'. 'In line with the Commission's gradual approach, and in order to minimize resource constraints and avoid a proliferation of bodies dealing with supervisory [or other] reporting, existing structures will be used to the extent possible'. 'By 2023, working together with the relevant EU authorities, the Commission will formalize the governance arrangements as the work on the strategy progresses'. 'Given the global nature of the financial system, the need for cooperation and coordination extends beyond the EU to our international partners. International alignment and more consistent reporting across jurisdictions allows for more effective supervision at global level and helps avoid arbitrage between jurisdictions'[16].

Delivering these building blocks will enable a more effective and efficient use of modern technologies, including RegTech and SupTech[17]. The use of such solutions will further reduce the compliance burden for reporting entities and increase the accuracy and timeliness of the data received by authorities and improve their capacity to analyze it.

While modernizing and improving reporting is expected to bring significant long-term benefits and cost savings, it also inevitably implies an investment of resources during the transition. Therefore, the modernization of reporting in the EU will require a gradual approach to reduce implementation risks and costs.

Specifically, the Commission will use the regular review process to introduce the following targeted improvements: '(i) removing identified overlaps or inconsistencies within a legislative act or across multiple acts; (ii) removing redundant or outdated reporting requirements; (iii) providing or clarifying specific definitions; (iv) streamlining data flows between authorities; and (v) where appropriate, strengthening the proportionality of the reporting requirements'.

Regarding the possibility to increase, where feasible, the level of granularity for the reporting requirements in the context of an integrated reporting system, as a way to further increase the efficiency of the reporting process, it has to be considered that this is not as a pre-condition for having such a system. Further investigations on the scope and possible design of such a solution and a thorough cost/benefit assessment should be conducted before any change is proposed[18]. Reporting requirements in statistical, prudential and resolution are defined using a mix of different levels of granularity, ranging from very granular (e.g., item-level data) to highly aggregated. Having in mind that there could exist some similarities in the concepts requested across frameworks, further efficiency gains could potentially be obtained if some concepts defined at a more aggregated level could potentially be derived from more granularly reported ones by means of transformations. Preliminary evidence gathered during the EBA feasibility study shows that 'in many cases it may not be cost efficient or feasible to report with a level of granularity which ensures that data is reported at one single (highest) granularity level and more aggregated views could be derived, especially in the area of prudential and resolution reporting' (European Banking Authority, 2021).

[16] 'Standard setters and authorities may wish to consider evaluating the scope for common data standards and taxonomies for relevant regulatory areas, including the potential for international collaboration, in order for reporting solutions to be made more scalable and interoperable'. (Financial Stability Board (2020), 'The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions', October.

[17] Supervisory technology (SupTech) is the use of innovative technology by supervisory authorities to support supervision. It helps supervisory authorities to monitor risk and compliance of the financial institutions more efficiently and proactively.

[18] Also the banking industry has expressed a preference for a transition to a data-driven approach from a template-based one rather than data to be reported at one single (highest) granularity level: 'A stepwise design of the transition to the data driven approach from the current template driven approach is a key matter to avoid a high-cost rework that could be the consequence of a poorly designed or hasty race for granularity' (European Banking Federation, 2021).

Even if data at the highest level of granularity is not collected by the authorities, it should be available in the input layers of the institutions and its governance and operational management is crucial to grant effective and efficient reporting processes. Institutions need to have in place and implement strong (risk) data aggregation and (risk)-reporting capabilities, which are also assessed as part of their internal governance within the SSM supervisory priorities 2023-2025[19] (which for large institutions might also include compliance with the BCBS 239 principles[20]).

To facilitate these complex tasks and to reduce the room for interpretation of reporting regulations, which have a significant impact on data quality, the Bank of Italy implemented a strategy, based on a structured and voluntary collaboration with reporting entities (PUMA), which has been producing remarkable results over the past thirty years (Casa *et al*, 2022b). In particular, 'the value of this collaboration is two-fold: *ex ante*, it makes it possible to carefully scrutinize new proposals for reporting regulations in order to identify options that favour the production of high-quality data while containing their costs; *ex post*, through the preparation of the PUMA documentation, it describes the calculation steps from banks' input data to the outputs requested by the authorities, combining participants' regulatory knowledge and operational experience with regulators' fundamental contribution to clearing any interpretation doubts. The concrete results achieved by the PUMA cooperation project have contributed to improve the quality of reporting and support banks and other financial intermediaries in the production of statistical information. Furthermore, the spirit of cooperation and frequent interactions between the industry and the regulator have enabled the Bank of Italy to become more aware of the costs borne and challenges faced by reporting entities'.

Italy's PUMA project has been an example for European authorities. In 2015, in fact, the European System of Central Banks (ESCB) launched the Banks' Integrated Reporting Dictionary (BIRD), a cooperation initiative involving a significant number of European NCBs and commercial banks[21]. Both experiences are based on the assumption that, with few exceptions (both in the statistical field, e.g. AnaCredit and SHSG, and in the supervisory one, e.g. large exposures), reporting requirements are defined by the authorities at an aggregate level and that the calculation processes and related responsibilities for the preparation of the regulatory reports remain a prerogative of reporting entities.

This kind of documentation, which contains the transformation rules to generate the aggregate data, can also be considered a sort of RegTech solution - PUMA even *ante litteram* (see Signorini, 2018) - as it constitutes an important reference point for intermediaries in applying the reporting regulations. In spite of that, PUMA-like solutions and the machine-executable ones, represented in the previous section, actually differ substantially. While both approaches require a standardized input data layer for all recipients of reporting obligations, in the case of PUMA-like documentation the timely production, consultation and use of this data layer are entirely voluntary. On the contrary, in a pure

[19] See https://www.bankingsupervision.europa.eu/banking/priorities/html/ssm.supervisory_priorities202212~3a1e609cf8.en.html .

[20] See https://www.bis.org/publ/bcbs239.htm .

[21] In light of a wider integrated reporting system in Europe, it has to be noted that the EBA feasibility study claims that further assessment of the institutions' appetite for a common approach to compliance is needed. In particular: 1) the institutions' commitment to develop a common compliance solution, 2) the institutions' commitment to apply the outcomes of the common compliance solution, 3) the role and contribution of authorities in a solution that is in the sphere of the institutions' compliance obligations, 4) the expected cost of the common input approach for compliance, 5) the ways to assure a fair playing-field for the small Fintech, Regtech and SupTech improving the market competitiveness on compliance solutions, and 6) the possible decrease on the value of a common compliance solution due to: i) the benefits of having a common regulatory data dictionary with all reporting requirements defined under the same data dictionary; ii) the impact of potentially more granular regulatory frameworks in reducing the amount and costs of transformations needed.

RegTech solution, in which the reporting instructions are published in the form of an executable code, the production of the input data is regulated and therefore constitutes an obligation for the reporting agents, with all the cost implications discussed earlier in this paper.

Conclusions

The cost of traditional approaches to the collection of data from the banking (and financial) industry is rising for both the reporting agents and the authorities; due to the increasing number of reports, the current reporting system will become progressively less efficient. An important investment in how data are collected is urgently needed, which would also contribute to ensuring adequate quality assurance, a common interest of both financial institutions and authorities, despite their different roles.

The European and international debate has identified some possible solutions. The two most innovative ones - machine-executable reporting solutions and those based on smart contracts – require a high level of standardization and granularity, as well as a massive use of technology; as such, they would have a huge impact on the status quo for both authorities and institutions. A less disruptive approach points to the gradual standardization and harmonization of the current reporting system and is based on better governance and a general rationalization of the existing information requirements.

The two innovative approaches seem premature, mainly because of the insufficient standardization in the operational systems of financial institutions; they also imply a deep change in the roles and responsibilities of the different stakeholders in data processing, and some major legal obstacles must be overcome for their implementation. In addition, the available empirical case studies and trials are too simplistic and therefore not yet convincing. Smart contracts seem to be a more promising area for further investigation; they might lead to potentially huge savings and could make information available in near real-time. This is why it could be worth investing in proof-of-concepts for specific business cases, possibly in the context of international cooperation (involving or with a leading role for the Financial Stability Board and the Bank for International Settlements) and with the due involvement of the industry and the end users of the data.

A feasible path for changing the reporting system has already been undertaken at European level. It envisages the creation of an integrated reporting system for European banks (to then be extended to other segments of the financial market) by establishing appropriate governance and by standardizing, harmonizing and rationalizing the existing European and national frameworks, while leveraging technology as much as possible.

Dialogue and joint work with the banking industry will be crucial in this process. In this regard, continuing to invest in cooperation initiatives on the more operational and technical part of reporting, such as the strategic BIRD project launched and promoted by the ESCB, will provide fundamental support for maintaining a close connection between the data production of the reporting agents and the data collection of the authorities.

References

BaFin (2022): 'A reporting system for the future - Feasibility study on 'Redesign Options for Regulatory Reporting' (short version)', July.

Bank for International Settlements Innovation Hub and Monetary Authority of Singapore (2022): 'Project Ellipse - An integrated regulatory data and analytics platform', March

Bank of England (2020): 'Transforming data collection from the UK financial sector', Discussion Paper, January

Bank of England (2021): 'Transforming data collection from the UK financial sector: a plan for 2021 and beyond', February.

Casa, M., Carnevali, M., Giacinti, S. and R. Sabatini (2022a): 'PUMA cooperation between the Bank of Italy and the intermediaries for the production of statistical, supervisory and resolution reporting', Issues of Economics and Finance No. 734, Banca d'Italia, November.

Casa, M., Graziani Palmieri, L., Mellone, L. and F. Monacelli (2022b): 'The integrated approach of the Bank of Italy to the collection and production of credit and financial data', Issues of Economics and Finance No. 667, Banca d'Italia, February.

Crisanto, J.C., Kienecker, K., Prenio, J. and E. Tan (2020): 'From data reporting to data-sharing: how far can suptech and other innovations challenge the status quo of regulatory reporting?', BIS - FSI Insights on policy implementation, No. 29, December.

European Banking Authority (2021a): 'Study of the cost of compliance with supervisory reporting requirements report', EBA/REP/2021/15.

European Banking Authority (2021b): 'EBA report on a feasibility study of an integrated reporting system under article 430c CRR', EBA/REP/2021/38.

European Banking Federation (2018): 'Data reporting: European banks underline the need for an integrated and standardised EU framework', EBF Press Release, October.

European Banking Federation (2021): 'EBF response to the EBA discussion paper on the feasibility study of an integrated reporting system under article 430c CRR', EBF Press Release, June.

European Central Bank (2020): 'The ESCB input into the EBA feasibility report under article 430c of the Capital Requirements Regulation (CRR 2)', September.

European Commission (2019): 'Results of the fitness check of supervisory reporting requirements in EU financial services legislation', November.

European Commission (2021): 'Strategy on supervisory data in EU financial services', December.

European Commission (2022): 'MRER Proof of Concept: Assessing the feasibility of machine readable and executable reporting for EMIR (Final report)', April.

Financial Conduct Authority (2020): 'Digital Regulatory Reporting - Phase 2 Viability Assessment', January.

Gross, F., Colangelo, A. and F. Schuster (2021): 'Effective measurement of the economy in the emerging digital age', IFC Bulletin No. 55, November.

Kavassalis, P., Stieber, H., Breymann, W., Saxton, K. and F.J. Gross (2018), 'An innovative RegTech approach to financial risk monitoring and supervisory reporting', The Journal of Risk Finance, Vol. 19 No. 1, 2018, pp. 39-55.

Sel, M., Diedrich, H., Demeester, S. and H. Stieber (2017), 'How smart contracts can implement "report once"', Data for Policy 2017: Government by Algorithm (Data for Policy), London, 6-7 September.

Signorini, L. F. (2018), Introductory speech at the workshop 'Tra segnalazioni nazionali e reporting armonizzato europeo: rafforzare la cooperazione tra gli intermediari e le autorità' (Between national and harmonised reporting in Europe: strengthening the cooperation between reporting agents and authorities), May.

Triantafyllou, N., Ksystra, K., Stefaneas, P. and P. Kavassalis (2018), 'Blockchain and Financial Risk Reporting: design principles and formal reasoning', Artificial Intelligence in Industry and Finance (3rd European COST Conference on Mathematics for Industry in Switzerland), Winterthur, 6 September.

BANCA D'ITALIA
EUROSISTEMA

Connecting the dots of the international debate on the standardization and granularity of regulatory data

Massimo Casa
Senior Advisor - Banca d'Italia

BoC-IFC satellite seminar on "Granular data: new horizons and challenges for central banks"
Ottawa, 15 July 2023



Agenda



1. Current situation and its criticalities
2. The ongoing debate on standardization and granularity
   - STANDARDIZATION AND GRANULARITY
   - OPERATIONAL DATA STANDARDIZATION
   - INPUT LAYER STANDARDIZATION
3. Output layer standardization
4. Conclusions

CURRENT SITUATION

Proliferation of surveys

Survey	Authority	Granularity	Business scope
BSI/MIR	ECB	Aggregated	Assets and liabilities
AnaCredit/SHSG	ECB	Item-level	Loans and securities
Payments statistics	ECB	Aggregated	Payments services
EBA-ITS on Sup. Rep.	EC + EBA	Aggregated	Acc/Prud/Res data
SRB reporting	SRB	Aggregated	Resolution data
MMSR	ECB	Transactional	Money market
EMIR reporting/SFTR	EC + ESMA	Transactional	Derivatives/SFT
BIS loc/cons statistics	BIS	Aggregated	Assets and liabilities
National reporting	National authority		
Ad hoc data collect.	Competent authority		

Lack of effective *ex ante* coordination

MAIN CRITICALITIES

- Sharp increase in the reporting costs
- Pressure on the current system is never ending
- Cooperation between the authorities has to be strengthened
- Regulatory reporting is not suited to FinTech-induced changes in the financial landscape

STANDARDIZATION AND GRANULARITY



Since 2016 there has been **a proliferation of studies trying to find solutions** to the these criticalities. Three main streams of discussion:

- **operational data standardization** - granularity at the level of operational data (**smart contracts** in pure business language);
- **input layer standardization** - granularity at the level of **common input data** (business-oriented in its design but also contaminated by regulatory definitions);
- **output layer standardization - item-level reporting combined with** multidimensional structures for more **aggregated data**.

OPERATIONAL DATA STANDARDIZATION (1/2)

The prerequisite of these studies is the use of **smart financial contracts**. These are **computer programs** that can run on a ledger. They represent and execute multi-party workflows, such as trading and reporting.

In the **EC machine-readable and executable reporting** end-to-end trading system the **smart contracts are used for the execution of trades**, based on a common data model and executed on a distributed ledger.

Other solutions propose a bearer service, which generates and maintains a "digital doppelgänger" for every financial contract in the form of a **dynamic transaction document (DTD)**; it is a standardized "data facility" which automatically makes important contract data from the transaction counterparties available to relevant authorities.

OPERATIONAL DATA STANDARDIZATION (2/2)

PROS	CONS
Automation	Need for public intervention
Operational data standardization	Need for global vision and strategy
Efficiency	Lack of economic design of smart contracts as public good
Accuracy	Lack of design principles of a network
Transparency	Some data more difficult to standardize
Time-to-market	Legal issues
Consistency	New skills and competencies
Transparency	
Compliance	
New opportunities	

INPUT LAYER STANDARDIZATION (1/2)

The most interesting aspect of these studies (**EC MRER, BoE/FCA Digital Regulatory Reporting, BIS Ellipse, BaFin study**) is that a reporting regulation would be drafted in a form that allows for its automatic implementation by the intermediaries ('instruction as a code').

Two main features characterize these new regulatory reporting solutions:

- a very **high level of granularity** of the information requests from the authorities;
- the **regulatory authorities bear the responsibility to carry out the transformations** of input layer data into the aggregated information necessary for their analyses.

INPUT LAYER STANDARDIZATION (2/2)

PROS	CONS
Consistency across reports	Implementation costs
Consistency across institutions	Adaptation costs
Flexibility	Lack of incentives
Compliance	Hetereogeneous data sources
Transparency	Legacy lock-in
Regulatory efficiency	Only partial feasibility
Common language	DQM costs
	Regulatory costs
	Legal issues
	New skills and competencies

SOME ADDITIONAL CONSIDERATIONS



It has to be underlined that:

- the **available case studies and trials are not yet convincing**. The proof-of-concepts carried out so far have taken into consideration quite standardized financial products and very granular or non-complex requirements;
- **final users (but also the industry) are not yet deeply involved** in these explorations that would change dramatically the way they should approach data;
- **these innovative solutions have been explored in a little or uncoordinated way between authorities**, there is a lack of a global reflection of the pros/cons.

OUTPUT LAYER STANDARDIZATION…

EU authorities have already identified **some areas fo** **ement** to modernize and integrate the current EU reporting fr **se are the** main building blocks:

1. **Consistent and standardized data**
2. **Data sharing and reuse**
3. **Improved design of report** **nts**
4. **Joint governance**

This system will **mpliance burden, increase the accuracy and timeliness of** **improve analytical capabilities**.

As the **ements** imply implementation risks and costs, a **gradual approac** **eeded**.

THE HIGHEST LEVEL OF GRANULARITY FOR THE REPORTING REQUIREMENTS IS NOT A PRE-CONDITION FOR HAVING SUCH A SYSTEM

…IN COOPERATION WITH THE INDUSTRY

Dialogue and joint work with the banking industry is crucial.

Even if **data at the highest level of granularity** are not collected by the authorities, they **are available in the input layers of the institutions'** and their governance and operational management is critical to grant effective and efficient reporting processes.

To facilitate these complex tasks and to reduce the room for interpretation of reporting regulations, we need to **invest further on Banks' Integrated Reporting Dictionary (BIRD)**, one of the pillars of the ESCB strategy for standardizing banks' reporting.

It is a fundamental support for maintaining a **close connection between the data production of the reporting agents and the data collection of the authorities**.

OUTPUT LAYER STANDARDIZATION - NEXT STEPS

- Development of a common data dictionary
- Explorations of possible areas for semantic integration
- Integration of reporting requirements
- Set up of a Joint Bank Reporting Committee
- Strengthen BIRD operational tasks

CONCLUSIONS



- **Full support to the output layer standardization**, which seems to be the most pragmatic and feasible solution in the medium term in Europe.

- **Continue exploring solutions for standardizing operational data**, as these could be real game changers.

- **Considering carefully the path of standardizing the input layer**, as it seems to add complexity to the current situation and require high costs.

REFERENCES

- BaFin (2022): “A reporting system for the future - Feasibility study on ‘Redesign Options for Regulatory Reporting’ (short version)”, July.
- Bank of England (2021): “Transforming data collection from the UK financial sector: a plan for 2021 and beyond”: London, February.
- BIS Innovation Hub (2022): “An integrated regulatory data and analytics platform”, March.
- European Banking Authority (2021), “EBA report on a feasibility study of an integrated reporting system under article 430c CRR”, EBA/REP/2021/38.
- European Central Bank (2020), “The ESCB input into the EBA feasibility report under article 430c of the Capital Requirements Regulation (CRR 2)”, September.
- European Commission (2021): “Strategy on supervisory data in EU financial services”: Brussels, December.
- European Commission (2022): “MRER Proof of Concept: Assessing the feasibility of machine readable and executable reporting for EMIR (Final report)”: Brussels, April.
- Gross, F.J., Colangelo, A. and F. Schuster (2021): “Effective measurement of the economy in the emerging digital age”: The Hague, 63rd ISI World Statistics Congress (WSC) / IFC-sponsored Invited Paper Sessions (IPS) related to new developments in central bank statistics, July
- Kavassalis, P., Stieber, H., Breymann, W., Saxton, K. and F.J. Gross (2018), “An innovative RegTech approach to financial risk monitoring and supervisory reporting”, The Journal of Risk Finance, Vol. 19 No. 1, 2018, pp. 39-55.



BANCA D'ITALIA
EUROSISTEMA

THANKS FOR YOUR ATTENTION

Massimo Casa
massimo.casa@bancaditalia.it

BoC-IFC satellite seminar on "Granular data: new horizons and challenges for central banks"
Ottawa, 15 July 2023

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

International trade in creative economy: an analysis with granular data[1]

Andrea Carboni, Marco Di Bella and Flavia Ciccioli,
Bank of Italy

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

GRANULAR DATA NEW HORIZONS AND CHALLENGES FOR CENTRAL BANKS

International Trade in Creative Economy: an analysis with granular data

Andrea Carboni – Marco Di Bella – Flavia Ciccioli

OTTAWA 15 July 2023

BANCA D'ITALIA
EUROSISTEMA

AGENDA

- Definitions: from creative economy to creative products
- Trade in creative products in Italy:
  - Macrodata evidences
  - Granular data analysis (export in creative services)
- Some conclusions

What is creativity?



What is creativity?

A commonly accepted classification is to consider three areas of creativity development:

- **artistic creativity** involves imagination and a capacity to generate original ideas and novel ways of interpreting the world, expressed in text, sound and image;
- **scientific creativity** involves curiosity and a willingness to experiment and make new connections in problem solving; and
- **economic creativity** is a dynamic process leading towards innovation in technology, business practices, marketing, etc., and is closely linked to gaining competitive advantages in the economy.

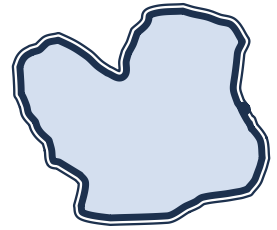
Kea European affairs Microsoft Word - final_version_020709.doc (keanet.eu) and UNCTAD (2008) ditc20082cer_en.pdf (unctad.org)

Creative industry and creative economy

The **creative industry** refers to a broad range of sectors and economic activities that are based on the creation, production, and distribution of goods and services that have a high degree of creativity and artistic value



Creative economy is a broader concept, that encompasses not only the creative industry but also the economic and cultural impact generated by creativity in general

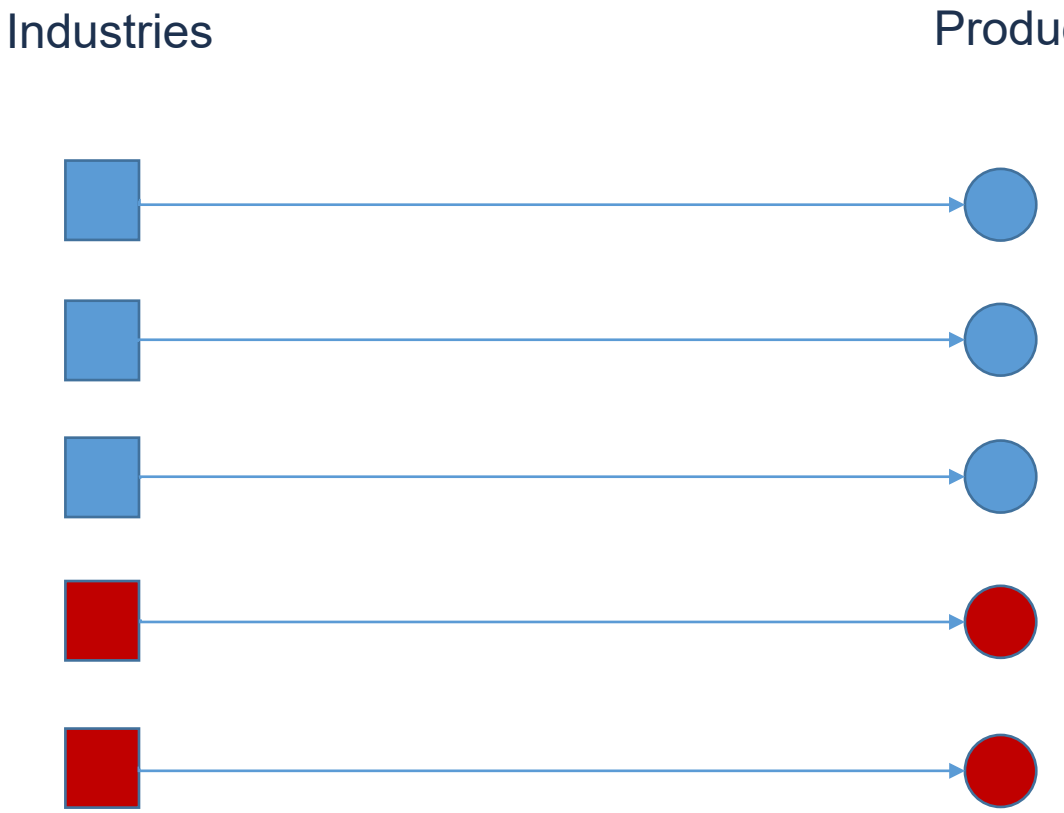


How measure creativity (with our instruments)?



Definitions: from creative economy to creative products

Mono-product economy



Creative Industries and products



Other (non- creative) industries and products

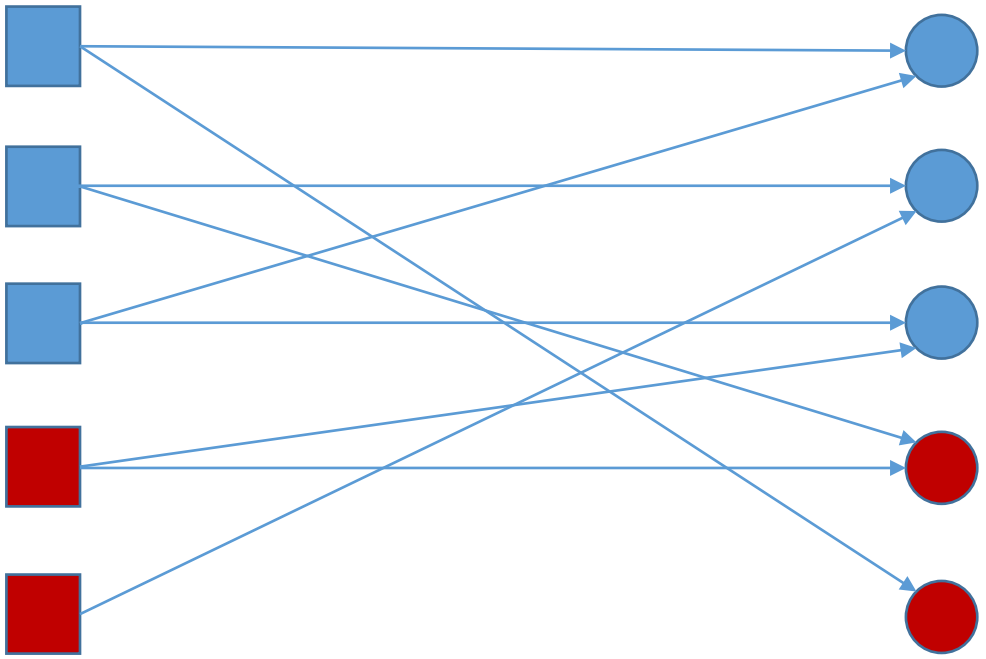


Definitions: from creative economy to creative products

Globalization and diversification economy

Industries

Products



Creative Industries and products



Other (non- creative) industries and products



Definitions: from creative economy to creative products

Product classification: goods

UNCTAD has a list of 197 creative goods identified at the Harmonized System (HS) 6-digit level that fall under the following sub-categories:

- *art crafts* (carpets, products related to celebration, other crafts, paperware, wickerware, and yarn);
- *audio-visuals* (films and CDs, DVDs and tapes);
- *design* (architecture, fashion, glassware, interior, jewellery, and toys);
- *new media* (recorded media and video games);
- *performing arts* (musical instruments and printed music); *publishing* (books, newspapers, and other printed matter); and
- *visual arts* (antiques, painting, photography, and sculpture).

For a detailed list of the creative economy product groups and hierarchy see https://unctadstat.unctad.org/EN/Classifications/DimHS2012Products_Creatives_Hierarchy.pdf.

Definitions: from creative economy to creative products

Product classification: services

A precise classification (based on E-Bops) is annually provided by UNCTAD in its Annual Report

Code	EBOPS description
SH	CHARGES FOR THE USE OF INTELLECTUAL PROPERTY N.IE.
SI	TELECOMMUNICATIONS, COMPUTER AND INFORMATION SERVICES
SI3	INFORMATION SERVICES
SI31	NEWS AGENCY SERVCIES
SI32	OTHER INFORMATION SERVICES
SJ	OTHER BUSINESS SERVICES
SJ1	RESEARCH AND DEVELOPMENT SERVICES
SJ11	WORK UNDETAKEN ON A SYSTEMATIC BASIS TO INCREASE THE STOCK OF KNOWLEDGE
SJ111	PROVISION OF CUSTOMISED AND NON-CUSTOMISED R&D SERVICES
SJ112	SALE OF PROPRIETARY RIGHTS ARISING FROM R&D
SJ12	RESEARCH AND DEVELOPMENT SERVICES OTHER THAN WORK UNDERTAKEN ON A SYSTEMATIC BASIS TO INCREASE THE STOCK OF KNOWLEDGE
SJ2	PROFESSIONAL AND MANAGEMENT CONSULTING SERVICES
SJ22	ADVERTISING, MARKET RESEARCH AND PUBLIC OPINION POLLING
SJ3	TECHNICAL, TRADE-RELATED AND OTHER BUSINESS SERVICES
SJ31	ARCHITECTURAL, ENGINEERING, SCIENTIFIC AND OTHER TECHNICAL SERVICES
SJ311	ARCHITECTURAL SERVICES
SK1	AUDIOVISUAL AND RELATED SERVICES
SK2	OTHER PERSONAL , CULTURAL AND RECREATIONAL SERVICES
SK23	HERITAGE AND RECREATIONAL SERVICES

Definitions: from creative economy to creative products

Product classification



HS Nomenclature (Convention on the Harmonized Commodity Description and Coding System) is adopted for International Merchandise Trade (or Foreign Trade Statistics)



EBOPS classification is used for balance of payments and ITSS (International Trade in Services Statistics)

FTS and BOP goods are different from a conceptual perspective, but highly correlated

Definitions: from creative economy to creative products

Product classification: the sources



In Italy, NSI provides data about international trade statistics with the HS breakdown…

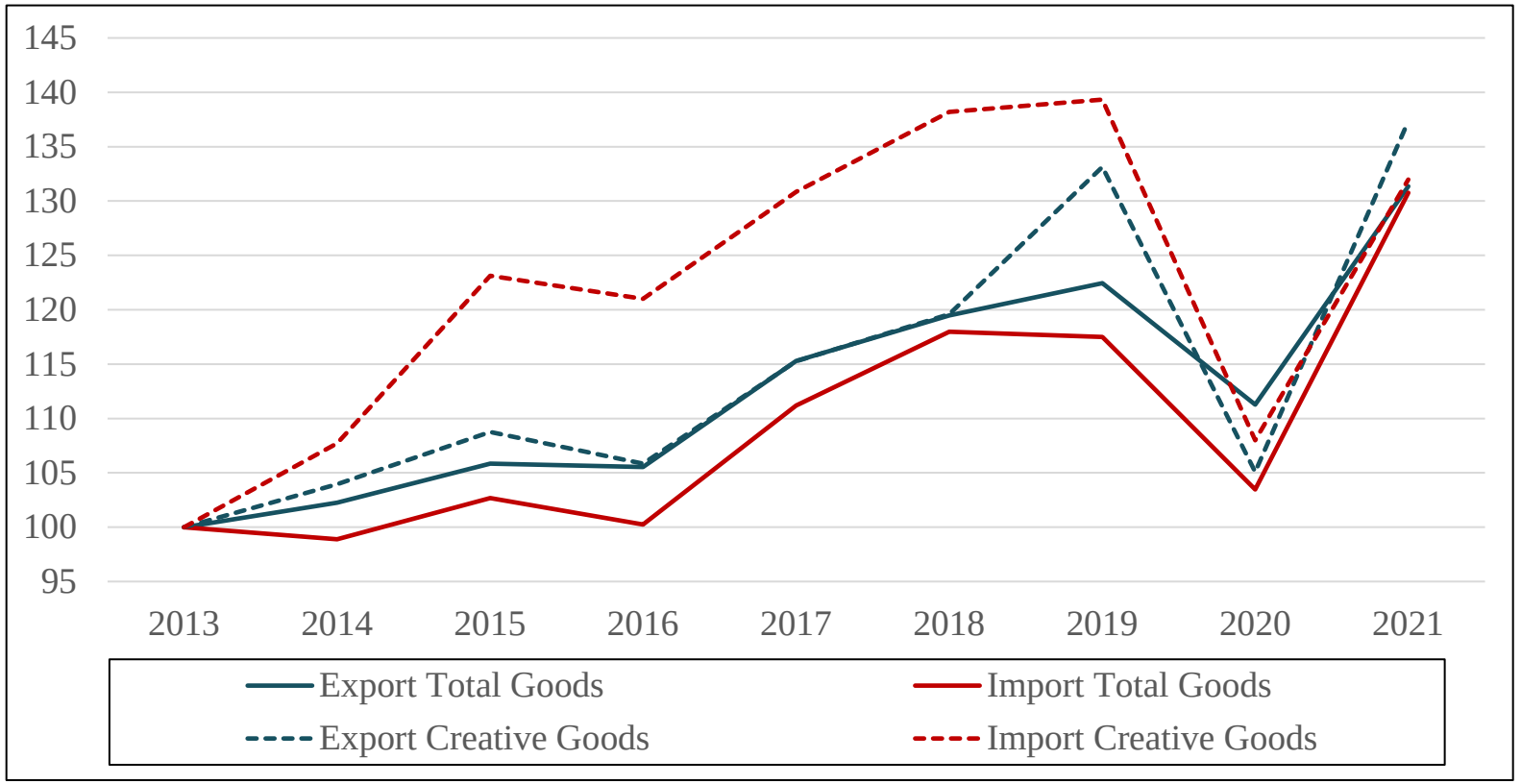
While Banca d'Italia collects the trade in services EBOPS classification, (adopted for balance of payment and ITSS) with an extensive survey



Granular data for trade in services statistics are available

The trade of creative products in Italy: macroeconomic evidence

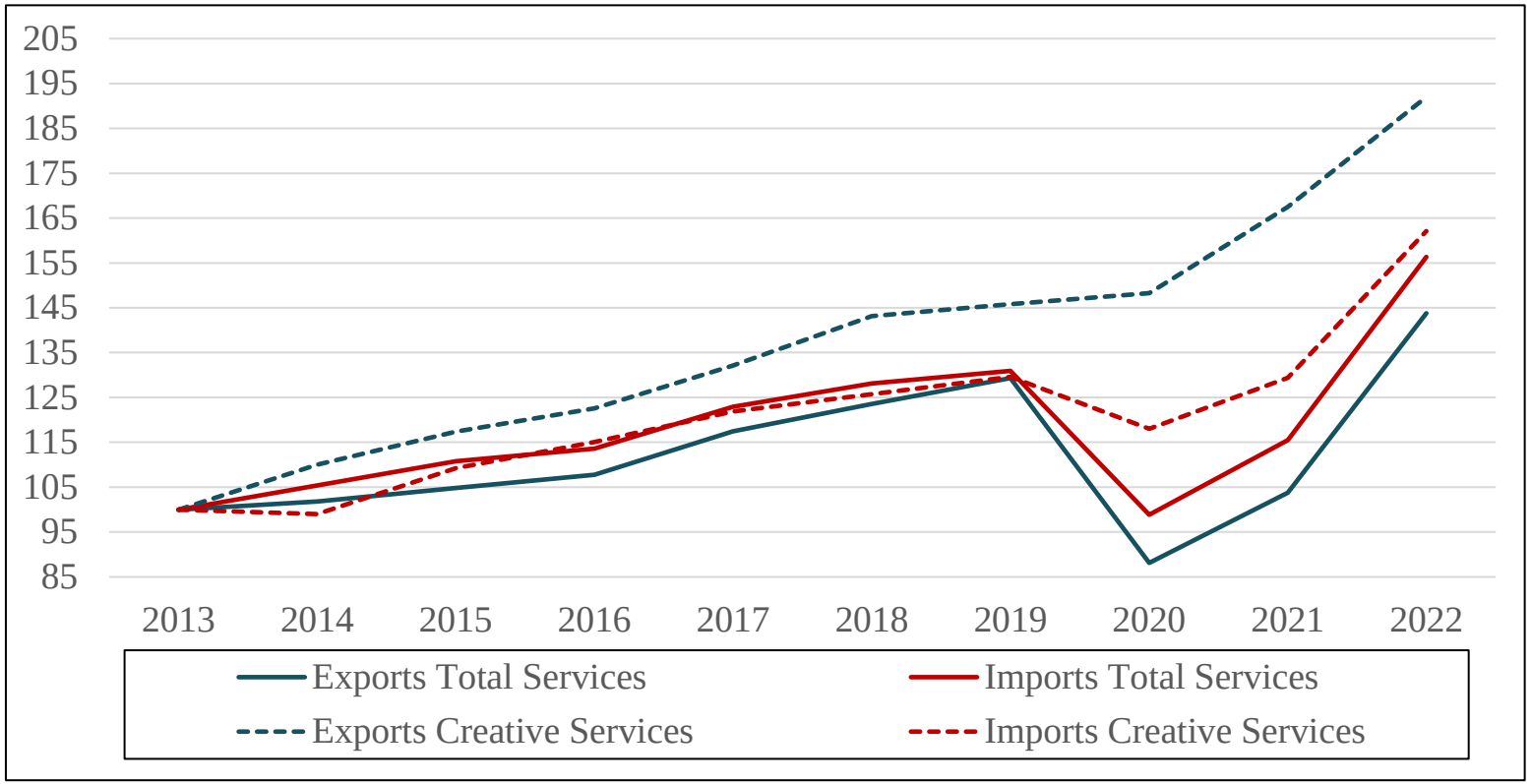
Creative Goods: exports and imports in Italy (2013 base 100)



Foreign Trade Statistics (Istat)

The trade of creative products in Italy: macroeconomic evidence

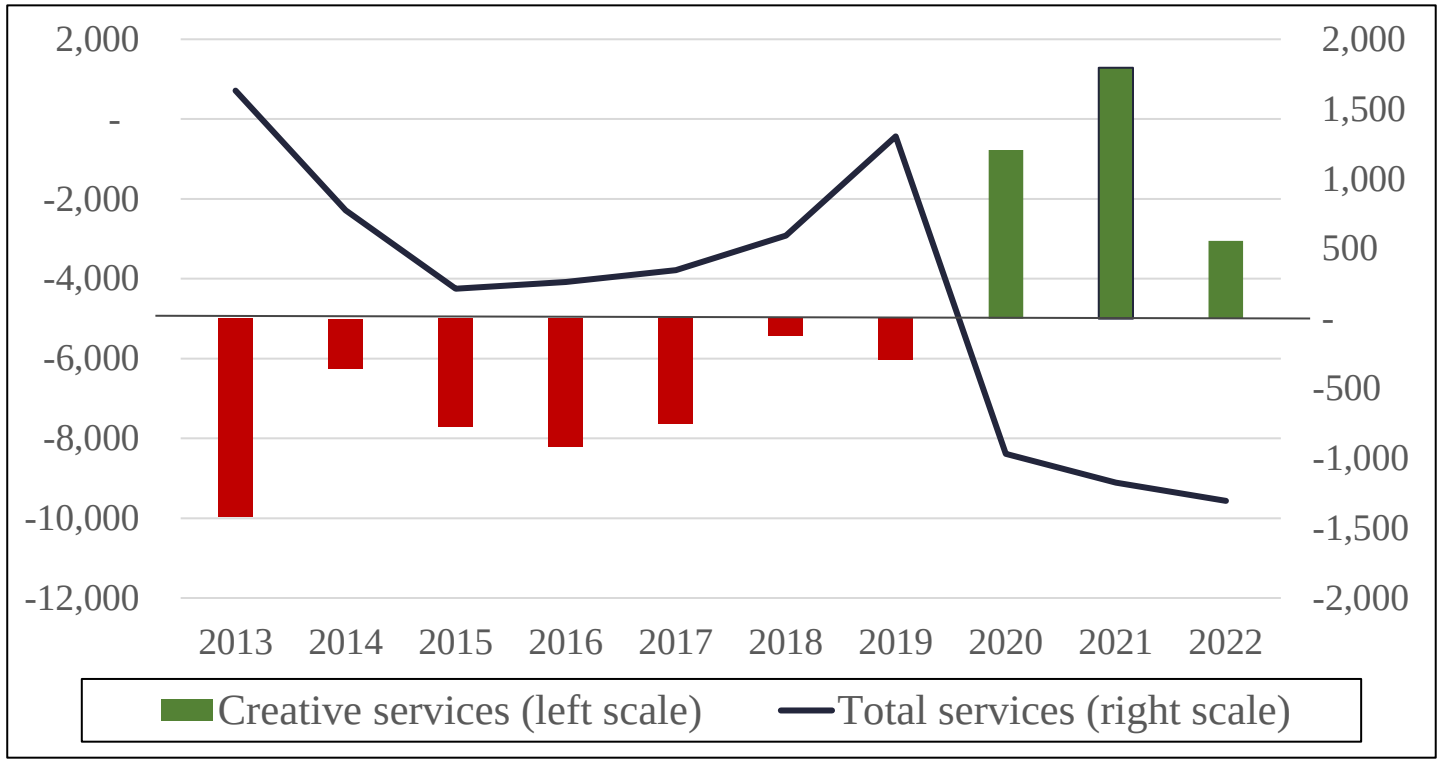
Creative Services: exports and imports in Italy (2013 base 100)



Balance of payments (Banca d'Italia)

The trade of creative products in Italy: macroeconomic evidence

Total Services and Creative Services: the balance



The large growt of export in creative services has generate a positive surplus on the last three years in Italy. At the opposite for the other services the balance is negative.

Balance of payments (Banca d'Italia)

The trade of creative products in Italy: granular data analysis

The Bank of Italy ITSS survey

A ML approach based on a logit model

The trade of creative products in Italy: granular data analysis

The Bank of Italy Bop survey for non-financial transactions

Quarterly frequency

A sample of around 3,500 non financial enterprises (increased to 5,000 units in 2023). The response rate is around 95%

Import / Export in services: breakdown counterpart country – invoicing currency

Several stratification variables: dimension (total turnover), FDI , cross-border payments (through a resident bank).

The sample is renewed every 2 years, with a rotation of around 20% enterprises.

The trade of creative products in Italy: granular data analysis

The enterprises that led the growth

Around 1,400 enterprises have participated to the survey from 2017 to 2022.

166 companies have registered a growth rate greater than 30 per cent in the export in creative services in this period.

These companies accounted for approximately 80 percent of exports in creative services.

WHO ARE THESE COMPANIES?



The enterprises that led the growth: a logit model

We developed a logit model to identifying the main features of the 166 enterprises that led the growth in creative services exports.

We tested several approaches included a ML model (a training and a validation dataset).
We have used balance samples: half units in target and half not.
In general the correct classification is around 80%; for our purposes (the companies' description) is enough

The trade of creative products in Italy: granular data analysis



WHO ARE THESE COMPANIES?

- These companies are mainly located in the northeast and central regions of Italy.
- They usually have FDI outward at the opposite of the other enterprises in the sample (while FDI inward are less relevant for distinguishing them from other companies).
- They import services.
- In general, they have recorded a significant growth in number of employees over the past 5 years
- These companies are usually of large size, particularly focusing on the turnover range between 500 million and 2.5 billion Euros. It is rare for them to have a turnover below 10 million Euros.

The trade of creative products in Italy: granular data analysis

WHO ARE THESE COMPANIES?



- They do not belong to a specific sector of economic activity, but typically they are associated with the following sectors:

NACE	*Description*
53	Postal and courier activities
91	Libraries, archives, museums and other cultural activities
60	Programming and broadcasting activities
74	Other professional, scientific and technical activities
18	Printing and reproduction of recorded media
13	Manufacture of textiles
59	Motion picture, video and television programme production, sound recording and music publishing activities
93	Sports activities and amusement and recreation activities
66	Activities auxiliary to financial services and insurance activities
32	Other manufacturing

Not all sectors are related to creative industries (i.e. Postal services, textile sector, and other manufacturing industries.)

Some conclusions

In recent years, Italy has experienced a particularly intense international trade of creative products.

In particular, the export of creative services has been significant.

It has been possible to analyze the companies that have contributed the most to this growth and outline their major characteristics (size, geographical, Nace, FDI, turnover growth)

Their challenge for the future is to understand whether the current business model will need to undergo changes due to their remarkable development.

Some conclusions

Macro-economic statistics have always provided an essential contribution to policy choices.

The analysis of granular data has also become important in providing a complementary view of the economy.

The integration of trade data and business statistics is an urgency for the statistical community (EU-EBS Regulation).

Central banks (as BoP compilers) will need to seek granular information (i.e. trade in services surveys, customs data) to develop a comprehensive economic analysis.

GRANULAR DATA NEW HORIZONS AND CHALLENGES FOR CENTRAL BANKS

International Trade in Creative Economy: an analysis with granular data

THANK YOU !!!

andrea.carboni@bancaditalia.it

OTTAWA 15 July 2023

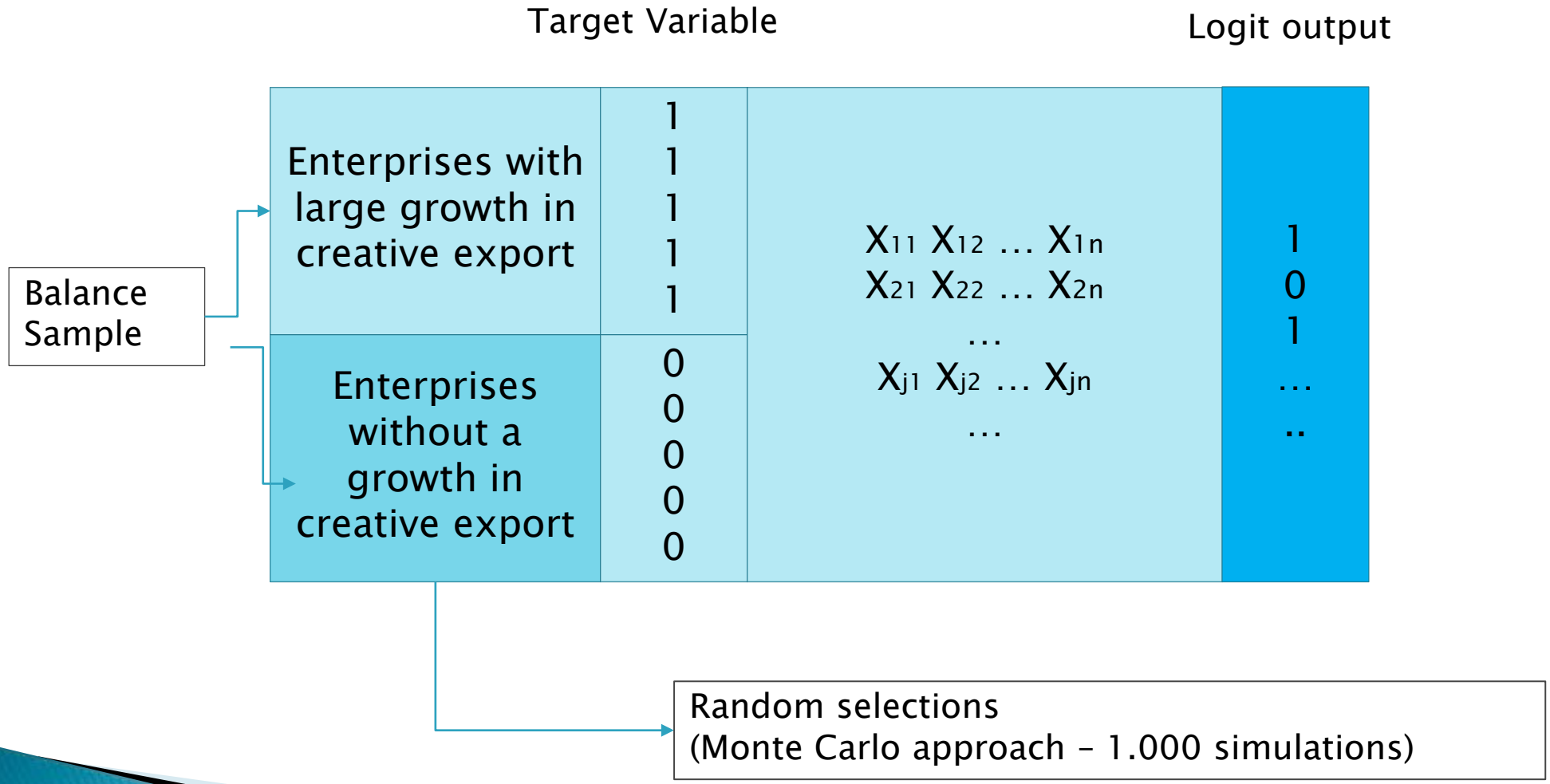
BANCA D'ITALIA
EUROSISTEMA

ADDITIONAL SLIDES

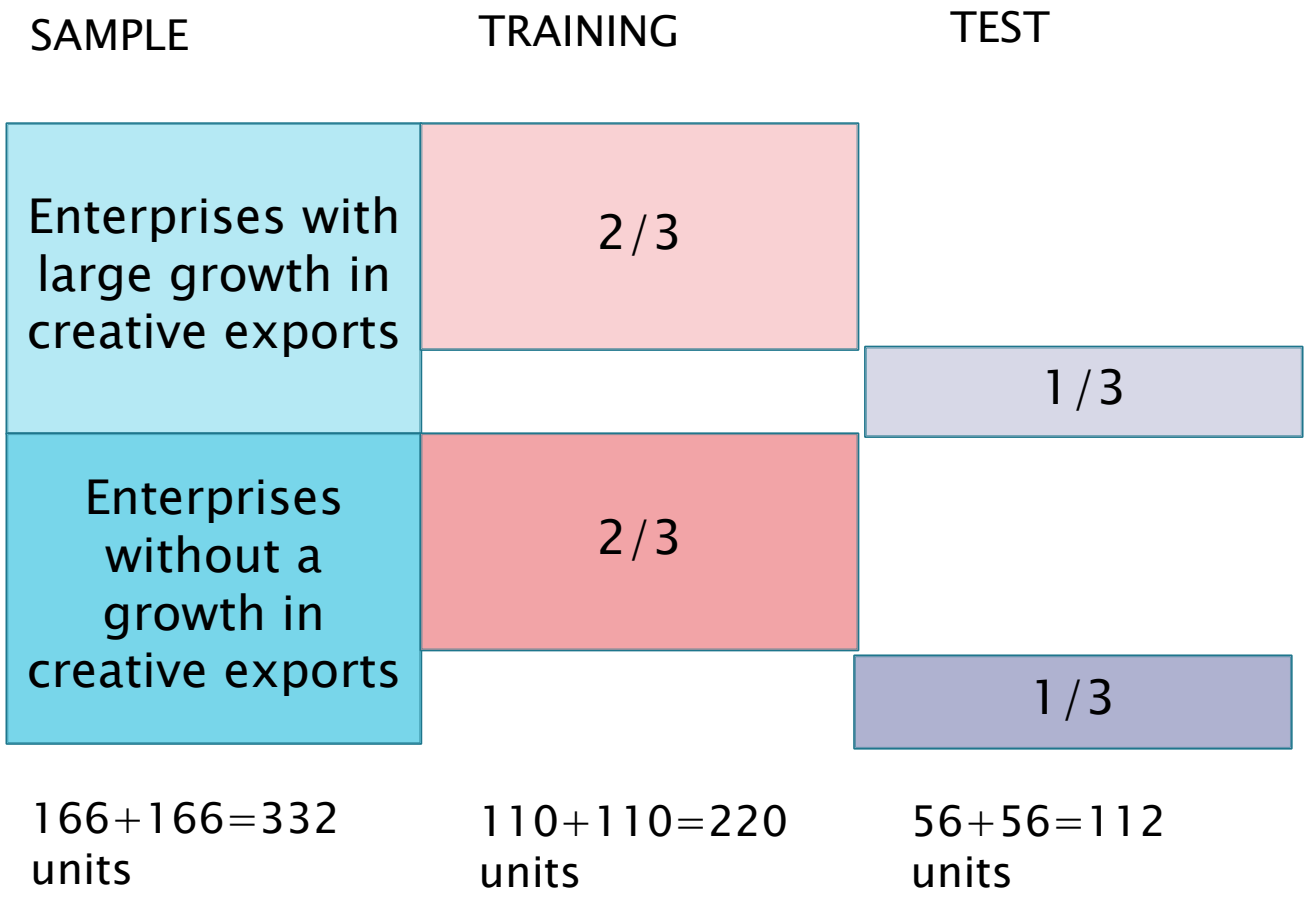
Table 11 – Number of enterprises with significant growth rate in creative services exports in the period 2017-2022
Source: ITSS Survey (Banca d'Italia)

(1) Growth rate (2017-2022)	(2) Number of enterprises	(3) % of enterprises	(3) Share of export in creative services (2022)
30	166	12.1	78.3
50	137	10.0	71.3
100	102	7.4	48.8

ADDITIONAL SLIDES



ADDITIONAL SLIDES



ADDITIONAL SLIDES

The corner point approach

Class	Description	Parameter
1	Turnover lower than 10 million euro	-0,39417483
2	Turnover between 50 and 100 million euro	-0,732568843
3	Turnover between 100 and 500 million euro	0,019538032
4	Turnover between 500 million and 1 billion euro	0,123339257
5	Turnover between 1 and 2.5 billion euro	0,453097813

For class 6 -enterprises with sales greater than 2.5 billion- the parameter is equal to 0 (this is the corner!).

Thus the coefficients in the table are positive if class membership increases the probability of being an enterprise with large creative export growth (target=1) compared to class 6; negative if it decreases the probability.

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Statistical Disclosure Control (SDC) for results derived from combined confidential microdata[1]

Jannick Blaschke, Christian Hirsch and Robin Kollmann,
Deutsche Bundesbank

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Statistical Disclosure Control (SDC) for results derived from combined confidential microdata

Jannick Blaschke, Christian Hirsch, Robin Kollmann[1]

Abstract

Most researchers combine two or more datasets in their projects. In addition to technical feasibility and the meaningfulness of content, it is also necessary for researchers working with confidential Bundesbank microdata to consider possible consequences for Statistical Disclosure Control (SDC). This paper highlights three possible SDC challenges that may arise from appending and merging datasets: (1) loss of rows, (2) newly generated missing values in SDC variables, and (3) newly generated duplicates. It then presents possible solutions and provides simplified examples and rules of thumb.

Keywords: statistical disclosure control, sdc, output control, research data centre, rdc, appending, merging, joining

[1] The authors gratefully acknowledge the contributions from our colleagues at the RDSC and Matthias Gomolka. All views expressed in this report are the authors' personal views and do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem. The paper was completed while Kollmann was an intern at the RDSC of Deutsche Bundesbank.

Contents

Statistical Disclosure Control (SDC) for results derived from combined confidential microdata 1

1. Introduction 3

2. Brief theoretical foundation 3

2.1 Appending 4

2.2 Merging 4

3. What hampers SDC? 6

3.1 Challenge 1: Loss of rows 7

3.2. Challenge 2: Missing values in SDC variables 8

3.3 Challenge 3: Duplicates 10

4 When does appending or merging hamper SDC? 11

4.1 Appending 11

4.2 Merging 12

5. A rule of thumb to resolve the SDC challenges 13

5.1 Appending 14

5.2 Merging 14

Glossary 16

References 17

1. Introduction

Most empirical research papers contain results based on multiple datasets. We distinguish two kinds of data operations in this paper. "Appending" datasets, i.e. extending a dataset with additional rows from another dataset, and "merging" datasets, i.e. combining columns from two datasets.

Whenever merging two datasets, researchers must consider not only technical feasibility (i.e. that there is at least one overlapping column in both datasets that can be used to identify the rows that should be combined), but also the meaningfulness of the resulting dataset. If, for example, we are merging two datasets with company information, we should first ask ourselves whether the companies in both datasets are defined in the same way. E.g. if one dataset contains information on the entire group and the second only on the German part, the company names may be the same, but caution is advised when interpreting the variables of the merged dataset.

In addition, when working at the Research Data and Service Centre (RDSC) of the Deutsche Bundesbank, all possible effects on Statistical Disclosure Control (SDC) must be kept in mind at all times[2]. Results that do not pass SDC cannot be taken out of the RDSC's secure environment or used in publications. Confidential microdata differ in this regard from other data sources where researchers can focus exclusively on obtaining results.

The purpose of this paper is to raise awareness among researchers that working with confidential microdata means focusing on two equally important objectives: (i) achieving interesting research findings while also (ii) checking that these results comply with SDC. Furthermore, these two objectives arise concurrently as researchers who only focus on results usually risk time-consuming retroactive adjustments to their code to ensure compliance.

This paper is structured as follows. In Section 2, we briefly introduce the concepts of appending and merging. Section 3 explains the three SDC challenges loss of rows, missing values in SDC variables and newly generated duplicates, and presents a solution for each challenge. Bringing together the two previous sections, Section 4 shows under which conditions, these challenges occur when appending or merging datasets. Section 5 concludes with some rules of thumb.[3]

2. Brief theoretical foundation

Before we take a detailed look at the challenges of combining different datasets, we distinguish the most important terminologies from each other. We are aware that the shape of the resulting dataset can be influenced by additional options in the append

[2] More information on the RDSC's SDC rules can be found in the the "*Rules for visiting researchers at the RDSC*" (Research Data and Service Centre (2021)), which are available at https://www.bundesbank.de/resource/blob/826176/ffc6337a19ea27359b06f2a8abe0ca7d/mL/2021-02-gastforschung-data.pdf

[3] The technical report available from the RDSC website also contains two appendices with further examples and a formal proof of the statements of Table 8.

or merge command[4]. However, for the sake of clarity, we assume that the respective commands are used in their simplest form.

2.1 Appending

When a dataset B is added from below to a dataset A, we speak of "appending" dataset A with dataset B (see Figure 1). For a successful append, the columns of interest should be included in both A and B. The shape of the resulting dataset C will be as follows:

- The number of rows will be the sum of the rows in A and B.
- The number of columns will be the union of the columns in A and B.



Figure 1: Appending two datasets A and B

2.2 Merging

The "merging" command can be used to combine the columns from two datasets to a single dataset (see Figure 2)[5]. As the datasets are combined sidewards, we often speak of a left and a right dataset After merging two datasets A and B, the shape of the resulting dataset C will be as follows:

- The number of rows will be smaller, equal or larger than the number of rows in A and B.
- The number of columns will be equal or larger than the number of columns in A and B.

[4] For example, Stata allows the "keep(varlist)" option in their append command, which allows the user to specify that the variables from the "using" dataset should be kept (StataCorp. (2021a)).

[5] Similar to merging, "joining"allows to combine the columns of two datasets. Depending on the programming language used, there can be slight differences between merge and join commands. In Python, for example, the method join() is used to combine two DataFrames based on their indexes whereas merge() requires specifying columns as a (merge) key (McKinney et al. (2010)). However, for the purpose of this paper, we will use merge as a synonym for join.



Figure 2: Merging two datasets A and B

Merges are usually done on indexes or key variables. When merging over an index, the first observation in A will be matched to the first observation in B, the second in A with the second in B, and so on. When merging on one (or multiple) key variables(s), this variable has to be included in both datasets so that the program can match similar values (e.g., if the key variable is a company ID, then information from A and B is merged per instance, i.e. per company, of this ID). If merging on key variables, the indexes will be ignored. For the remainder of this paper, we assume a merge on key variables.

Depending on the dataset structure, we distinguish between the following **relation types of the key variables**, which describe how the key variables of two datasets are related to each other when they are merged:

- **one-to-one (1:1)**: Each expression of the key variables used to merge occurs once in each of the two datasets.
- **Many-to-one (m:1)**: Each expression of the key variables used for the merge occurs 1-m times in the left dataset and - exactly once in the right dataset.
- **Many-to-many (m:m)**: Each expression of the key variables used for the merge occurs 1-m times in both datasets.
- **One-to-many (1:m)**: Each expression of the key variables used for the merge occurs exactly once in the left dataset and 1-m times in the right dataset.

In addition, we distinguish the following four merge types. They determine which rows will be kept and which ones will be dropped in the merged dataset:

- **Left outer merge**: Returns all rows from the left dataset and only the merged ones from the right dataset.
- **Right outer merge**: Returns all rows from the right dataset and only the merged ones from the left dataset.
- **Inner merge**: Returns only merged rows from both datasets.
- **Full (outer) merge**: Returns all rows from both datasets regardless if they were merged or not.

Figure 3 visualizes the four merge types using dataset A (left) and B (right) as an example:

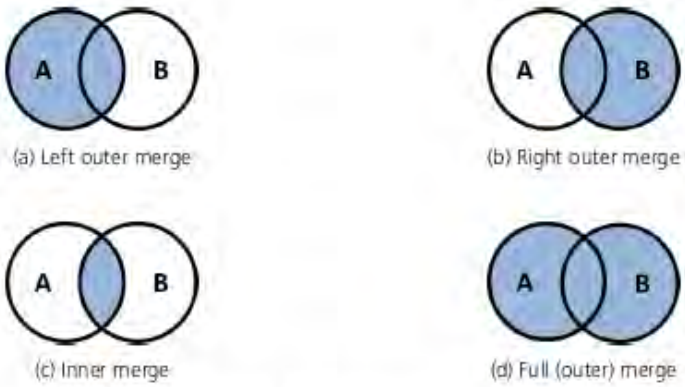


Figure 3: The four merge types

3. What hampers SDC?[6]

As part of their projects, researchers generate results based on confidential Bundesbank microdata. These could be descriptive statistics (e.g. a mean or a frequency table) as well as regressions. When performing SDC on such results, the dataset that was used to compute them (analysis dataset) ideally has the following properties:

1. Usually, results are not calculated directly on the unmodified Bundesbank data (original dataset), but may also be based on previously self-generated variables. Thus, for the SDC of results based on previously self-generated variables, it is imperative that all **rows** that were used for the generation of this variable are also still contained in the analysis dataset.
2. SDC variables do not contain **missing values**.
3. The dataset does not contain **duplicated observations**.

In reality, however, dataset transformations often lead to one or more of these assumptions being violated[7]. In this section, we show why a violation becomes a challenge for SDC and how researchers can address it. In Section 4, we explain whether and when this is the case when appending and merging two datasets.

[6] At this point, we would like to highlight that we try to describe SDC challenges in a short and concise way in this report. We explicitly point out that there will of course always be special cases for which the rules mentioned do not fully apply. We therefore strongly recommend thinking through the particular structure of the data provided by the RDSC and any changes made to it. All resulting SDC requirements should be kept in mind at all times.

[7] Even some of the datasets provided by the RDSC contain SDC variables with missing values or duplicated observations. This is not a sign of data quality issues but can have good reasons (e.g. missing values in the ID of a party that is only involved in certain transactions and therefore missing otherwise).

3.1 Challenge 1: Loss of rows

We describe this challenge and possible solutions in detail in a technical report (Blaschke et al. (2022)). Therefore, we will only give a brief overview in this paper.

Assume we have a dataset that contains company information with two SDC variables "Company_ID" and "Group_ID" (see Table 1a). For our analysis, we are only interested in the companies' groups rather than the individual companies. Therefore, we aggregate the data on the group level (see Table 1b). However, if we want to publish any results computed on "Group_Value", we cannot perform SDC proedures based on Table 1b as we no longer know the number of individual companies that belong to each group (see Principle O.1.3 "Adherence to minimum sample size" in Research Data and Service Centre (2021)) and we further cannot check for dominance (see Principle O.1.4, "Adherence to p% or dominance rule" in Research Data and Service Centre (2021)).

As described in Blaschke et al. (2022), we have to ensure that no SDC variables get lost when using a summary function on a dataset. Summary functions calculate results based on multiple values from multiple rows (e.g. an aggregation as in Table 1). Therefore, we have the following three options to ensure SDC compliance:

1. Perform a SDC before using the summary function and ensure that each row in the aggregated dataset will satisfy the SDC criteria.
2. Use summary functions only when taking into account the full set of SDC variables.
3. Do not drop the duplicated rows after using the summary function. Compute the results by using a dummy variable while performing SDC on the full dataset.

Company_ID	Group_ID	Value
111	AAA	1000
222	AAA	2000
333	AAA	1400
111	BBB	2100
111	CCC	100
111	DDD	100

(a) Company dataset before aggregation

Group_ID	Group_Value
AAA	4400
BBB	2100
CCC	100
DDD	100

(b) Company dataset after aggregation

Table 1: Loss of rows - Challenge

In our example, we choose the third option and define a dummy that is 1 each time "Group_ID" takes on a new value for the first time. Table 2 shows the result.

Company_ID	Group_ID	Value	Group_Value	Dummy
111	**AAA**	1000	**4400**	**1**
222	AAA	2000	4400	0
333	AAA	1400	4400	0
111	**BBB**	2100	**2100**	**1**
111	**CCC**	100	**100**	**1**
111	**DDD**	100	**100**	**1**

Table 2: Loss of rows - Solution

3.2. Challenge 2: Missing values in SDC variables

Assume we are using a dataset on transactions where company A is selling a product to company B and for some transactions, the sale is through an intermediary. In addition, the dataset also includes the ID of company A's group (ownership information). All four parties (companies A, B, the intermediary as well as company A's group) have to be protected in the SDC (i.e. "Company_A_ID", "Group_A_ID", "Inter_ID" and "Company_B_ID" in Table 3 are SDC variables). Tables 3a and 3b show two examples of this dataset with missing values in one of the SDC variables.

Let us assume, we would like to compute the sum of column "Value" in Table 3a and perform an SDC afterwards. The result shows that the aggregate would be based on six different IDs in "Company_A_ID", "Group_A_ID", and "Company_B_ID" and that the dominance criteria would not be violated[8]. However, we only have a single observation with an intermediary and thus, the overall SDC check would result in a disclosure problem[9].

Company_A_ID	Group_A_ID	Inter_ID	Company_B_ID	Value
F05	C		L01	1100
G02	D	Y	V73	20
J08	E		R99	9800
P11	F		C22	3200
Z21	G		K09	4000
L99	H		Q33	6600

(a) Missing ID values are uncorrleated with other SDC variables

Company_A_ID	Group_A_ID	Inter_ID	Company_B_ID	Value
F05	C	X	L01	1100
G02		Y	V73	20
J08	C	Y	R99	9800
P11	D	Z	C22	3200
Z21	E	X	K09	4000
L99		Y	Q33	6600

(b) Missing ID values are correlated with other SDC variables

Table 3: Newly generated missing values - Challenge

[8] Computation of dominance: (9800 + 6600)/(1100 + 20 + 9800 + 3200 + 4000 + 6600) = 66.34% < 85%.

[9] Computation of dominance: 20 / 20 = 100% > 85%.

Does this mean that it is generally very unlikely to get results through the SDC if a dataset contains a rarely filled SDC variable? This question should be answered from the context. To do so, RDSC researchers must proceed as follows:

1. Test whether the missing values of the SDC variable can be filled by interpolation. Some SDC variables are correlated with other SDC variables and can thus be adequately explained by them. In the example in Table 3b, the "Group_A_ID" is missing for some transactions. The reason for this is that not all companies may belong to a group. In these cases, the missing group IDs may be imputed from the related ID variable (here "Company_A_ID"). Please refer to the respective data report to understand whether this procedure can be applied to an SDC variable in a specific dataset.

2. If interpolation is not possible and the SDC variable with the missing values is not sufficiently correlated with any other SDC variable, the missing values can be chronologically numbered. As previously mentioned, not all transactions in Table 3a are processed via an intermediary. As "Inter_ID" cannot be imputed from any other SDC variable, all missing values in "Inter_ID" can be assigned to random ID values that are unique per missing value. To avoid any confusion with the real IDs, the following naming convention must be observed: "sdc_" + consecutive number (e.g. "sdc_15").

Table 4 shows how an SDC following the two approaches would look for the cases presented in Table 3. The two additional variables indicate the sample size ("SDC_N") and the dominance ("SDC_Dom") for "Group_A_ID"[10] and "Inter_ID"[11]:

					for "Group_A_ID"	
Company_A_ID	Group_A_ID	Inter_ID	Company_B_ID	Value	SDC_N	SDC_Dom
F05	C	X	L01	1100		
G02	**G02**	Y	V73	20		
J08	C	Y	R99	9800	6	70.79%
P11	D	Z	C22	3200		
Z21	E	X	K09	4000		
L99	**L99**	Y	Q33	6600		

(a) Interpolated missing IDs

					for "Inter_ID"	
Company_A_ID	Group_A_ID	Inter_ID	Company_B_ID	Value	SDC_N	SDC_Dom
F05	C	**sdc_1**	L01	1100		
G02	D	Y	V73	20		
J08	E	**sdc_3**	R99	9800	6	66.34%
P11	F	**sdc_4**	C22	3200		
Z21	G	**sdc_5**	K09	4000		
L99	H	**sdc_6**	Q33	6600		

(b) Consecutively numbered missing IDs

Table 4: Newly generated missing values - Solutions

[10] Computation of dominance: (1100 + 9800 + 6600) / (1100 + 20 + 9800 + 3200 + 4000+ 6600) = 70.79% <85%.

[11] Computation of dominance: (9800 + 6600) / (1100 + 20 + 9800 + 3200 + 4000 + 6600) = 66.34% < 85%.

3.3 Challenge 3: Duplicates

Table 5a shows a dataset on company groups. Since we are also interested in the associated values for the individual companies, we enrich this dataset with data from Table 5b. As some groups own more than one company, we use a 1:m inner merge[12]. Table 5c shows the merged dataset. Although there are no missing values or deleted rows, there are duplicates in the rows previously contained in Table 5a.

Group_ID	Group_Value
AAA	1000
BBB	2100

(a) Company group dataset

Group_ID	Company_ID	Company_Value
AAA	F22	40
AAA	G54	50
AAA	O99	60
AAA	W71	70
AAA	A01	80
BBB	J77	10
BBB	C30	20

(b) Company dataset

Group_ID	Group_Value	Company_ID	Company_Value
AAA	1000	F22	40
AAA	1000	G54	50
AAA	1000	O99	60
AAA	1000	W71	70
AAA	1000	A01	80
BBB	2100	J77	10
BBB	2100	C30	20

(c) Merged dataset with company and group information

Table 5: Newly generated duplicates - Challenge

If we now calculate the mean of the variable "Group_Value", we cannot simply divide the sum of all values in "Group_Value" by the number of rows in "Group_ID". Instead, we must find a way to account for the duplicates in "Group_ID" and "Group_Value" that were generated when merging. For this, we best follow the approach presented in *Challenge 1: Loss of rows* where we used a dummy variable (see Table 2). Hence, the mean of "Group_Value" would have to be (1000+2100)/2 = 1550. Table 6 shows the resulting dataset including the SDC variables for sample size ("SDC_N") and dominance ("SDC_Dom")[13] for the SDC variable "Group_ID":

[12] Note, that since all Group_IDs are contained in both datasets, the result would have been the same for a left and a right outer merge or a full (outer) merge.

[13] Computation of dominance: (1000 + 2100) / (1000 + 2100) = 100% > 85%.

					for "Group_ID"	
Group_ID	Group_Value	Company_ID	Company_Value	Dummy	SDC_N	SDC_Dom
AAA	**1000**	F22	40	**1**		
AAA	1000	G54	50	0		
AAA	1000	O99	60	0		
AAA	1000	W71	70	0	2	100%
AAA	1000	A01	80	0		
BBB	**2100**	J77	10	**1**		
BBB	2100	C30	20	0		

Table 6: Newly generated missing values - Solution

4 When does appending or merging hamper SDC?

Readers should differentiate between existing challenges in datasets and newly created challenges that appear because of appending or merging datasets. This section explains origins of the latter. For SDC all challenges need to be adressed irrespective of their origin.

4.1 Appending

A dataset C that is generated by appending two datasets A and B will always contain all columns and rows from A and B. Hence, no columns or rows get lost. However, appending can generate **duplicates** and **missing values**. Duplicates occur if at least one row exists in both A and B. Missing values are generated, if there is at least one column that is only included in A or in B:

- Columns only included in A will lead to missing values in all rows in C that came from B.
- Columns only in B will lead to missing values in all rows in C that came from A.

Table 7 shows, how data for a new month (Dataset B) are appended to an existing dataset (dataset A). The length of dataset C is the sum of the rows in datasets A and B while the number of columns in C is the union of A and B. Variable "Group_ID" was only included in dataset B and therefore, all previous months in C are empty in "Group_ID". Finally, note, that no observations from SDC variables ("Company_ID", "Group_ID") were dropped. If "Group_ID" is an SDC variable, we have to address this challenge as described in the previous section.

Date	Company_ID	Company_Value
2021-01	F55	10
2021-01	L91	20
2021-02	F55	20
2021-02	L91	30

(a) Dataset A

Date	Company_ID	Group_ID	Company_Value
2021-03	F55	AAA	10
2021-03	L91	BBB	20

(b) Dataset B

Date	Company_ID	Group_ID	Company_Value
2021-01	F55		10
2021-01	L91		20
2021-02	F55		20
2021-02	L91		30
2021-03	F55	AAA	10
2021-03	L91	BBB	20

(c) Dataset C

Table 7: Example for appending datasets A and B

If we generated missing values in "Company_Value" we must not fill them with a value (e.g. zeros) but preserve their character as missing. Otherwise, we could not distinguish between a reported value (even if zero) and a missing value that must not be included in the SDC computation.

4.2 Merging

Whether SDC challenges occur after merging datasets depends in the

1. Relation of the key variables (e.g. 1:m vs. 1:1)
2. Merge type (e.g. left outer merge vs. inner merge)
3. Content of the datasets (e.g. if both datasets contain the same codes in the merge key variable(s))

As we have these three dimensions determining whether SDC challenges have to be considered. Table 8 gives a first indication about possible challenges. Note that the table only shows merge-induced SDC challenges but does not say anything about pre-merge characteristics of the individual datasets (e.g. dataset A could already have had SDC variables with missing values before the merge). [14]

[14] The technical report available from the RDSC website contains a formal proof of the statements of Table 8 in an appendix.

	Left dataset (dataset A)			Right dataset (dataset B)		
	Deleted rows	Duplicates	Missing values	Deleted rows	Duplicates	Missing values
1:1	None	None	None	Non-merged rows	None	Possible
m:m		Possible			Possible	
1:m					None	
m:1		None			Possible	

(a) Left outer merge

	Left dataset (dataset A)			Right dataset (dataset B)		
	Deleted rows	Duplicates	Missing values	Deleted rows	Duplicates	Missing values
1:1	Non-merged rows	None	Possible	None	None	None
m:m		Possible			Possible	
1:m					None	
m:1		None			Possible	

(b) Right outer merge

	Left dataset (dataset A)			Right dataset (dataset B)		
	Deleted rows	Duplicates	Missing values	Deleted rows	Duplicates	Missing values
1:1	Non-merged rows	None	None	Non-merged rows	None	None
m:m		Possible			Possible	
1:m					None	
m:1		None			Possible	

(c) Inner merge

	Left dataset (dataset A)			Right dataset (dataset B)		
	Deleted rows	Duplicates	Missing values	Deleted rows	Duplicates	Missing values
1:1	None	None	Possible	None	None	Possible
m:m		Possible			Possible	
1:m					None	
m:1		None			Possible	

(d) Full (outer) merge

Table 8: Possible SDC challenges by merge type

For example, in case of a left outer 1:m merge, Table 8 (a) states that when merging A and B, we keep all rows that are either only in A or both in A and B, while we drop those only occurring in B. For those rows that only exist in A, missing values are generated in SDC variables only existing in B. As we are using a 1:m merge we know that each expression of the key variables used for the merge occurs exactly once in A and 1-m times in B. Therefore, duplicates in A will be generated for each expression of key variables that exists multiple times in B.

5. A rule of thumb to resolve the SDC challenges

This section concludes the main findings on addressing SDC challenges from appending and merging datasets. Note again, that SDC challenges might already exist in the original data as provided by the RDSC. They have to be addressed irrespective of their origin.

5.1 Appending

Appending two datasets usually does not lead to big challenges when performing SDC. However, you should keep in mind that you might generate missing values in SDC variables. Furthermore, you should make sure not to drop any SDC variables even when they are rarely filled and not the focus of your research.

5.2 Merging

It is important that researchers understand that the three SDC challenges are not equally serious. While an analysis dataset with missing values or duplicates can usually be accounted for during SDC, this is no longer possible when rows are missing which are needed for SDC. In the latter case, the previous program code would have to be reworked. Therefore, it can be said that **missing rows are by far the biggest SDC challenge**.

As a general rule, we learn from Table 8 that:

- Deleted rows and missing values depend on the merge type.
- Duplicates depend on the relation of the key variables.

Thus, we conclude the following rules of thumb:

- **Understand your data structure**. Key variables identify rows in datasets. The origin of many SDC challenges lies in the relation between these key variables in the original datasets. Once you are aware if you have e.g. a 1:m or a 1:1 relationship in your set of merge key variables, you can use Table 8 to understand which SDC challenges you are likely to encounter.
- **Understand the impact of your selection of merge key variables**. Merging datasets will never create new missing values in SDC variables that are in the set of merge key variables. However, sometimes (e.g. when the two datasets have different SDC variables) we cannot merge over all SDC variables in a dataset. This often happens when one of the datasets contains multiple parties (e.g. buyer and seller, or borrower and lender) or reflects ownership relations (e.g. parent and subsidiary). In these cases, we will most likely create new missing values in an SDC variable.
- **If you cannot reverse it try to avoid it**. As mentioned above deleted rows may not be reversed. When researchers decide to first compute a summary function based on multiple rows and then drop some of those rows, they should carefully plan in advance how to address the challenge of deleted rows. We strongly recommend to first merge the original datasets and then follow the approach from Blaschke et al. (2022).
- **Use full (outer) merges**. Except for full (outer) merges all merge type involve deleting rows from one or both datasets if key variables do not overlap perfectly.

- **Use 1:1 merges**. For all merge types, 1:1 merges perform best. However, we are well aware that 1:1 merges are rarely possible in reality. On the other end, m:m merges always perform worst and even in the Stata reference manual for the **merge** command, it says "*m:m specifies a many-to-many merge and is a bad idea. (...) If you think that you need an m:m merge, then you probably need to work with your data so that you can use a 1:m or m:1 merge.*" (StataCorp. (2021b)). By definition, 1:m (m:1) merges can duplicate a single row in the left (right) dataset up to m times.
- **Duplicated rows make untidy data**. Duplicated rows and tidy data make for a formidable trade-off (Wickham (2014)). A dataset is considered tidy if the same set of key variables identifies all variables in the dataset. By its very definition tidy datasets do not have duplicated rows. However, merging tidy datasets can lead to untidy datasets. For these cases, we recommend sticking with untidy dataset rather then trying to make them tidy again. Tidying a dataset almost always requires deleting rows, which may severely impair SDC.
- **Consider using dummy variables to navigate between different key variables.** Navigating untidy data for SDC requires understanding which key variables identify an observation (e.g. as done in Table 2). For these cases we recommend using a dummy variable to switch different key variables as described in Blaschke et al. (2022).

Glossary

- **Analysis dataset**. The "analysis dataset" is the dataset that is used to compute results. Normally, the analysis dataset differs from the original dataset (see below) due to data transformation steps (e.g. dropping of variables and rows, merging with other datasets, or generation of additional variables).
- **Key variables**. One or more variables (combined) enable the distinct identification of each cell of the other columns. As these variables are the key to identifying the other cells, we refer to them as "key variables". We refer to the key variables in their entirety as the "set of key variables". Database terminology may also call the set the "compound key" or "unit of observation". Merge key variables are key variables that are used to identify rows in a merge.
- **Original dataset**. The "original dataset" is the unmodified dataset as provided by the RDSC.
- **SDC variables**. Variables containing the IDs of sensitive entities are referred to as "SDC variables". As the classification of SDC variables is purely context-related and depends on whether the entities contained therein need to be protected or not, the RDSC specifies SDC variables for each original dataset in its data reports. These reports identify SDC variables as such for each individual dataset, which is why a variable may be an SDC variable for one original dataset but not for another.
- **Summary functions**. Some functions may calculate results based on multiple values from multiple rows, which leads to an aggregation of information. Examples include sums or means across several rows. We refer to these functions as "summary functions".

References

Blaschke, J., Gomolka, M., and Hirsch, C. (2022). *Statistical Disclosure Control (SDC) for results derived from aggregated confidential microdata, Technical Report 2022-01 – Version 1.0*. Deutsche Bundesbank, Research Data; Service Centre. Retrieved from https: //www.bundesbank.de/resource/blob/745168/3501912ca41dc894f605a9d24413dee1/mL/ 2022-01-sdc-data.pdf

McKinney, Wothers. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *445*, 51–56. Austin, TX.

Research Data and Service Centre. (2021). *Rules for visiting researchers at the RDSC, Technical Report 2021-02 - Version 1-0*. Deutsche Bundesbank, Research Data; Service Centre. Retrieved from https://www.bundesbank.de/resource/blob/826176/ffc6337a19ea27359b06f2a8abe0ca7d/mL/2021-02-gastforschung-data.pdf

StataCorp. (2021a). *Stata 17 Base Reference Manual- append*. College Station, TX: Stata Press. Retrieved from https://www.stata.com/manuals/dappend.pdf

StataCorp. (2021b). *Stata 17 Base Reference Manual - merge*. College Station, TX: Stata Press. Retrieved from https://www.stata.com/manuals/dmerge.pdf

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10), 1–23. https://doi.org/10. 18637/jss.v059.i10

DEUTSCHE BUNDESBANK EUROSYSTEM

Statistical Disclosure Control (SDC) for results derived from combined confidential microdata

Christian Hirsch and Jannick Blaschke, Deutsche Bundesbank, Research Data and Service Centre

Session 3b: Data governance and quality

BoC-IFC Satellite Seminar on “Granular data: new horizons and challenges for central banks”

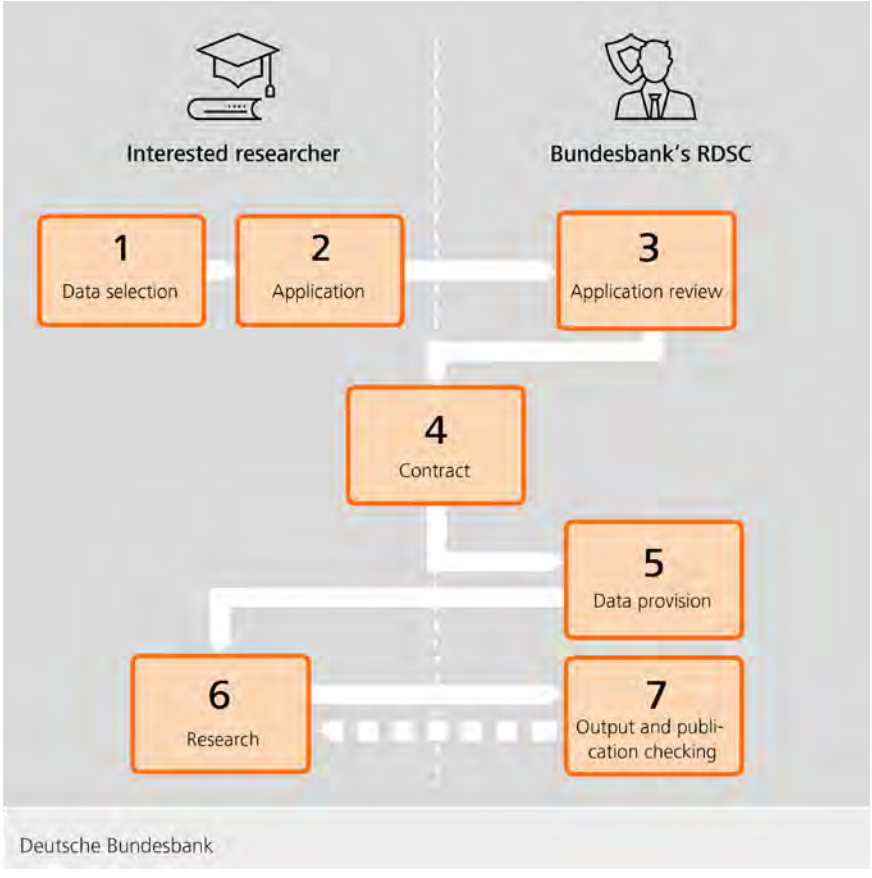
on 15 July 2023 in Ottawa, Canada

The views expressed here do not necessarily reflect the opinion of the Deutsche Bundesbank, the INEXDA network or the Eurosystem.

Research Data and Service Centre

1. Introducing the Research Data and Service Centre (RDSC)

The RDSC in a nutshell



The RDSC offers access for **non-commercial research** to (sensitive) **micro data** of the Bundesbank.

Currently, the RDSC counts around 300 active research projects

For more information please see the information on our **website**: https://www.bundesbank.de/en/bundesbank/research/rdsc/your-research-project-at-the-rdsc

2. Statistical Disclosure Control (SDC)

SDC principles and rules

All results that leave the RDSC must go through a **output control** that is conducted by the researcher and then **checked by two RDSC staff members**.

The two main rules are:

- **Minimum sample size**: Results intended for release must be based on at least **five** different observation units.
- **Dominance rule**: The combined shares of the two largest observation units should **not exceed 85%** of the total value under analysis.



All rules are outlined here:
Rules for visiting researchers: Principles and rules

2. Statistical Disclosure Control (SDC)

Example

	Company_ID	Company_Value
1	F22	40
2	G54	50
3	O99	60
4	W71	70
5	A01	80
6	J77	10
7	C30	20

The average of *Company_Value* is:

$$\frac{40 + 50 + 60 + 70 + 80 + 10 + 20}{7} = 47.1$$

Does this result comply with SDC?

Yes

- ✓ Based on 7 different companies
- ✓ Dominance criterion satisfied as $\frac{70+80}{40+50+60+70+80+10+20} \leq 0.85$

3. What hampers SDC?

Challenge 1: Loss of rows

Fixing a loss of rows might require **reworking previous code**

Company_ID	Group_ID	Value
111	AAA	1000
222	AAA	2000
333	AAA	1400
111	BBB	2100
111	CCC	100
111	DDD	100

(a) Company dataset before aggregation

Group_ID	Group_Value
AAA	4400
BBB	2100
CCC	100
DDD	100

(b) Company dataset after aggregation

- How many company IDs are behind each row?
- Is there a problem with dominance in *Company_ID*?

= SDC variable

3. What hampers SDC?

Challenge 2: Missing values in SDC variables

***Example**: Transaction between two companies with one intermediary party*

Company_A_ID	Group_A_ID	Inter_ID	Company_B_ID	Value
F05	C		L01	1100
G02	D	Y	V73	20
J08	E		R99	9800
P11	F		C22	3200
Z21	G		K09	4000
L99	H		Q33	6600

(a) Missing ID values are uncorrleated with other SDC variables

Company_A_ID	Group_A_ID	Inter_ID	Company_B_ID	Value
F05	C	X	L01	1100
G02		Y	V73	20
J08	C	Y	R99	9800
P11	D	Z	C22	3200
Z21	E	X	K09	4000
L99		Y	Q33	6600

(b) Missing ID values are correlated with other SDC variables

- How do we check the number of entities behind *Inter_ID* if some transactions simply do not have any intermediary party involved?
- How do we compute the dominance for *Inter_ID*?

= SDC variable

3. What hampers SDC?

Challenge 3: Duplicates

Group_ID	Group_Value
AAA	1000
BBB	2100

(a) Company group dataset

Group_ID	Company_ID	Company_Value
AAA	F22	40
AAA	G54	50
AAA	O99	60
AAA	W71	70
AAA	A01	80
BBB	J77	10
BBB	C30	20

(b) Company dataset

Group_ID	Group_Value	Company_ID	Company_Value
AAA	1000	F22	40
AAA	1000	G54	50
AAA	1000	O99	60
AAA	1000	W71	70
AAA	1000	A01	80
BBB	2100	J77	10
BBB	2100	C30	20

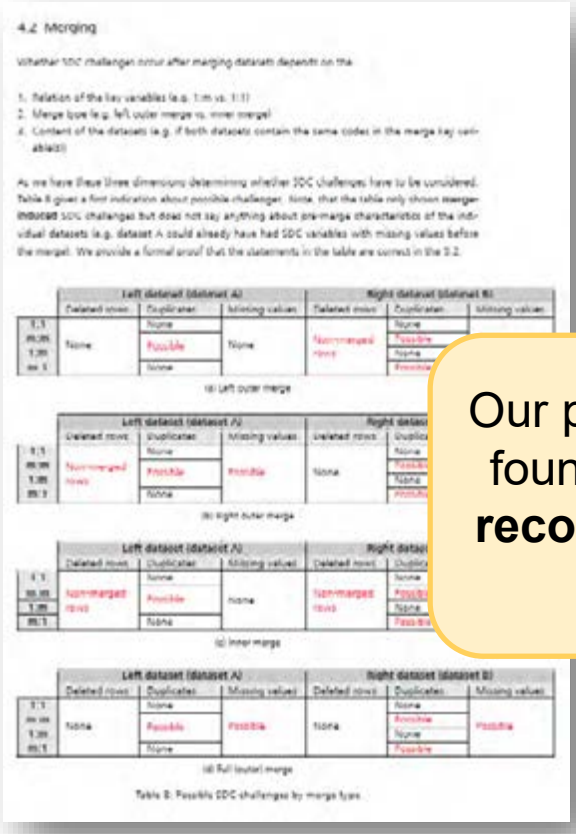
(c) Merged dataset with company and group information

When computing e.g. a mean on Group_Value, we should only consider rows 1 and 6. The SDC must do the same.

= SDC variable

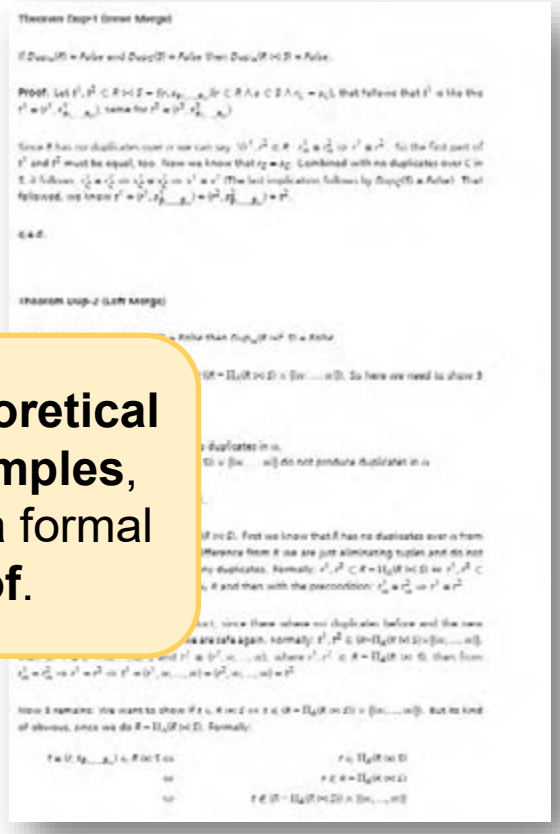
4. Contribution of our paper

The challenges are deterministic but complicated…





Our paper provides a **theoretical** foundation, detailed **examples**, **recommendations** and a formal **mathematical proof**.



4. Contribution of our paper
Merging

Loss of rows



Missing values in SDC variables



Duplicates



1. Understand your **data structure**
2. Understand the impact of your selection of merge **key variables**
3. If you cannot reverse it, try to **avoid** it
4. Use **(full) outer** merges
5. Use **1:1** merges
6. Understand that duplicated rows make **untidy data**
7. Consider using **dummy variables** to navigate between different key variables

4. Contribution of our paper

Merging - Example for solving a loss of rows

Company_ID	Group_ID	Value
111	AAA	1,000
222	AAA	2,000
333	AAA	1,400
111	BBB	2,100
111	CCC	100
555	DDD	200

m:1 inner merge via *Group_ID*

Group_ID	Group_Value
AAA	9,000
BBB	10,000
CCC	15,000

Company_ID	Group_ID	Value	Group_Value
111	**AAA**	1,000	**9,000**
111	**BBB**	2,100	**10,000**
111	**CCC**	100	**15,000**

Challenge 1: Loss of rows

- How many company *Company_IDs* are behind each row?
- Is there a problem with dominance in *Company_ID*?

= SDC variable

4. Contribution of our paper

Merging - Example for solving a loss of rows

Company_ID	Group_ID	Value
111	AAA	1,000
222	AAA	2,000
333	AAA	1,400
111	BBB	2,100
111	CCC	100
555	DDD	200

m:1 inner merge via *Group_ID*

Group_ID	Group_Value
AAA	9,000
BBB	10,000
CCC	15,000

Company_ID	Group_ID	Value	Group_Value	Dummy
111	**AAA**	1,000	**9,000**	**1**
222	AAA	2,000	9,000	0
333	AAA	1,400	9,000	0
111	**BBB**	2,100	**10,000**	**1**
111	**CCC**	100	**15,000**	**1**

SDC for *Group_ID*:

X *based on only 3 different groups*

✓ Dominance criterion satisfied as $\frac{10000+15000}{3*9000+10000+15000} \leq 0.85$

= SDC variable

Conclusion and contact



The full report is available on our website:
https://preview.buba.de/blueprint/servlet/resource/blob/912040/ff1b34cc84c41541f65fc98aae431e7f/mL/2023-02-sdc-data.pdf

Thank you.

Research Data and Service Centre

Christian Hirsch
(christian.hirsch@bundesbank.de)
Jannick Blaschke
(jannick.blaschke@bundesbank.de)

Website: www.bundesbank.de/rdsc

Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Missing values imputation for Central Balance Sheets Data Office (CSBO) accounting data[1]

Iker González Crespo and Pablo Jiménez Segovia,
Bank of Spain

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Missing values imputation for Central Balance Sheet Data Office

Iker González Crespo

Pablo Jiménez Segovia

Abstract

The scope of this paper is to impute missing accounting data in the Central Balance Sheet Data Office (CBSO) questionnaires.

Financial and non-financial companies are obliged to deposit their annual accounts and other administrative records, which are used for statistical purposes. The Bank of Spain receives this information from the non-financial companies through the Commercial Registry. Once this information is gathered an automatic filtering is applied in which missing values are identified so they cannot be used for economic analysis. Therefore, it is proposed to impute this information not to lose these questionnaires, and therefore gain a larger sample (and representativeness of Spanish companies). In order to achieve this goal a series of Machine Learning (ML) algorithms based on random forest have been used to predict these accounting items. Afterwards, an accounting adjustment has been made so that all these imputations balance accounting. Finally, AI explainability techniques have been used to understand the variables that have most influenced when making the imputations.

Keywords: Central Balance Sheet Data Office, CBSO, SME, ML, AI, imputation, miceranger, random forest, R, accounting adjustment, explainability

JEL classification: C53 (Forecasting and Prediction Methods - Simulation Methods), M41 (Accounting).

1. Introduction

The Central Balance Sheet Data Office (CBSO) of the Bank of Spain annually receives and cleans data, primarily accounting data, from nearly one million non-financial companies.

Since the questionnaires through which information is received from the Mercantile Registry sometimes have details of missing information in fields necessary for proper cleansing, and given that this is one of the reasons why approximately 20% of the questionnaires are discarded to date, it was proposed to investigate whether modern statistical techniques could impute those "missing" values. That is the goal of this project.

Different imputation algorithms have been studied, and as mentioned in Section 3.2, the algorithm called "miceranger" was chosen, based on chained random forests.

Once the values have been imputed, in a second iteration, they must balance in an accounting way so that the completed questionnaires pass the subsequent validations carried out in the CBSO. To achieve this, it is necessary to force this reconciliation by applying business rules (the accounting relationships

that exist between the reported items) since no current algorithm, as far as the authors of this article are aware, imputes and reconciles simultaneously.

To carry out this work, a Machine Learning model has been fitted for each information node in which the companies subject to imputation are classified. The nodes are formed as the intersection of each sector of activity (14 groups of activities in the monograph) with each size (according to EU Recommendation). The sizes of the analysed companies are: large, medium, small and micro-enterprise.

In 2019, the Central Balance Sheet Data Office began a proof of concept in collaboration with the Instituto de Ingeniería del Conocimiento (IIC), where a similar imputation was tested, focusing only on four totals (short-term debts, long-term debts, debtors, and creditors). One of the intentions of the current work has been to broaden that approach for four reasons:

- Increase the number of items that can be imputed.
- Attempt to incorporate the accounting logic of relationships between items, beyond their specific totals, to help the algorithm learn. For instance, the "Cash" item could influence another item in a different total, such as "Long-term financial investments."
- This more ambitious path can be extended to include items from the "Income and Expenses" section.
- The accounting explainability of the algorithm and the patterns that can be detected may open doors to deeper insights into accounting relationships.

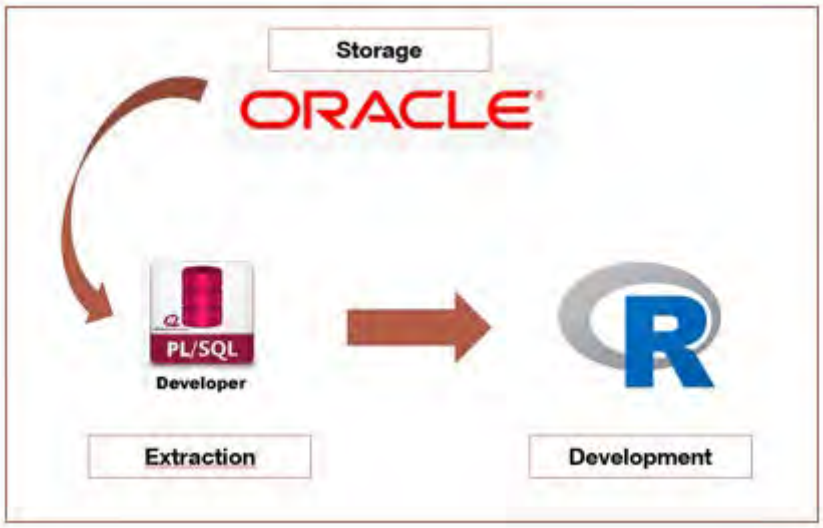
To sum up, this work is the outcome of a combination of data science (extraction, processing, data wrangling, and deployment of Machine Learning models) and business knowledge (understanding of accounting items, questionnaire frequencies, ranges of possible values, and accounting adjustments). After imputing the missing values, an analysis is conducted to assess the correctness of these imputations at an aggregate level, meaning by node, by sector, and by size.

2. The data

A database called GTB (which is the database of questionnaires received in an environment accessible to data scientists) is available for the years 2018 and 2020, hosted in Oracle. This database contains accounting data and assessments of questionnaire quality (reconciliations, checks, reliability of monetary units, cleansing dates, and other metrics analysing the status of questionnaires at each point in time).

This project is conducted for a specific type of questionnaire: those originating from the Mercantile Registry (CBB) and of the "reduced" type (i.e., those used by companies with fewer than 50 employees). In CBB, there is only one questionnaire for each company in each fiscal year.

To connect with the database, the PL/SQL tool is used, and the desired information is extracted, stored in a directory, and accessed using the R programming language. In our case, we select questionnaires that are perfectly filled out (passing all the quality controls used by the Balance Sheet Central, marking the company as consistent and suitable for use in studies). A procedure for removing outliers (an internal process developed in the CBSO) is applied to these questionnaires to discard anomalous cases. Although they comply with all business accounting rules, they have values that exceed the normal value of their node and are eliminated to avoid biases in the studies. These data serve as both the training set (train) and as a reference for algorithm verification (test).

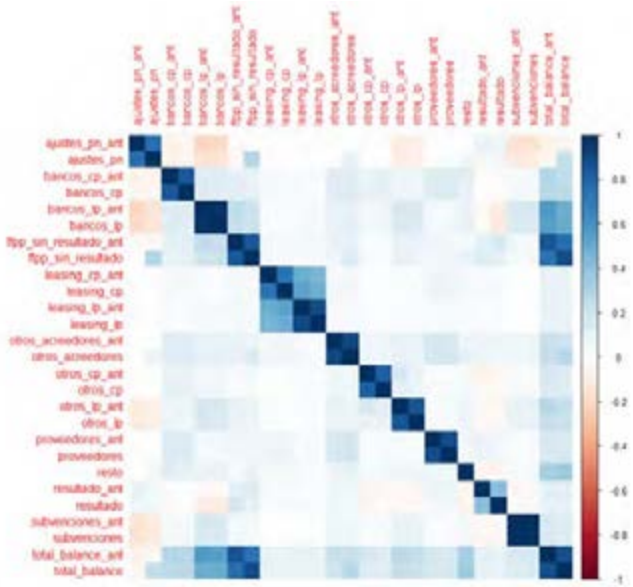


The data to be imputed corresponds exclusively to the balance sheet items. An example is provided to illustrate the type of information we have in the questionnaires for the asset categories of Non-Current Assets and Current Assets:

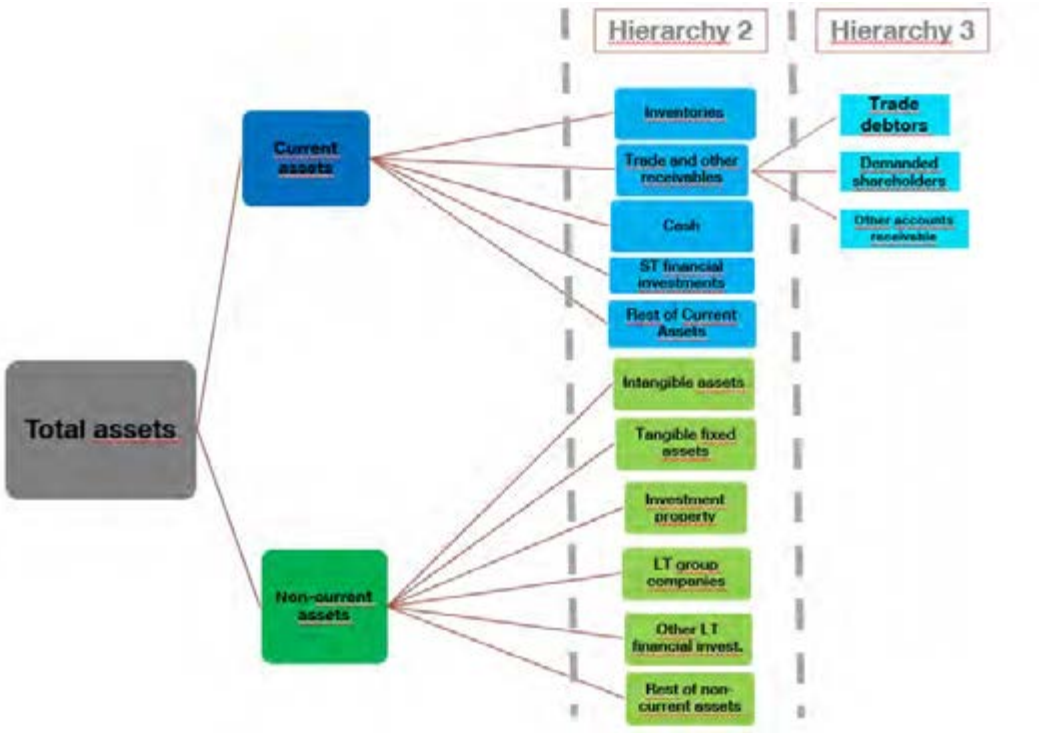
ACTIVO		NOTAS DE LA MEMORIA	EJERCICIO 2020 (2)	EJERCICIO 2019 (3)
A) ACTIVO NO CORRIENTE	11000		424,71460	424,38489
I. Inmovilizado intangible	11100			0,02288
II. Inmovilizado material	11200		358,34557	386,62683
III. Inversiones inmobiliarias	11300			
IV. Inversiones en empresas del grupo y asociadas a largo plazo	11400			
V. Inversiones financieras a largo plazo	11500			
VI. Activos por impuesto diferido	11600		68,36903	57,73518
VII. Deudores comerciales no corrientes	11700			
B) ACTIVO CORRIENTE	12000		1.173,05580	1.068,56633
I. Activos no corrientes mantenidos para la venta	12100			
II. Existencias	12200		527,76339	499,86441
III. Deudores comerciales y otras cuentas a cobrar	12300		501,45542	420,93524
1. Clientes por ventas y prestaciones de servicios	12380		490,71850	396,46022
a) Clientes por ventas y prestaciones de servicios a largo plazo	12381			
b) Clientes por ventas y prestaciones de servicios a corto plazo	12382		490,71850	396,46022
2. Accionistas (socios) por desembolsos exigidos	12370			
3. Otros deudores	12380		10,73692	24,47502
IV. Inversiones en empresas del grupo y asociadas a corto plazo	12400			
V. Inversiones financieras a corto plazo	12500			19,47000
VI. Periodificaciones a corto plazo	12600			
VII. Efectivo y otros activos líquidos equivalentes	12700		143,83699	128,29668
TOTAL ACTIVO (A + B)	10000		1.597,77040	1.492,95122

Note that the value of keys is available for each of the accounting items for the current year (2020 in this example) and the previous year (2019 in this example). Some of the items have been grouped into variables called "Rest" as they are considered too infrequent for algorithms to learn from (having very little variability).

When conducting analyses, data from the previous year's items has been taken into account. This is because having data from the previous year is considered an important variable for imputing values for the current fiscal year, and vice versa. The following image shows the correlation with the previous year's items.



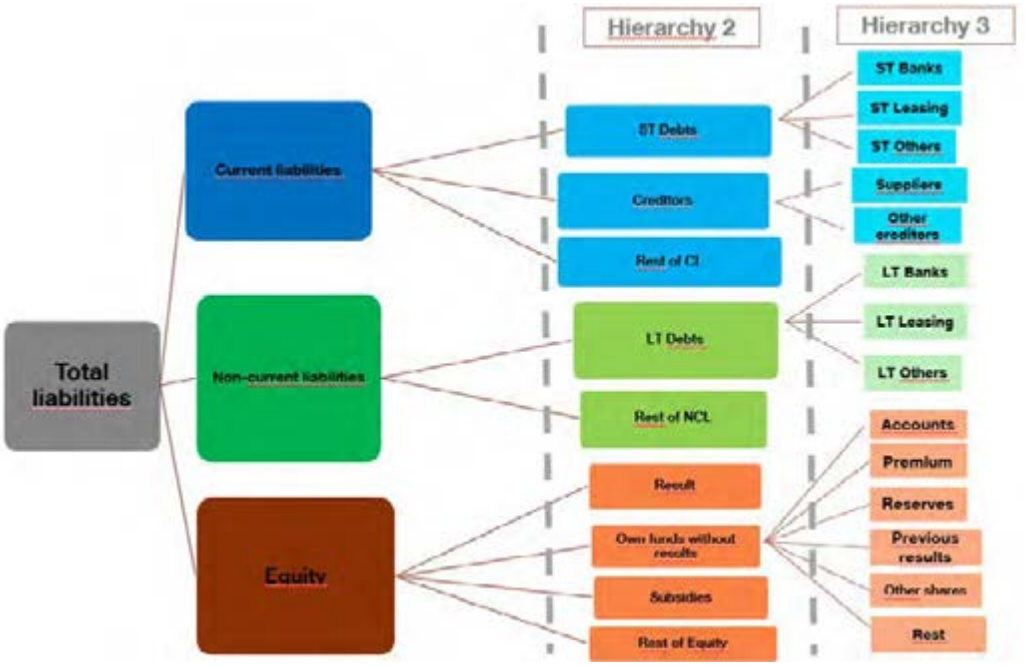
For the current year and the previous year, the following relationships of accounting items for assets exist:



The values of accounting items that are totals (total assets, current assets, and non-current assets) will always be provided due to the technical requirements of the CBSO. The rest of the accounting items are the ones that we will impute and then perform an accounting adjustment to ensure that their sum matches the total of their respective totals.

In the case of assets, we select questionnaires that have all positive accounting items since all their values must be greater than or equal to 0.

On the other hand, liabilities have the following relationships:



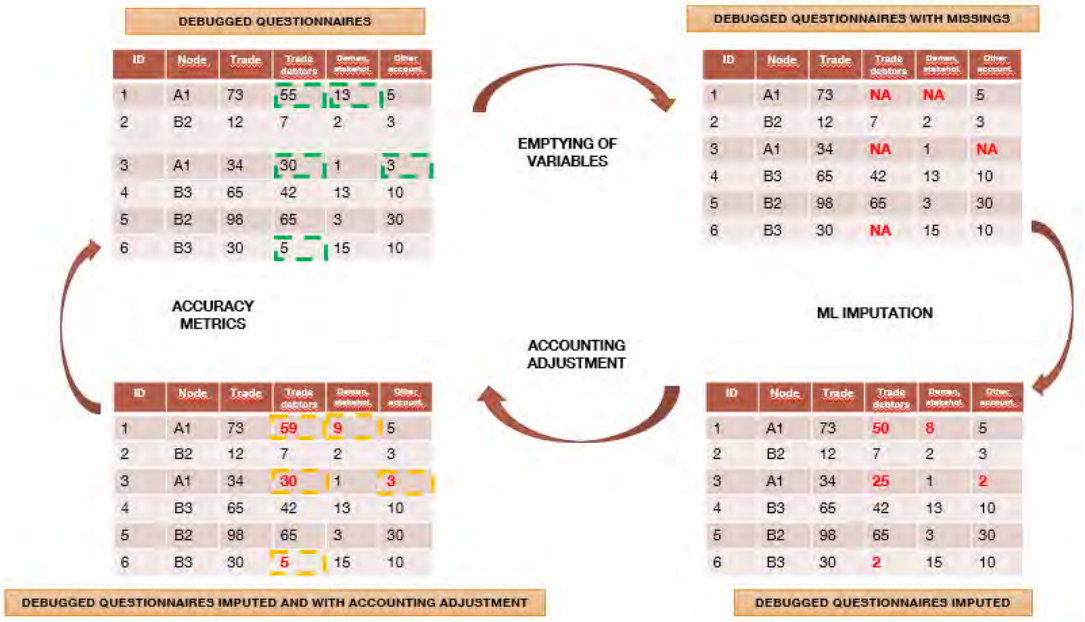
Where the items that are totals, namely, Current Liabilities, Non-Current Liabilities, Equity, will always be reported. The rest of the accounting items are the ones that will be attempted to be imputed. Long-Term Debt, Short-Term Debt, Creditors, and Own funds without results are also total items that must be taken into account when reconciling the corresponding sums.

In the case of liabilities, there are items that must always be positive, while others may be negative at times (such as the result for the year). We only consider questionnaires where those accounting items that must necessarily be positive are positive, with the aim of preventing the algorithm from learning from errors.

Additionally, it was decided not to impute items that are very infrequent, i.e., items that exist in only, for example, 1% of cases. This decision is based on three main reasons:

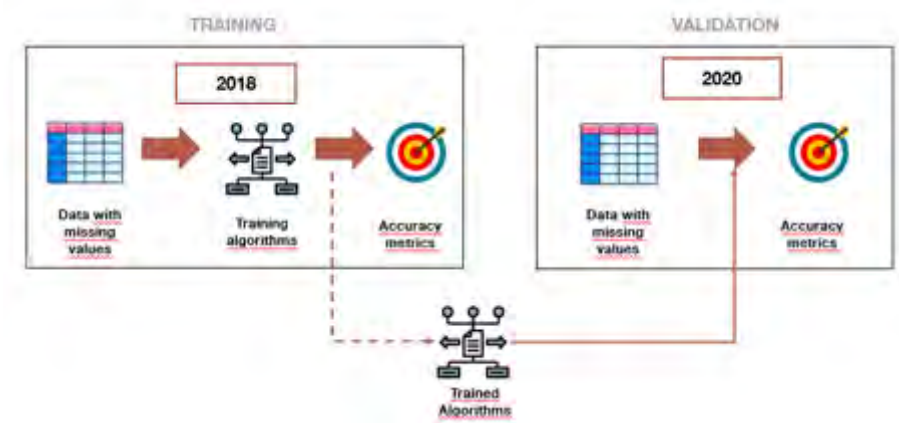
1. Artificial intelligence algorithms do not learn well from variables with little variability. In other words, they struggle to learn patterns effectively. This falls within the realm of "variable selection."
2. Having fewer items reduces the complexity of the development.
3. It seems reasonable not to impute a value for an item that, in practice, is almost always zero.

3. Training and validation method



This imputation scheme is carried out for each node (business sector and company size) as it is believed that the behaviour between each of them may differ. Therefore, an algorithm will be obtained for each node.

Data from the year 2018 is used to train the algorithm and select the one with fewer errors. The data from 2020 serves as validation to check if the final algorithm can generalize well for other years.



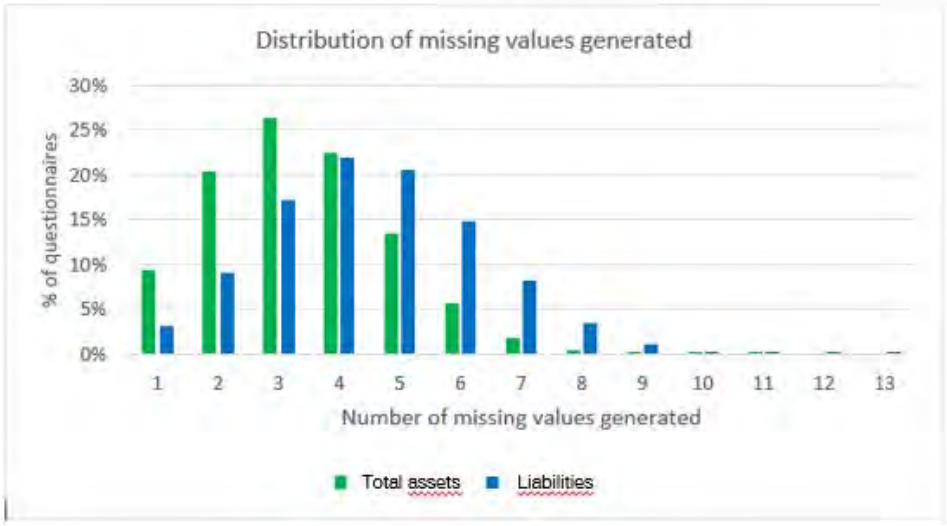
It is important to emphasize that real data from the questionnaires is not modified. In other words, the data reported in the questionnaires is accepted as valid and is not imputed.

Additionally, it is worth mentioning a step performed before imputation: if the "Result" item is reported in the Income Statement and does not appear in the Balance Sheet, that "Result" is not imputed but is transferred directly. This prioritizes the actual data received over imputations.

Going into more detail on each of the steps in the methodology:

3.1 Variable emptying

For both the year 2018 and the year 2020, there are perfectly reconciled and publishable questionnaires for the CBSO. Thirty percent of the items are emptied in a set of over 800,000 questionnaires used in the model. In 2020, the distribution of missing values generated for each questionnaire is as follows:



Where we can see that the distribution of the number of missing values does not concentrate on any specific item, and moreover, the number of blanks generated in the questionnaires is distributed randomly.

This emptying is decided to be done randomly because there is no intention to bias the algorithms, and it is assumed that in production, any combination of missing values may occur. In other words, no hypothesis is made about the patterns of missing values that the algorithm should learn.

3.2 Imputation algorithm

Due to the high computational cost required for this project to read, process, and launch various ML algorithms, a workstation was used to reduce project times.

Different algorithms of various natures were tested to find the one with the best fitting metrics. Among others, the following were tested:

- KNN: a distance-based algorithm. Different metrics and "k" values were tested to obtain the best fit. The distance calculation significantly slowed down the execution of these algorithms. The results were good but below the performance of other algorithms.
- MICE: a classic algorithm for imputation using chained equations. This algorithm proved to be computationally expensive, and the results were also below those of other algorithms.
- Missranger: an iterative algorithm based on random forest. This algorithm achieved very good results but did not explicitly return a model for interpretability. References to this model can be seen here aquí.

- Miceranger: another iterative algorithm based on random forest. The operation of the algorithm is described in the following link vínculo. Two variants of the algorithm were tested (with and without Predictive Mean Matching).

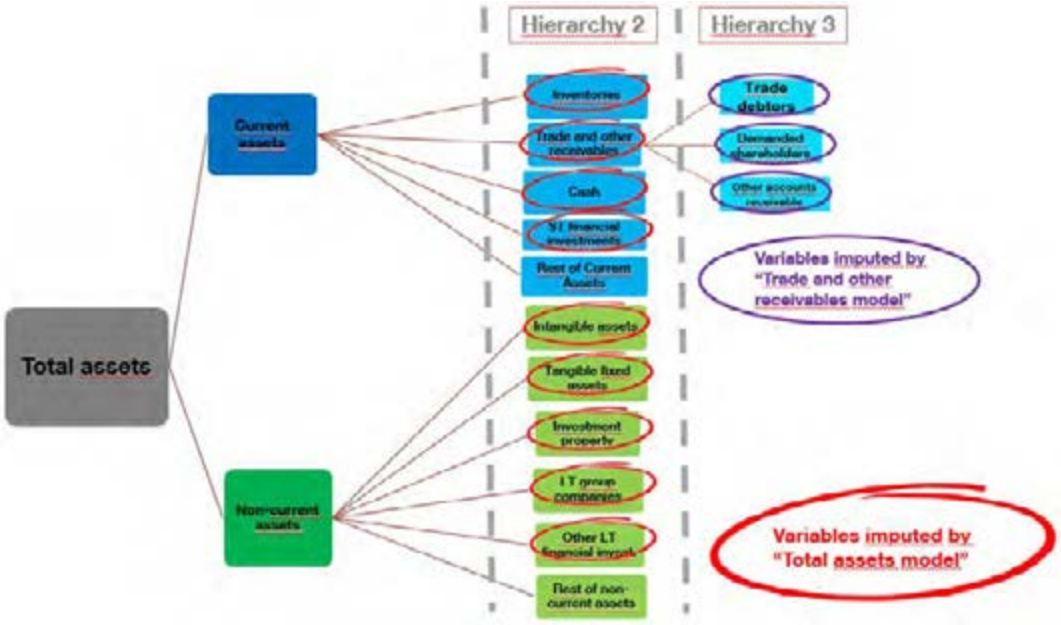
Ultimately, the chosen model was Miceranger without PMM (Predictive Mean Matching) since its fitting metrics were the best among the previously mentioned algorithms. This algorithm works very well because it is parallelizable (faster in execution) and is also an interpretable model (which is necessary for explaining its results). The fitting metrics that have been analysed (some aggregated, some accounting item-specific, and some questionnaire-specific) are as follows:

Type of metric	Metric	Definition
Aggregated	**Total aggregate variation**	sum of all predicted values divided by the sum of all actual values
Aggregated	**Total aggregate variation, broken down by variable**	sum of all predicted values divided by the sum of all actual values, broken down by variable
Accounting item	**Average error**	average of the errors committed in all accounting items
Accounting item	**Average error, broken down by variable**	average of the errors committed in all accounting items, broken down by variable
Accounting item	**Error deviation**	standard deviation of errors made in all accounting items
Accounting item	**Error deviation, broken down by variable**	standard deviation of errors made in all accounting items, broken down by variable
Accounting item	**Maximun error**	maximum error committed in all accounting items
Accounting item	**Correlation**	similarity of the structure of actual items versus imputed items
Accounting item	**% of accounting items with error**	% of accounting items where there has been an imputation error
Questionnaire	**% of questionnaire items with error**	% of questionnaires where there has been an imputation error in at least one accounting item

The training is carried out with questionnaires from 2018. During training, the procedure we use is to train 5 Miceranger models for each node and select the model with the lowest Out Of Bag (OOB) error, a common measure in this type of algorithm based on random forest.

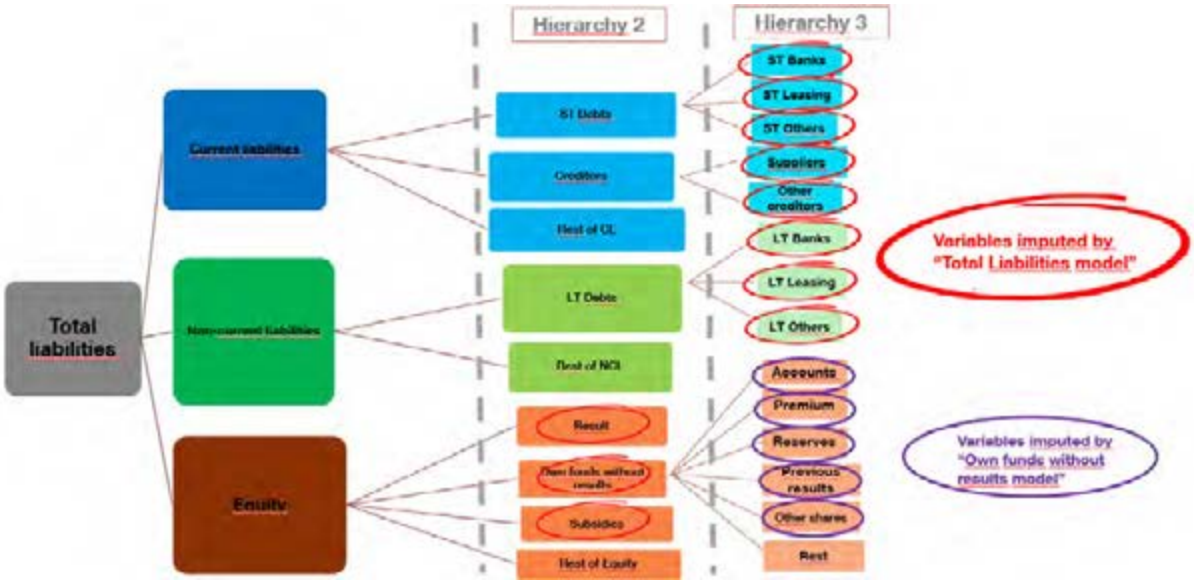
3.3 Training and validation of the algorithm

In the case of questionnaires for asset items, it is considered that there is not enough data for the items indicated in Figure 6 as hierarchy 3 (more detailed items) for the algorithms to learn patterns. Therefore, a model is built for accounting items of hierarchy 2 and another model for variables of hierarchy 3. In summary, for each year (current and previous) in the assets, we have:



In other words, two models are constructed: one for imputing items that fall under Current Assets and Non-Current Assets (hierarchy 2, marked in red), and another for imputing items located under Debtors (hierarchy 3, marked in purple). Accounting items related to "Rest" are not imputed, as they typically appear as zero and are less relevant.

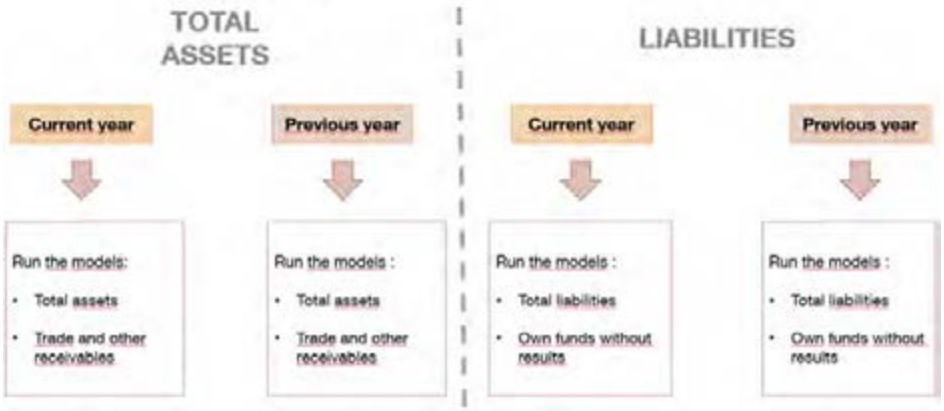
In the case of liabilities, questionnaires have more information for the more detailed items (hierarchy 3), and that's why the modelling approach is different. For each year (current and previous), the following scheme is used for liabilities:



The process began by creating a model to impute the variables marked in red (total liability model) because it was considered sufficient for the algorithm to learn, separating the Result item due to its accounting economic relevance. Later, for production purposes and to recover all the items in the questionnaire, it was decided that it was necessary to impute those belonging to the rest of the Net Equity items marked in purple (Net Equity model without result).

Similar to the approach for assets, accounting items related to "Rest" are not imputed, as they typically appear as zero and are less relevant.

In summary, for each year (current or previous) and each balance sheet category (assets or liabilities), the following models are executed:



3.4 Accounting adjustment

Once the imputations are completed, it is necessary to reconcile them accounting-wise, ensuring that the accounting items match their respective totals. For this, in each questionnaire, an adjustment factor is created for each total, defined as:

$$adjustment\ factor = \frac{\text{total} - \sum non - imputed\ variables}{\sum imputed\ variables}$$

This factor will multiply each of the imputations made by our algorithm. Depending on whether the items and totals can be negative or not, calculations had to be adapted to reconcile. If all items are positive, the corresponding adjustment factor for that total can be directly applied. However, when there are potentially negative items, it was necessary to adapt the programming to subtract those items before applying the adjustment factor.

This part has been particularly complex and has consumed a significant portion of the project's time due to the large number of cases involved. The assurance provided by this strategy is that ultimately, all accounting items are reconciled.

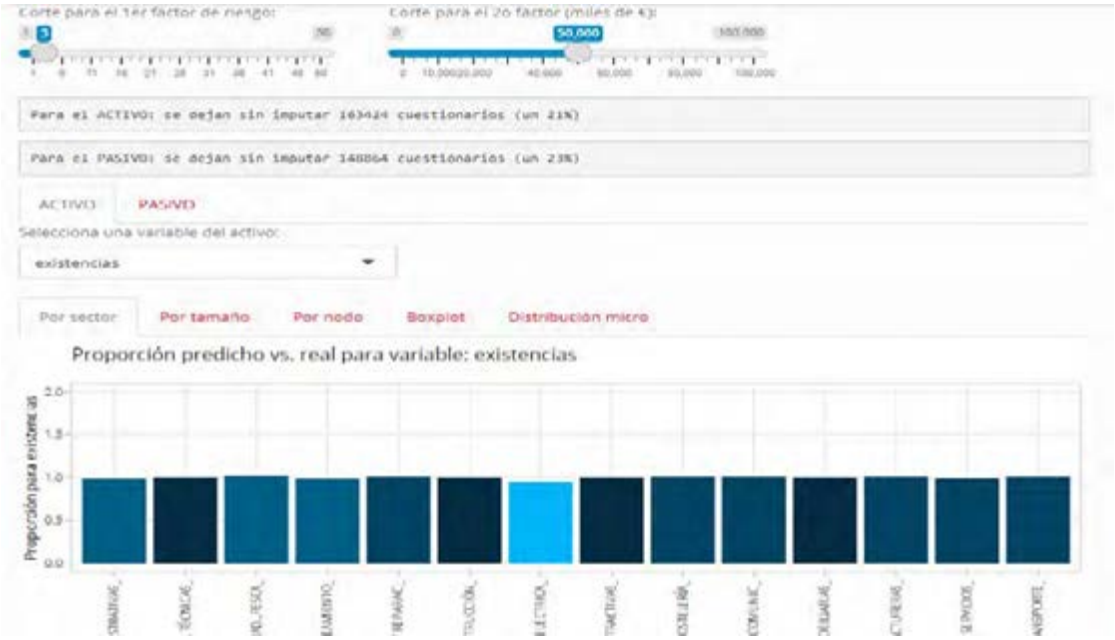
4. Risk factors

Taking into account these accounting adjustments, two risk factors have been identified when the imputation by the algorithms is not correct:

1. Adjustment factor: the quantity by which we need to multiply our imputations to reconcile the items with their totals. Note that if the algorithm imputes quantities that do not follow the actual proportion of accounting items and is multiplied by a high accounting adjustment, then errors will increase considerably.
2. Amount pending imputation: the difference between the total and the actual values (provided by the companies and therefore not imputed) associated with that total.

It is about balancing the risk associated with having a poor imputation with the attempt to impute the maximum number of questionnaires possible. Taking a conservative stance, a threshold of 3 was set for the first factor and 100,000 euros for the second factor. It is required that at least one of the two factors does not exceed the established threshold (in this case, that the first factor is less than 3 or the second factor is less than 100,000 euros) to be able to impute the questionnaire.

Simultaneously, a web application was developed using Shiny to perform quality control of aggregates (sector, size, and node) and make decisions about these thresholds. Additionally, this app serves to verify that the imputations are reliable for the data of the year 2020:



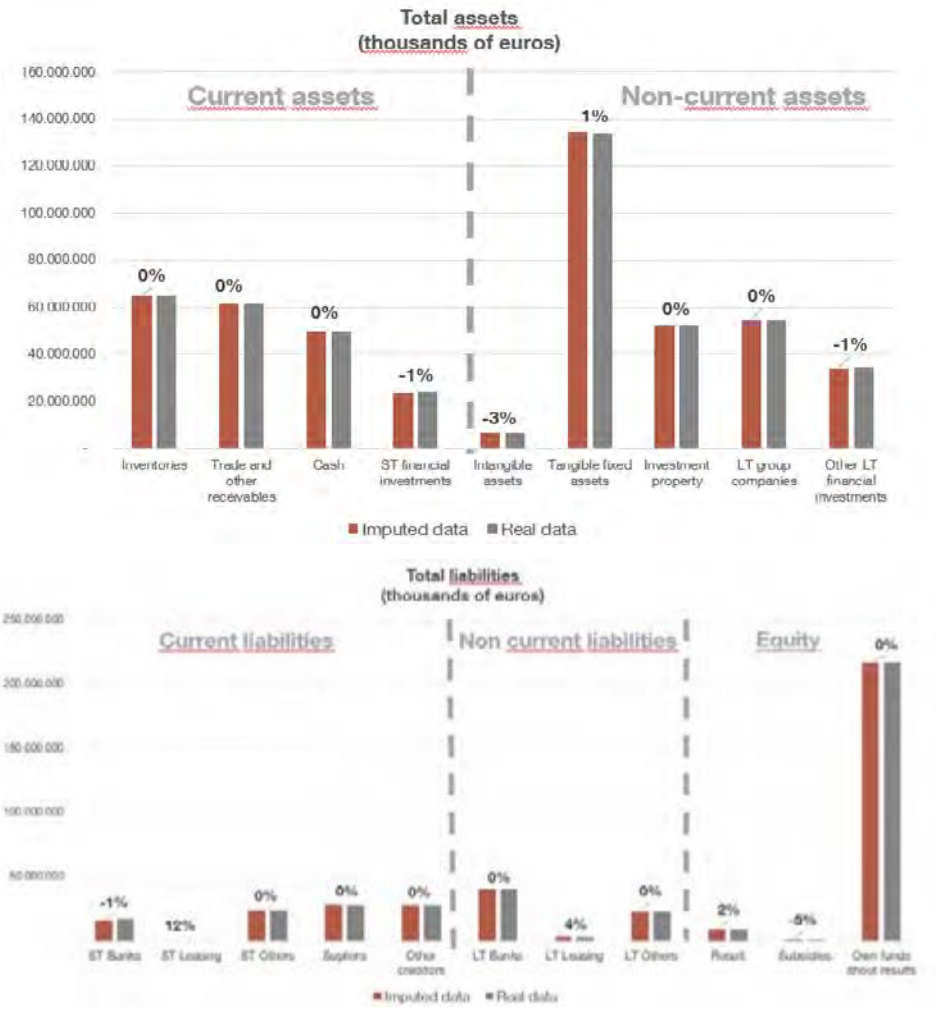
This interactive app will be made available to researchers so that they can consider which nodes have more certainty and which ones have less.

5. Results

As mentioned earlier, the primary goal is to try to analyze if the imputations are reliable at an aggregate level (for the statistics that the CBSO publishes in its reports). The data used for the results are those related to the year 2020, as our models were trained on 2018, and in 2020, we have real data to compare the imputed values with the actual values.

5.1 Aggregated imputations

In aggregate terms, the difference between the actual and imputed accounting items is minimal, as observed in the following graph:



Where the differences between the actual and imputed values are very low percentages.

However, since there are many combinations of accounting items and types of aggregation (size, sector, and node), in the aforementioned Shiny web application, results can be analysed interactively. The metric displayed in the application is a ratio of the actual value to the imputed value. In other words, the closer it is to 1, the closer the imputed and actual values are.

This is illustrated with three examples for the accounting item "Cash" for the three types of aggregation:



5.2 Questionnaire imputations

Despite not being the main goal of the project, the imputations by questionnaire yield acceptable results. To verify these results, an error metric has been created for each item i and questionnaire j as follows:

$$error_{ij} = abs(real_{ij} - imputed_{ij})$$

The absolute value is taken to avoid cancelling positive and negative errors when calculating aggregate metrics for this error.

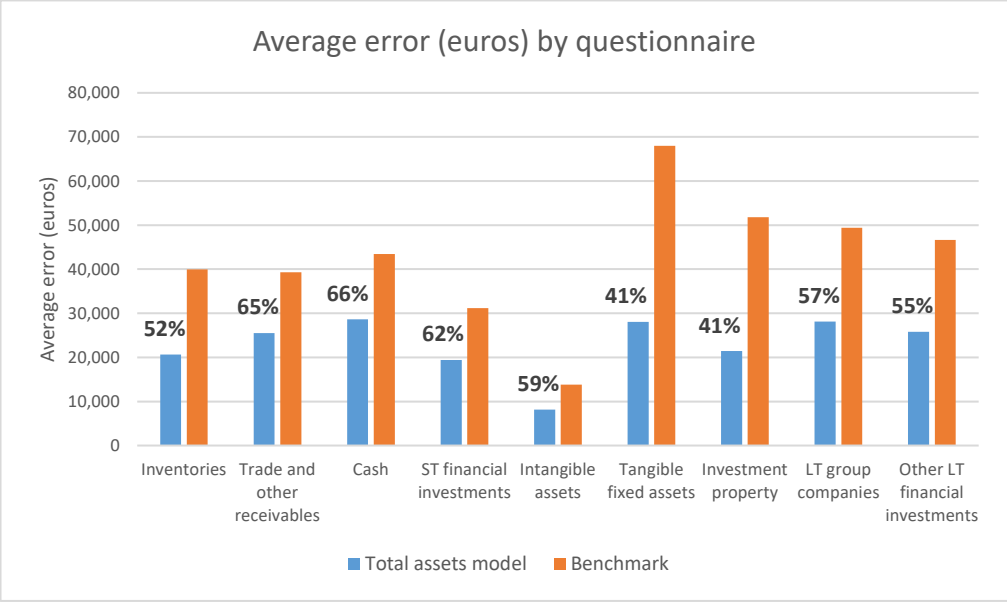
This way, we have the total error of the imputations made for each of the questionnaires in 2020.

5.2.1 Benchmark

When deciding whether an algorithm makes good predictions or not, it is common to use a "naive" model as a baseline or benchmark. In this case, the following model has been constructed as a benchmark:

- Calculate the mean for each accounting item for each of the nodes.
- Impute missing values with the mean of the node and the respective accounting item. This is precisely the system used by the Central de Balances when it wanted to recover specific companies: estimating the missing value based on the normal mean data of its aggregate (crossing business sector and size).
- Adjust those predictions to reconcile accounting-wise.

Comparisons have been made between the different models adjusted with Miceranger and this "benchmark" model. For example, for the Total Assets model, the following comparisons of the previously defined error (calculating the average error per questionnaire) have been obtained for each accounting item:



Where the percentages represent the percentage difference in error of our Total Assets model (obtained through ML) compared to the benchmark model (previously mentioned baseline model). In this graph, it is observed that for each accounting item, we obtain a significantly lower average error for our model compared to the benchmark model, using 610,103 imputed questionnaires for this comparison.

5.2.2 Distribution of errors

The distribution of errors (defined in section 5.2) for each of the models is as follows:

		Distribution of errors		
Model	**Imputed questionnaires**	**Mean**	**Median**	**P99**
Total assets	610.103	53.000 €	1.000 €	835.000 €
Trade and other receivables model	529.322	5.000 €	0€	75.000 €
Total liabilities	512.342	37.500 €	1.000 €	594.000 €
Own funds without results	484.093	33.500 €	0€	458.000 €

Finally, observing the P99 (99th percentile), we see that errors for the vast majority of questionnaires are not high. However, if we were to obtain the maximum errors for each model, they would be high amounts, and this is due to an incorrect imputation where the algorithm does not make a mistake in the amount but rather in the accounting item where it imputes it. This could be mitigated by lowering the threshold in the second risk factor (Section 4).

In other words, the distributions of these errors indicate that, for the vast majority of questionnaires, the models perform a good imputation.

5.3 Algorithm weight in respect to the accounting adjustment

In addition to the results, an analysis has been conducted on how much weight the algorithm has had compared to the adjustment in the final imputation for the year 2020. This metric has only been created for the asset models (total assets and receivables), as all items are positive, making it easier to create this metric and drastically reducing its possible scenarios.

1. For questionnaires that have any imputation, the sum of the breakdowns of the total (for Total Assets, the breakdowns would be current assets and non-current assets) is obtained. This sum is calculated for the algorithm's own predictions (SumPredP) and the predictions after the accounting adjustment (SumPredT).
2. Once these sums are obtained, depending on the scenario faced, the metric is calculated in one way or another:
    a. If SumPredP < SumPredT, then:

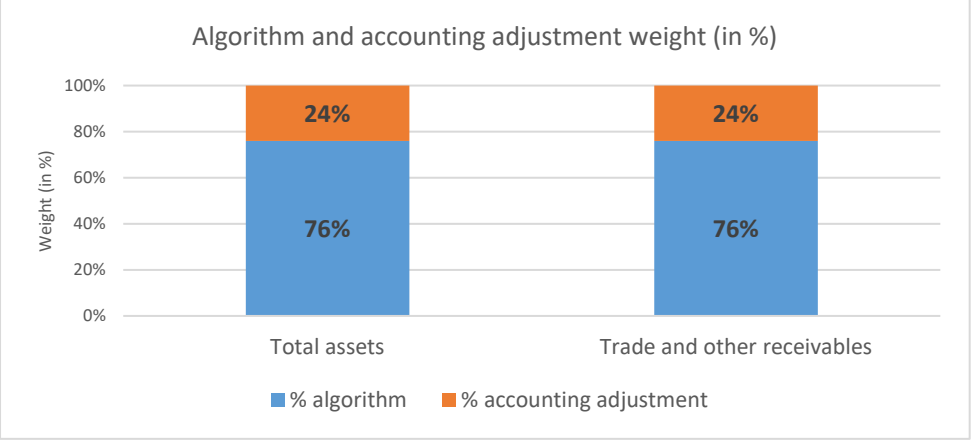
$$algorithm\ weight\ (\%) = \frac{SumPredP}{SumPredT}$$

    b. If SumPredP>=SumPredT, then:

$$algorithm\ weight\ (\%) = \frac{SumPredT}{SumPredP}$$

3. Finally, the average weight of the algorithm is calculated for each of the questionnaires.

Taking into account this described metric, the results for the Total Assets and Receivables models (items always positive) are as follows:



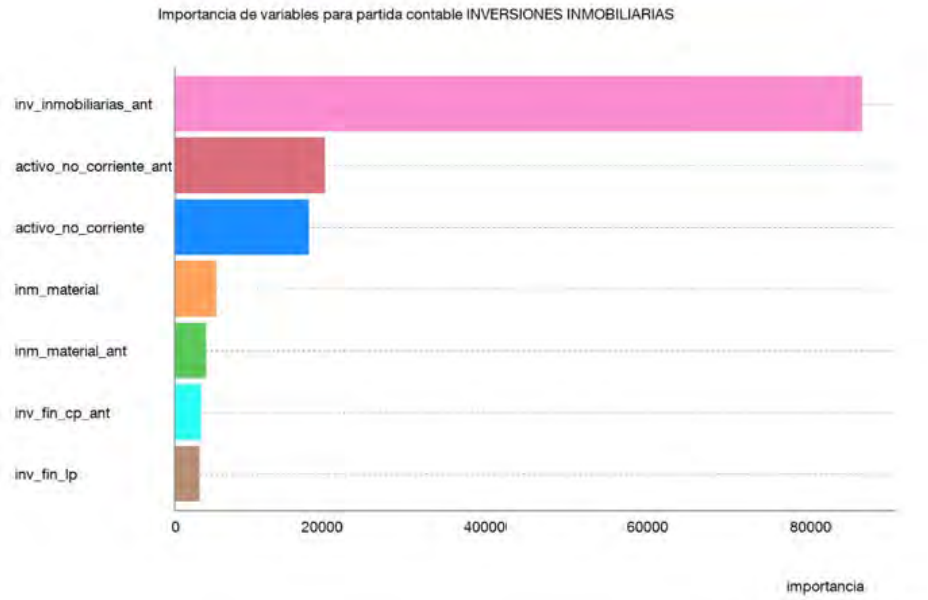
In conclusion, the results of the imputations at the aggregate level are very good. At the questionnaire level, the algorithms impute considerably well for the vast majority of questionnaires, although there are significant errors for a very small percentage of them.

6. Explainability

Once the final imputations have been obtained for each of the questionnaires, it is of interest to know which accounting items have contributed the most (and to what extent) when deciding the imputation. Since our Miceranger algorithm returns a random forest for each of the imputed items, explainability has been studied both globally and locally.

6.1 Global

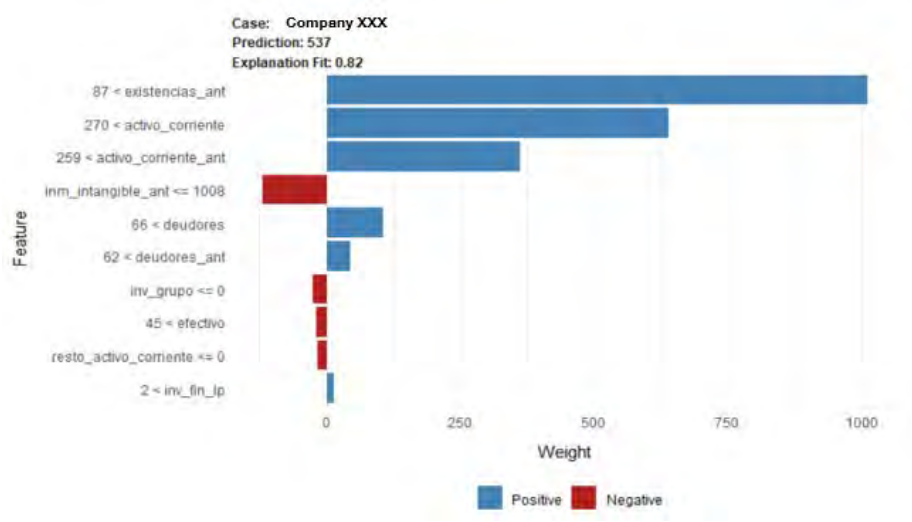
In this case, the accounting items that have contributed the most in a general sense to each of the imputations of the different accounting items have been analysed. For this purpose, variable importance has been used. To illustrate this, an example is provided for the random forest resulting from the accounting item "Inversiones Inmobiliarias" (Investment properties):



Where, in this example, it can be observed that the "Inversiones Inmobiliarias" (Investment properties) item depends on the real estate investments from the previous year and also on items closely related to it.

6.2 Local

In the case of local explainability, LIME (Local Interpretable Model-Agnostic Explanations) has been used to analyse, at the questionnaire level, which accounting items have had the most impact on that imputation and in what direction. As an illustration, the following example is provided:



In this case, it is a small-sized company where it can be observed that the high inventory from the previous year contributes very positively to the imputation of current inventory for the current year. Similarly, a higher volume of current assets (for the current year and the previous year) also contributes to a higher imputation of inventory. Therefore, it seems that the algorithm provides reasonable local explanations about which items influence and in what direction.

7. Production

The last part of this project is the deployment of this entire procedure to automatically carry out this imputation in the internal applications of the Bank of Spain.

7.1 Method

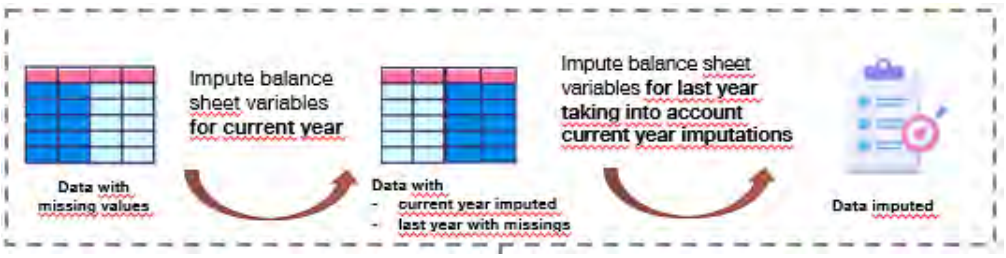
In the database, there is no distinction between a value reported as zero and an empty value. Therefore, the first step is to transform those empties corresponding to square summations into zeros since any imputation would lead to a mismatch.

Subsequently, the saved models are applied to the current year's data. This should be done following the order in which the flow has been designed for both assets and liabilities (the process where imputations are chained):

1. First, impute current year's data:
    - a. Impute and reconcile hierarchy 2 items (total assets model for assets/total liabilities model for liabilities).
    - b. Considering the imputations made in a., impute and reconcile hierarchy 3 items (debtors model for assets/ equity without result model for liabilities).
2. With the imputed information from the current year, impute the previous year's data:
    1. a. Impute and reconcile hierarchy 2 items (total assets model for assets/total liabilities model for liabilities).

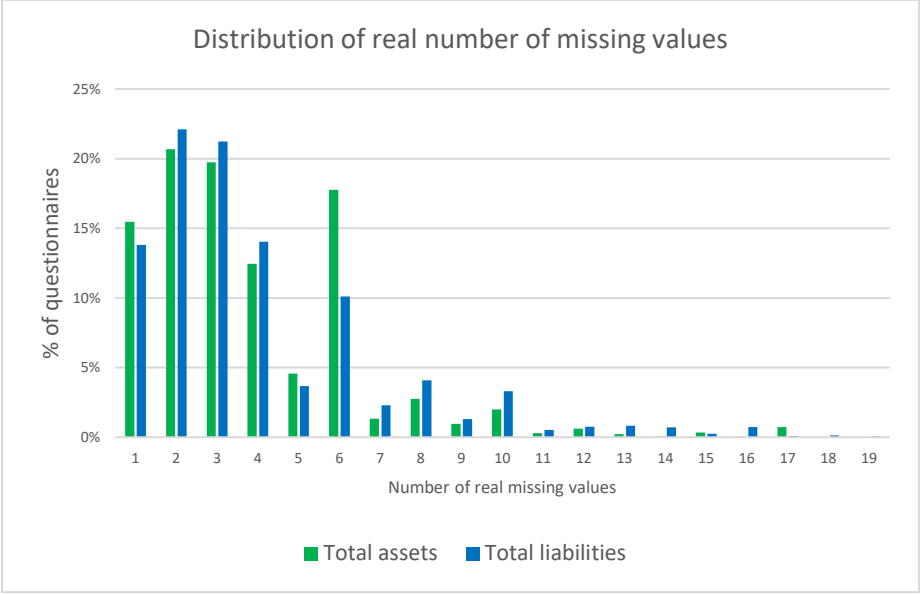
2. b. Considering the imputations made in a., impute and reconcile hierarchy 3 items (debtors model for assets/ equity without result model for liabilities).

Finally, once the imputations for the entire questionnaire are obtained, the output format is modified to fit the requirements of the Bank of Spain's accounting database (GTC). It can then follow its usual course of subsequent validations and quality control. This format is of a key-value type.



7.2 Real data into production

Having the data from 2020, the steps of the method described earlier have been followed. Firstly, taking into account the validations from the Accounting Treatment Manager of the Bank of Spain (GTC), those items that are empty and whose sum does not match have been emptied (if the item is empty but the sum matches, it is understood that the item is zero). In this way, the distribution of the real number of missing values is as follows:



The data we have after applying these GTC validations is as follows:

- 1,899 questionnaires to be imputed for the assets (7,562 entries to be imputed)
- 3,627 questionnaires to be imputed for the liabilities (14,849 entries to be imputed)

Once the data is obtained, values are predicted using the pre-trained algorithms, and an accounting adjustment is performed for both asset and liability entries. By applying the two mentioned risk factors, 94% of the questionnaires for assets and 90% of the questionnaires for liabilities are imputed. In other words, the imputation results are as follows:

- 1,791 questionnaires imputed for assets (7,141 entries imputed)
- 3,268 questionnaires imputed for liabilities (13,015 entries imputed)

The final output format (key-value) for recording in internal applications is as follows:

id	clave	valor
14029130257	123009	47
14029130257	127009	89
14029130257	125009	68
14029130257	111009	0
14029130257	112009	194
14029130257	113009	441
14029130257	114009	0
14029130257	115009	86
14030078636	122009	9
14030078636	125009	9
14028856637	122009	40
14029120474	122009	1

With the idea of verifying how the algorithms perform imputations, the business team has analysed a series of questionnaires (randomly selected). These analysts have checked that, in general, the imputations made by the algorithms make sense and are based on logical criteria (values of entries from the previous year, logical relationships, etc.).

Once this output is defined, collaboration with the Information Systems Department is needed to launch this process on internal servers periodically (BATCH).

8. Next steps

The lines of research that can be continued are as follows:

- Deployment on the internal servers of the Bank of Spain.

- Search for better algorithms: Clearly, the more accurate the imputation algorithm is, and therefore the better it imputes microdata, the better the entire procedure will be. The current pace of development in the field of algorithms and data science requires staying vigilant for innovations that could enhance any of the projects already developed.

- Apply these techniques to the Income Statement questionnaires.

BANCO DE ESPAÑA
Eurosistema

MISSING VALUES IMPUTATION FOR CENTRAL BALANCE SHEET DATA OFFICE (CBSO) ACCOUNTING DATA

64TH WORLD STATISTICS CONGRESS

Iker González Crespo
Data Scientist at Central Balance Sheet Data Office
Statistics Department

15th July 2023



BANCO DE ESPAÑA
Eurosistema

INDEX

1. Business background

2. Data to impute

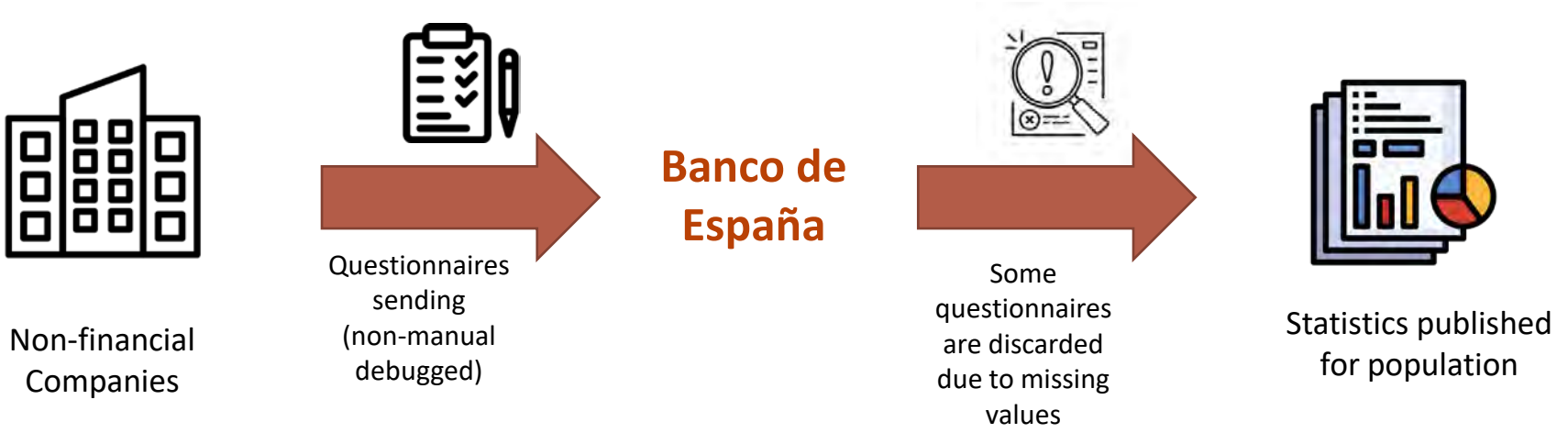
3. Methodology

    3.1 Train and validation
    3.2 Accounting adjustment
    3.3 Results

4. Production



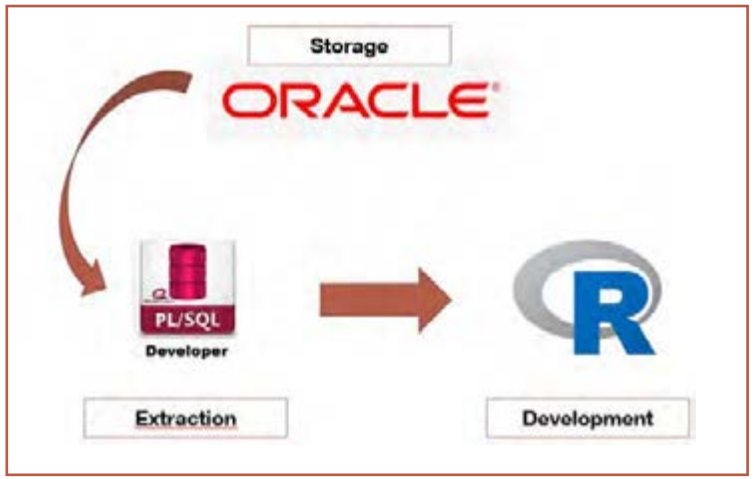
Banco de España receives almost 1 million of accounting questionnaries from non-financial companies. This information needs to be debugged by Central Balance Sheet Office (CBSO) in order to publish statistics for the spanish population.



What if we were able to impute these missing values, and thus:

- Enrich our databases
- Publish much more accurate statistics

GOAL OF THE PROJECT



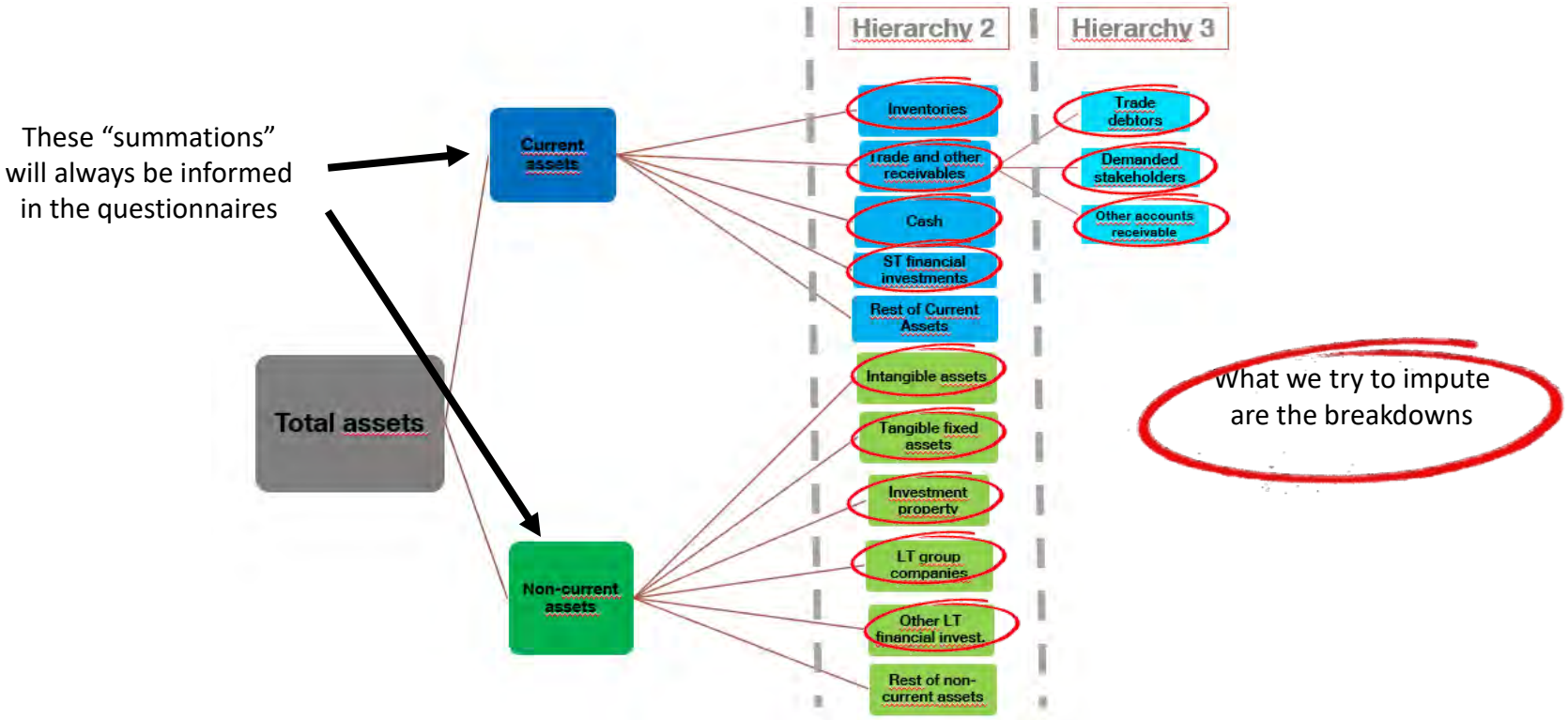
- The databases are stored in Oracle
- Information of questionnaires for years 2018 and 2020
- Usage of a Workstation due to computational cost
- Development has been conducted in R

- Focus in one specific type of questionnaire (only for SMEs)
- Select the balance variables (total assets, total liabilities and equity)
- There are 2 years to impute (current and last)

Example of questionnaire

ACTIVO		NOTAS DE LA MEMORIA	EJERCICIO 2020 (2)	EJERCICIO 2019 (3)
A) ACTIVO NO CORRIENTE	11000		424,71460	424,38489
I. Inmovilizado intangible	11100			0,02288
II. Inmovilizado material	11200		356,34554	366,62683
III. Inversiones inmobiliarias	11300			
IV. Inversiones en empresas del grupo y asociadas a largo plazo	11400			
V. Inversiones financieras a largo plazo	11500			
VI. Activos por impuesto diferido	11600		68,36903	57,73518
VII. Deudores comerciales no corrientes	11700			
B) ACTIVO CORRIENTE	12000		1.173,05586	1.068,56633
I. Activos no corrientes mantenidos para la venta	12100			
II. Existencias	12200		527,76339	499,86441
III. Deudores comerciales y otras cuentas a cobrar	12300		501,45543	420,93524
1. Clientes por ventas y prestaciones de servicios	12380		490,71858	396,46022
a) Clientes por ventas y prestaciones de servicios a largo plazo	12381			
b) Clientes por ventas y prestaciones de servicios a corto plazo	12382		490,71858	396,46022
2. Accionistas (socios) por desembolsos exigidos	12370			
3. Otros deudores	12390		10,73692	24,47502
IV. Inversiones en empresas del grupo y asociadas a corto plazo	12400			
V. Inversiones financieras a corto plazo	12500			19,47000
VI. Periodificaciones a corto plazo	12600			
VII. Efectivo y otros activos líquidos equivalentes	12700		143,83691	128,39668
TOTAL ACTIVO (A + B)	10000		1.597,77040	1.492,95122

The aim in this project is to impute all the balance sheet variables with missing values.
However, to be illustrative, it is only shown the relations between assets variables:



We get the debugged questionnaires in order to let our algorithms learn relations between all the variables in a proper way.

Debugged questionnaires

ID	Node	Trade	Trade debtors	Deman, stakehol.	Other account.
1	A1	73	55	13	5
2	B2	12	7	2	3
3	A1	34	30	1	3
4	B3	65	42	13	10
5	B2	98	65	3	30
6	B3	30	5	15	10

Generate 30% of missing values (randomly) for each accounting variable

Debugged questionnaires with missing values

ID	Node	Trade	Trade debtors	Deman, stakehol.	Other account.
1	A1	73	NA	NA	5
2	B2	12	7	2	3
3	A1	34	NA	1	NA
4	B3	65	42	13	10
5	B2	98	65	3	30
6	B3	30	NA	15	10

For each node (sector and size), **fit a ML model to impute values**

Debugged questionnaires imputed

ID	Node	Trade	Trade debtors	Deman, stakehol.	Other account.
1	A1	73	50	8	5
2	B2	12	7	2	3
3	A1	34	25	1	2
4	B3	65	42	13	10
5	B2	98	65	3	30
6	B3	30	2	15	10

Accounting adjustment (to match all the sums)

Debugged questionnaires imputed and with accounting adjustment

ID	Node	Trade	Trade debtors	Deman, stakehol.	Other account.
1	A1	73	59	9	5
2	B2	12	7	2	3
3	A1	34	30	1	3
4	B3	65	42	13	10
5	B2	98	65	3	30
6	B3	30	5	15	10

Accuracy metrics between imputed and real variables

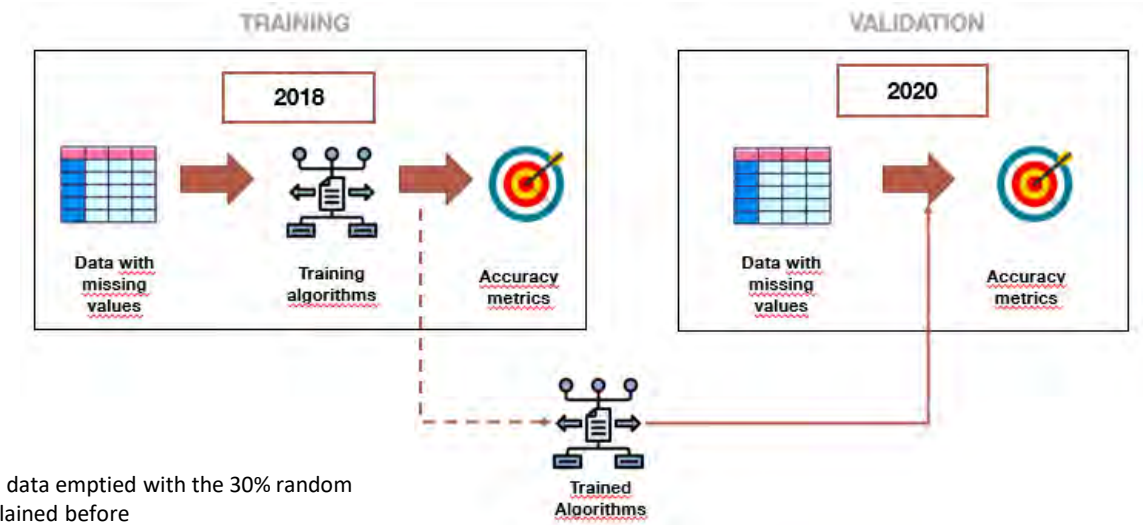


The ML algorithm used is **miceranger (link).** This is a technique based on ensembles of random forests. Some advantages of this model are:

- Fits very well for any kind of data (high accuracy metrics)
- Runs in parallel (speed up the process)
- Interpretable model (explain the imputations)

We **fit one model per node** (sector and size) **and for each year** (current and last).

Data from 2018 is used to train the best algorithms and save them. Afterwards, these models are used to predict data from 2020 (and check if they still fit well).



(*) Data with missing values: data emptied with the 30% random missing values technique explained before

Once all the imputations are set, there is a need to force all the sums to fit.
For every "summation" we created a factor to adjust the imputations:

$$factor = \frac{total - \sum non\ imputed\ variables}{\sum imputed\ variables} = \frac{difference}{\sum imputed\ variables}$$

Example

ID	Trade	Trade debtors	Demanded stakehol.	Other acc.	Difference	Factor
1	50	2	2	10	48	4
2	30	5	15	8	15	1,15

Real data Algorithm imputations

We **use this factor as a metric to balance the risk** we want to take when imputing questionnaires

Sensitivity analysis with interactive app to set the optimal factor

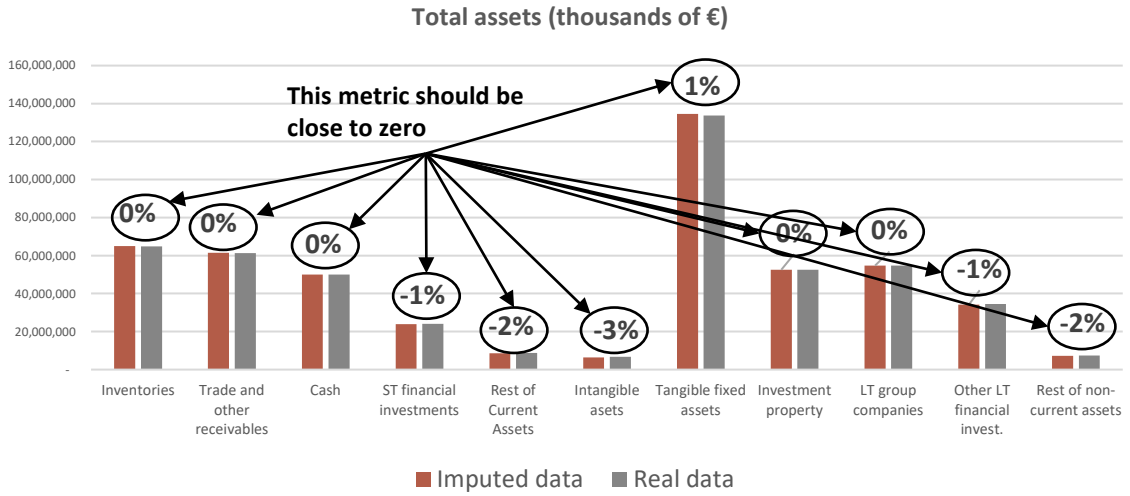
Note that depending on the sign of the variables, this metric can suffer different approaches. In the case of assets, all variables must be positive, so is easier.

A simulation has been conducted to analyze how well the statistics will hold after adding the imputations (and thus gaining sample).

The **results are significantly accurate** by many aggregations:

- Variable
- Size
- Sector
- Node (sector and size)

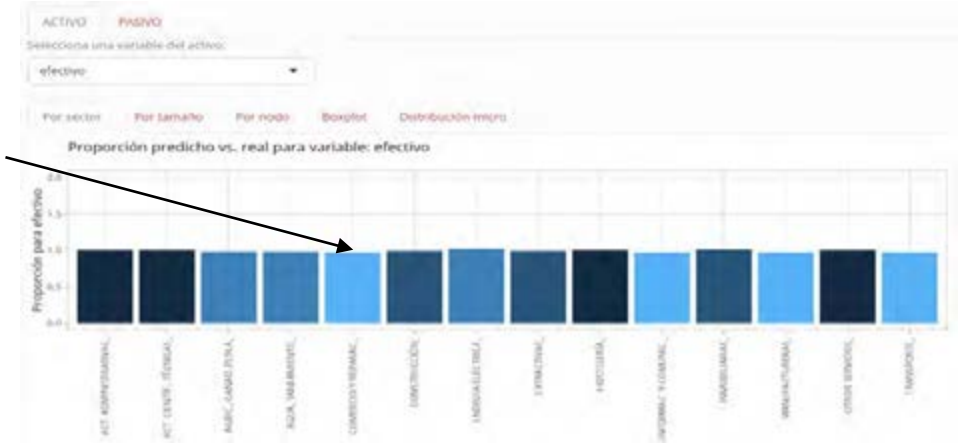
Variable-level



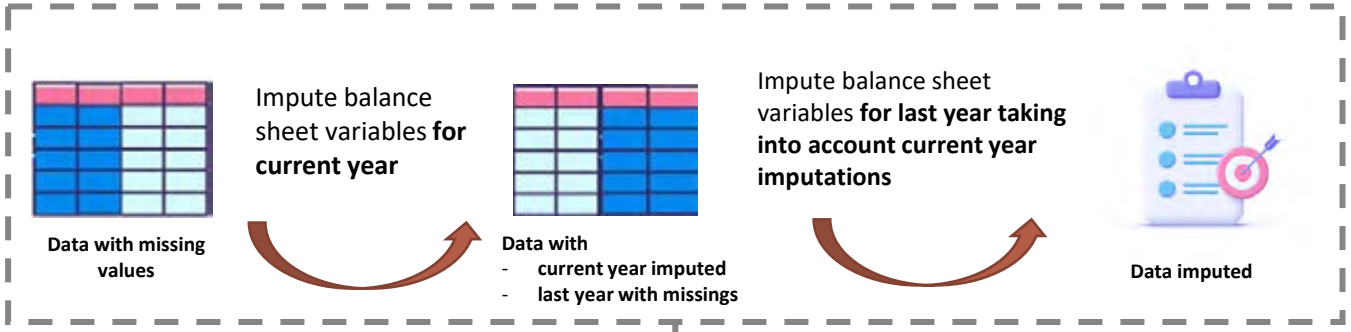
Size-level



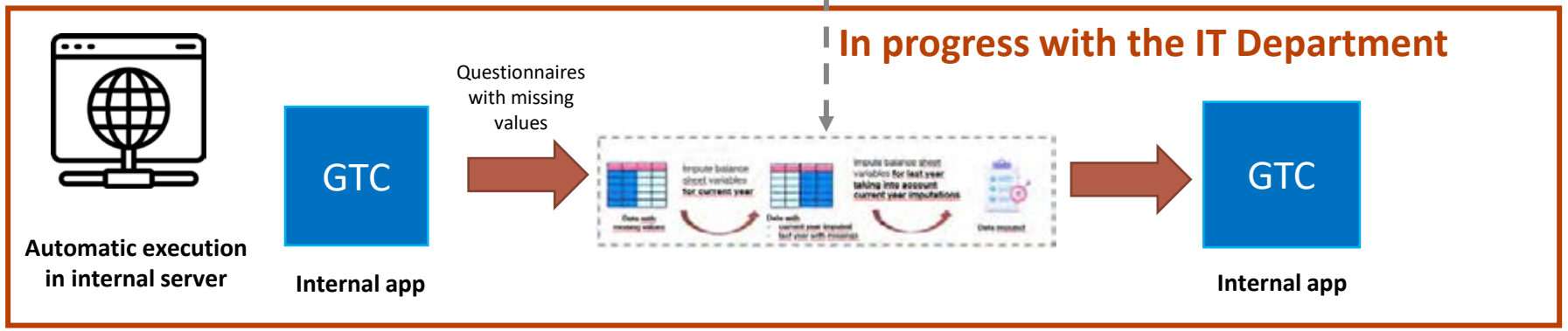
Sector-level



Dealing with streaming real data, we need to link these models (current and last year) as a way to impute the whole variables of the questionnaires:



This development should be linked to the internal processes and applications of Banco de España, in order to communicate one with each other:



BANCO DE ESPAÑA
Eurosistema

THANKS FOR YOUR ATTENTION

STATISTICS DEPARTAMENT



IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Utility and confidentiality assessment of synthetic company data[1]

Eugenia Koblents Lapteva,
Bank of Spain

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

BANCO DE ESPAÑA
Eurosistema

UTILITY AND CONFIDENTIALITY ASSESSMENT OF SYNTHETIC FINANCIAL DATA

Synthetic data pilot in collaboration with the European Commission

Eugenia Koblents
Banco de España, Statistics Department, BELab data laboratory

BOC-IFC SATELLITE SEMINAR ON "GRANULAR DATA: NEW HORIZONS AND CHALLENGES FOR CENTRAL BANKS"

Bank of Canada, July 15, 2023

STATISTICS DEPARTMENT





INDEX

1. Introduction:

- Synthetic data pilot with the European Commission
- Data description and performed synthetization

2. Utility assessment:

- Utility metrics for fully synthetic data
- Fit-for-purpose and outcome-specific utility metrics
- Utility assessment performed at Banco de España

3. Privacy assessment:

- Disclosure risk in fully synthetic data
- Unique matches between real and synthetic data samples

4. Summary and conclusions

INTRODUCTION

Synthetic data pilot with the European Commission

- This presentation outlines Banco de España's participation in the **synthetic data pilot** launched by the European Commission in the summer of 2022, to support the creation of the EU Digital Finance Platform's **Data Hub**.

- The Data Hub aims to provide participant FinTech companies with access to non-public datasets to test their solutions. However, the data is often **highly confidential** and cannot be shared with third parties due to legal restrictions.

- **Synthetic data** generation was proposed as an alternative solution to enable **sharing of sensitive information without compromising privacy**. Synthetic data preserves the characteristics of the original data while ensuring confidentiality.

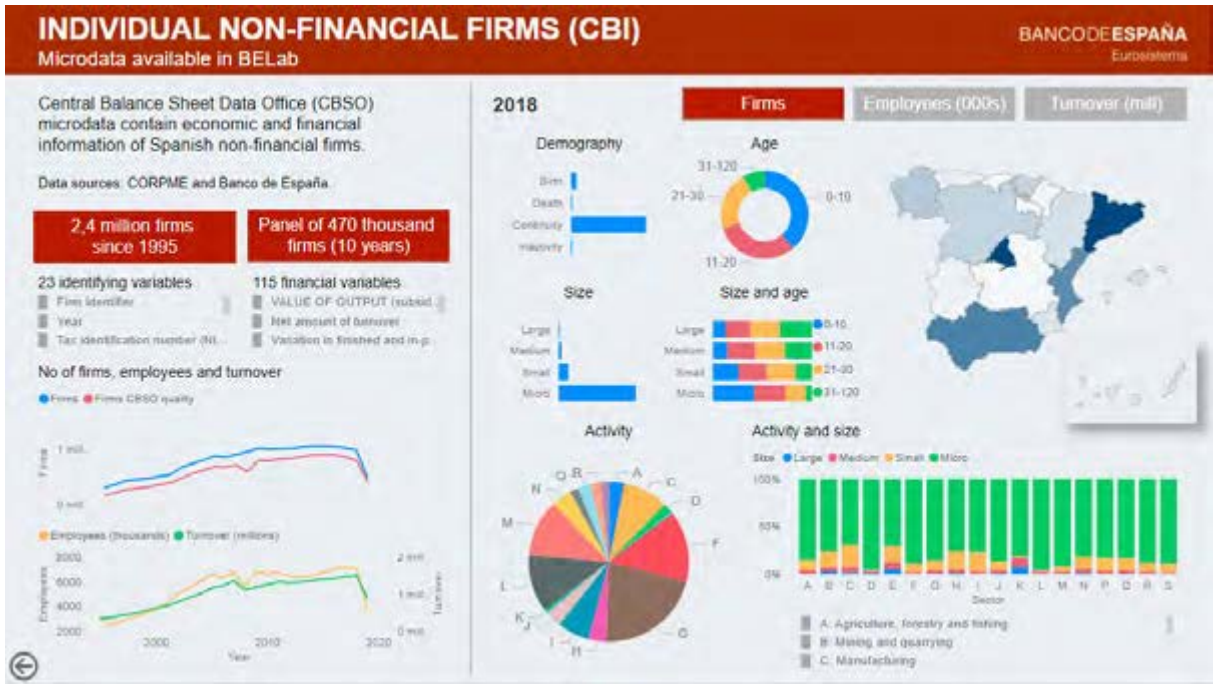
- Therefore, the pilot aimed to assess the **accuracy and privacy** of synthetic data and determine its suitability for different applications such as research, statistical analysis, IT development testing, and auditing.

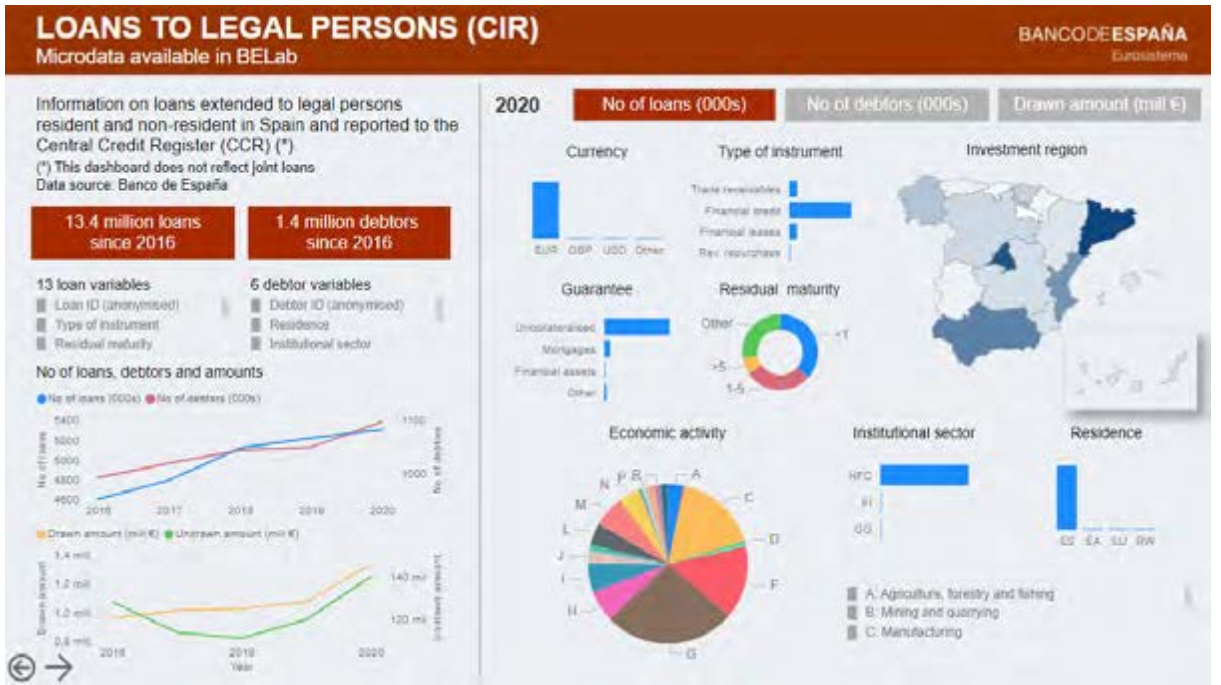
- **Pilot participants**:
  - **European Commission**: Pilot leader and coordinator, contracting authority.
  - **AI company:** Synthetic data software and training provider.
  - **Banco de España**: Data provider, synthetic data assessment.



BANCO DE ESPAÑA
Eurosistema

- **BELab** data laboratory and datasets used: https://www.bde.es/bde/en/areas/analisis-economi/otros/que-es-belab/
- **CBI** contains economic and financial variables built from the **annual accounts** reported by Spanish non-financial firms. Almost one million samples per year, 70 selected categorical and numerical variables.
- **CIR** contains sensitive information on **loans to legal entities** resident and non-resident in Spain. 5 million samples per year, 19 categorical and numerical variables. Multiple time series per debtor and loans with multiple debtors (joint loans).





INTRODUCTION

Data description and performed synthetization

- The software tool used implements the **full pipeline of fully synthetic data generation:** data upload, model training parameters configuration, execution supervision, and results analysis. The tool integrates a web-based user interface.

- According to the utility and privacy metrics computed by the software, **CBI and CIR datasets** have been successfully synthesised **independently and jointly**:

Synthesised datasets		Number of variables	Number of samples	Accuracy	Privacy tests
CBI	Static	8 numeric, 3 categorical	1,238,393	98%	Passed
	Dynamic	55 numeric, 1 categorical	4,067,062	95%	Passed
CIR (anon)	Static	5 categorical	699,480 (sample)	99%	Passed
	Dynamic	5 numeric, 7 categorical	10,434,349	98%	Passed
CBI-CIR	Static	8 numeric, 8 categorical	100,000 (sample)	99,4%	Passed
	Dynamic	59 numeric, 8 categorical	560,056	97,5%	Passed

- **Additional tests** have been conducted at Banco de España to complement the evaluation analysis.

- CBI and CIR datasets have been used for **utility and privacy assessment**, respectively.



INDEX

1. Introduction:
   - Synthetic data pilot with the European Commission
   - Data description and performed synthetization

2. Utility assessment:
   - Utility metrics for fully synthetic data
   - Fit-for-purpose and outcome-specific utility metrics
   - Utility assessment performed at Banco de España

3. Privacy assessment:
   - Disclosure risk in fully synthetic data
   - Unique matches between real and synthetic data samples

4. Summary and conclusions

UTILITY ASSESSMENT

Utility metrics for fully synthetic data

- Evaluating the utility of the generated data is a pivotal step in any synthetic data project, to identify the **most suitable synthesis strategy** and to decide whether the data are of **sufficient quality** to be released.

- **Utility metrics for fully synthetic data can be broadly divided into three categories:**

  - **Global utility metrics** assess the utility by directly comparing the original and the synthetic data (Kullback-Leibler divergence, propensity score matching, etc.), making no assumptions on the type of analyses to be conducted.

    Utility is measured on a **very aggregated level**, so good results do not guarantee high utility for specific analyses.

  - **Outcome-specific utility metrics** measure the utility for a specific analysis (means or regression coefficients for original and synthetic data, confidence interval overlap, etc.).

  - **Fit-for-purpose measures** provide a first impression of the quality of the synthetic data (marginal and bivariate distributions, consistency checks, goodness-of-fit measures, etc.).

[1] Drechsler, J., & Haensch, A. C. (2023). 30 Years of Synthetic Data. *arXiv preprint arXiv:2304.02107*.

[2] Drechsler, J. (2022, September). Challenges in Measuring Utility for Fully Synthetic Data. PSD 2022, Paris.

[3] Synthetic Data for Official Statistics, UNECE, 2022: https://unece.org/sites/default/files/2022-11/ECECESSTAT20226.pdf

UTILITY ASSESSMENT

Fit-for-purpose and outcome-specific utility metrics

- The synthetic data software implements several **graphical fit-for-purpose measures** (comparison of univariate and bivariate distributions, correlation matrices and pairwise accuracy matrix), yielding very good results in this experiment.

- Furthermore, the team at Banco de España has conducted additional **outcome-specific utility analysis** to evaluate the quality of the generated CBI synthetic data, with a focus on Banco de España's use cases.

- Two critical factors determine the utility of the synthetic **CBI dataset** for common use cases:

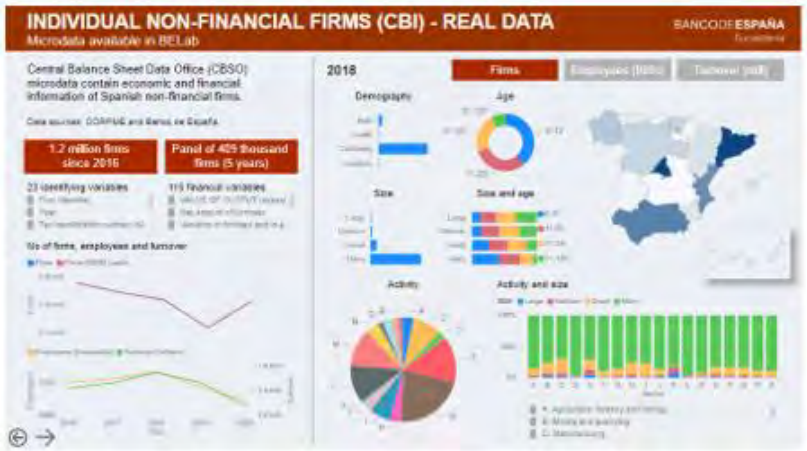
  - Accurate replication of the original **statistical distributions**: To validate this, various aggregated **summary statistics** obtained from real and synthetic data have been compared.

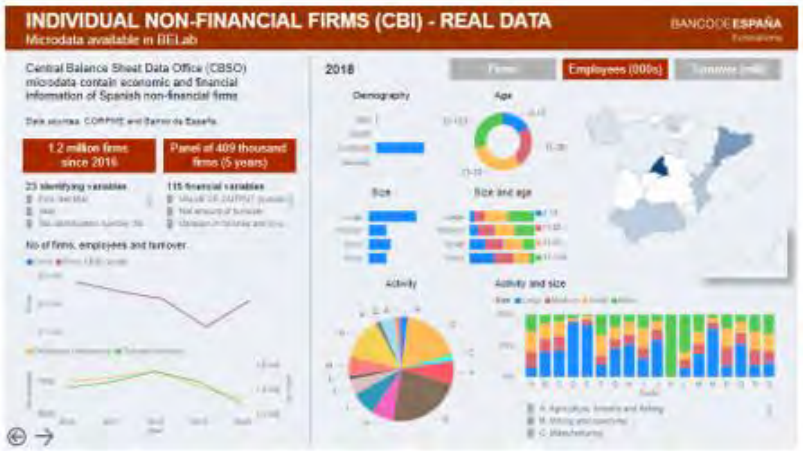
  - Preservation of **arithmetic relationships** among variables describing balance relationships: This aspect has been evaluated by comparing the **unbalance** between real and synthetic variables, respectively.

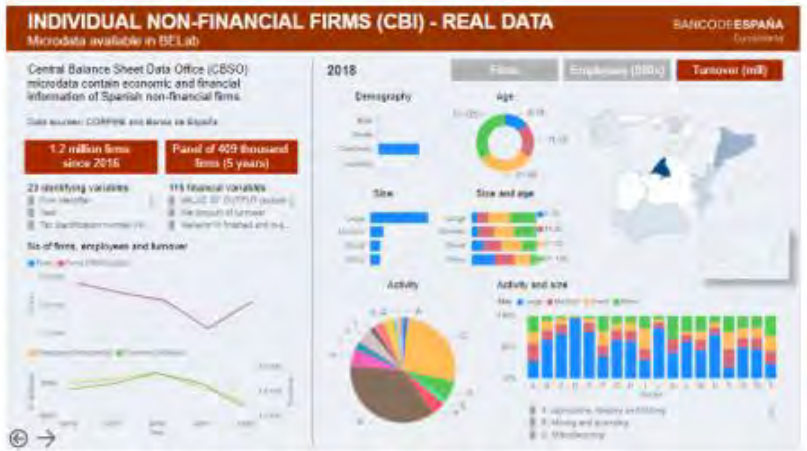
- Additional **comprehensive utility assessments** have been conducted by the **Suptech** and **Financial Innovation** teams at Banco de España, involving the comparison of machine learning models trained on real and synthetic data, descriptive time series analysis, and adversarial validation [4], reaching similar conclusions to the ones presented here.

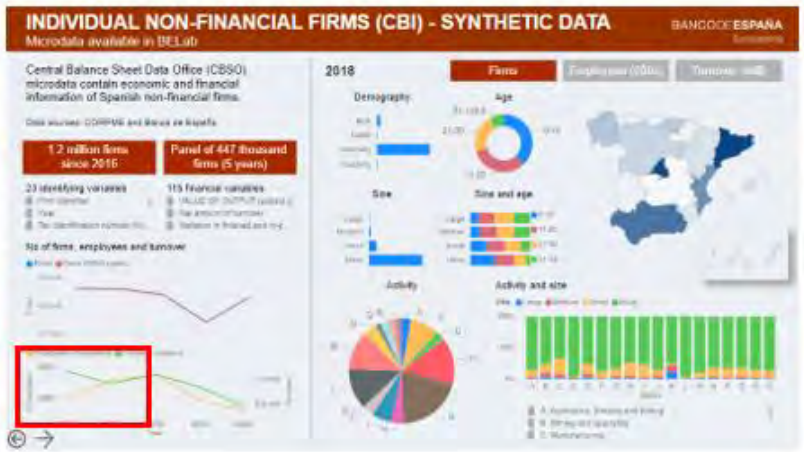
[4] Alonso, A., & Carbó, J. M. (2022). Accuracy of Explanations of Machine Learning Models for Credit Decision.

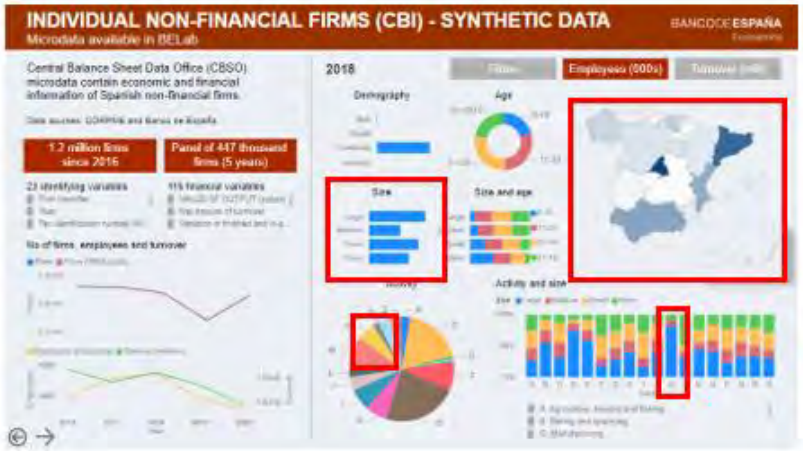
- **Global aggregates computed for BELab dashboards**: deviations in total amounts are observed, specially for large companies. This is due to **outlier suppression,** which is performed in order to avoid membership inference attacks.

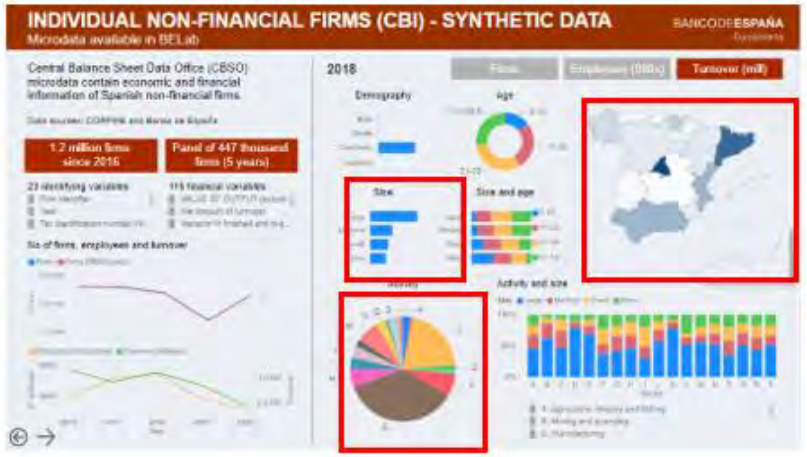




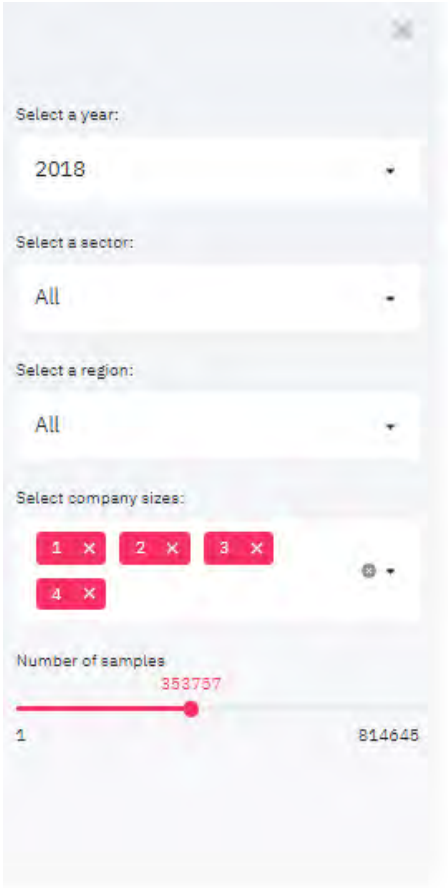


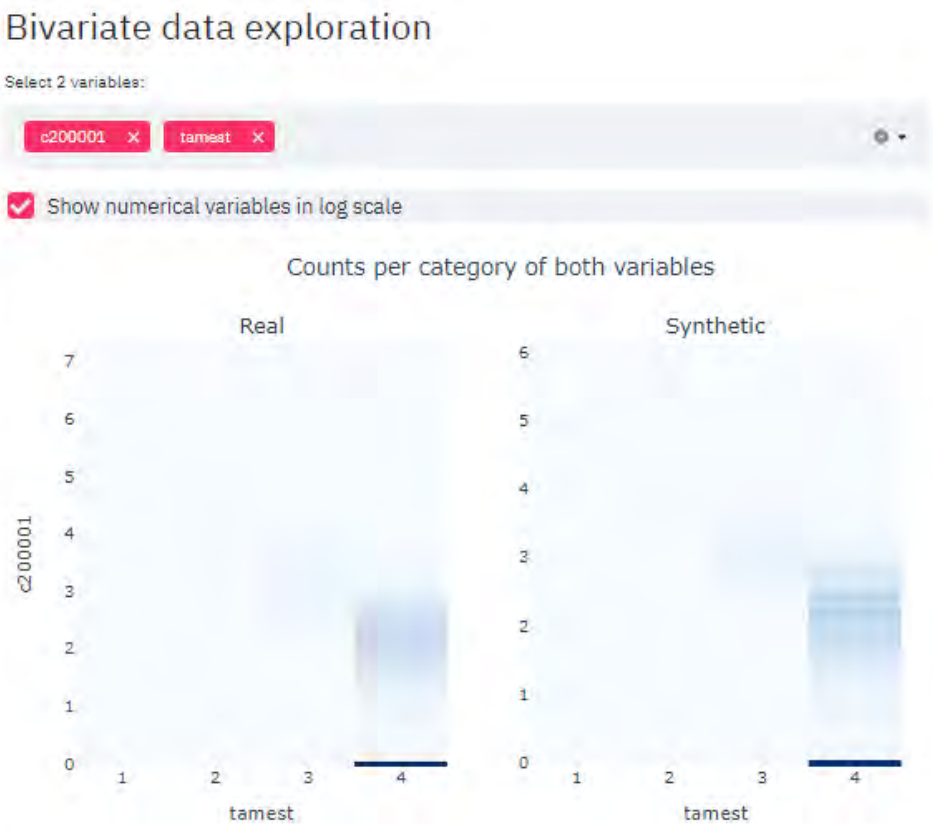


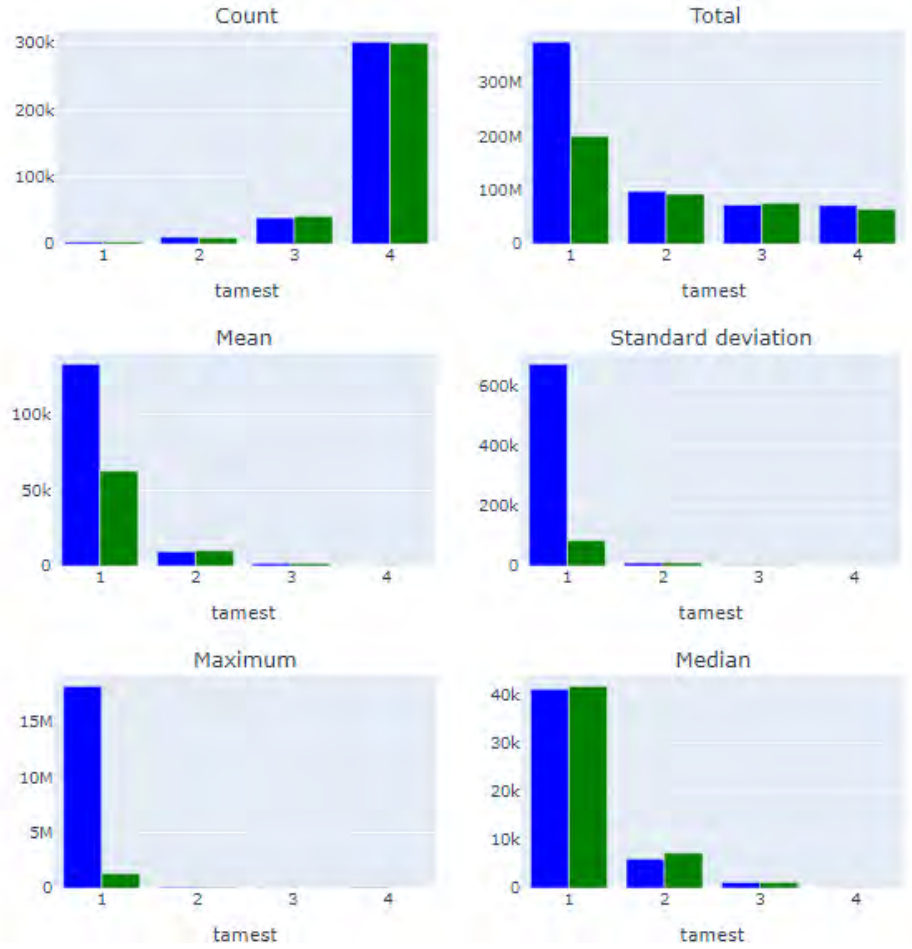




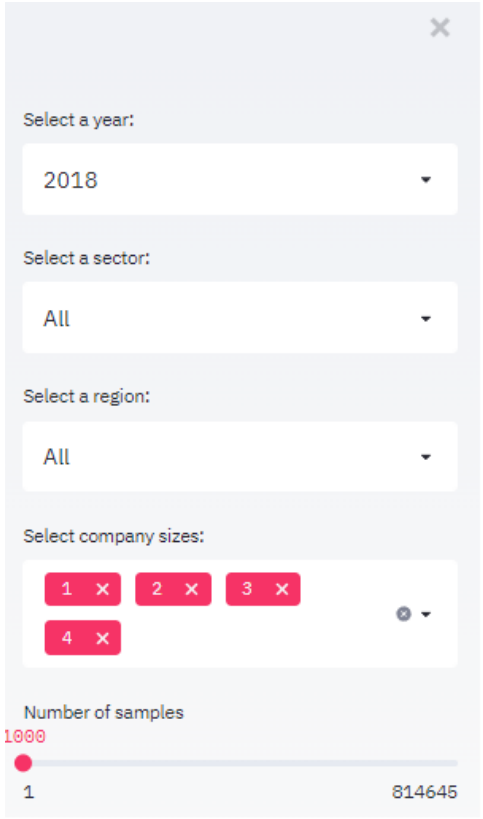
- **Analysis of summary statistics for sales per company size**. The number of companies (count) and the median are well reproduced but deviations are observed in the total, mean, maximum and std values.

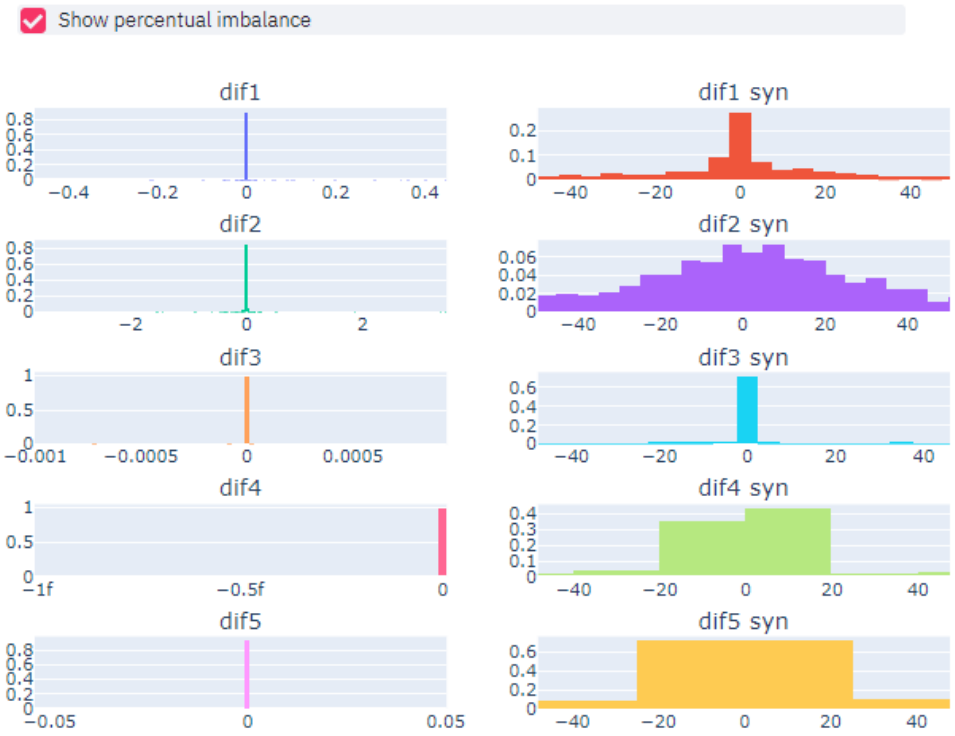






- **Analysis of linear relationships:** c200135 (total assets) = c200115 (non-current assets) + c200134 (current assets)
- In the real dataset most deviations are below 0.2% while in the synthetic dataset many samples present deviations above a 20%. Only **6% of samples** present an imbalance below 10% for these five linear relationships (many more exist).





Analysis of linear relationships in the data

Total imbalance 1 (c200115 = c200089 + c200098 + c200103): 835 vs 560346601

Total imbalance 3 (c200145 = c200142 + c200143 + c200144): 72 vs 423889005

Total imbalance 4 (c200158 = c200146 + c200151): 1 vs 239324366

Total imbalance 5 (c200180 = c200159 + c209166 + c209179): 716 vs 266960469

Total imbalance 6 (c200135 = c200115 + c200134): 1668 vs 748815588

Discard null values

Percentage of companies with an imbalance below a thresold

Maximum percentage of imbalance

10

Balance 1: 99 vs 54

Balance 2: 99 vs 26

Balance 3: 99 vs 80

Balance 4: 100 vs 80

Balance 5: 99 vs 59

Balance 6: 99 vs 74

Combination of all: 99 vs 6



INDEX

1. Introduction:
   - Synthetic data pilot with the European Commission
   - Data description and performed synthetization

2. Utility assessment:
   - Utility metrics for fully synthetic data
   - Fit-for-purpose and outcome-specific utility metrics
   - Utility assessment performed at Banco de España

3. Privacy assessment:
   - Disclosure risk in fully synthetic data
   - Unique matches between real and synthetic data samples

4. Summary and conclusions

CONFIDENTIALITY ASSESSMENT

Disclosure risk in fully synthetic data

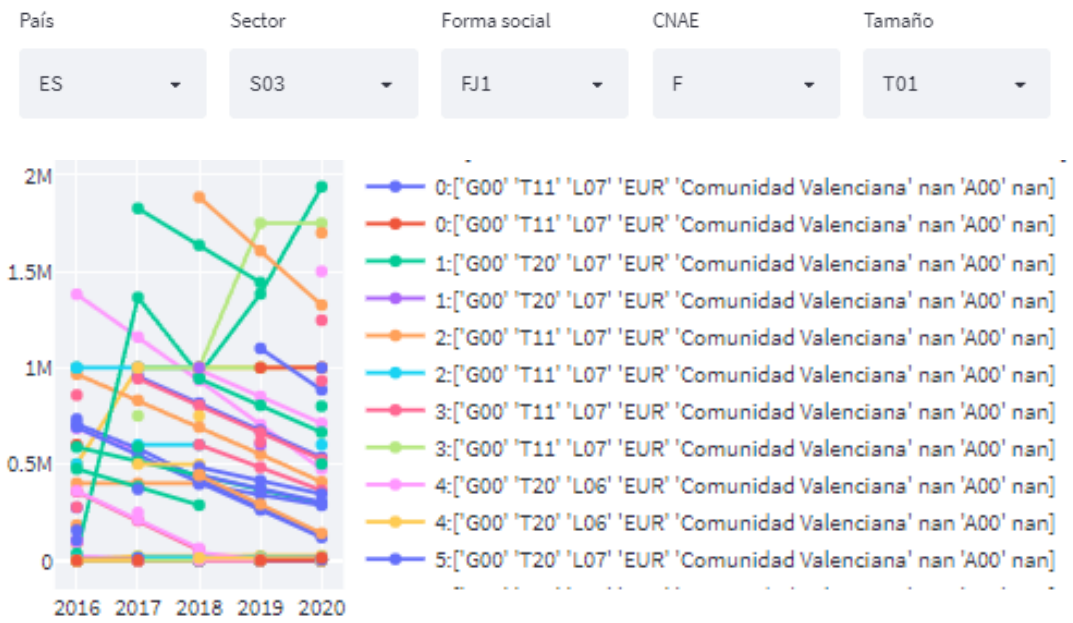
- No record in a fully synthetic data file corresponds to a real individual. Therefore, measuring the **risk of re-identification** is not applicable to fully synthetic data [1].
- However, National Statistical Authorities (NSA) often resort to disclosure risk metrics to avoid **reputational risk** [3]. Respondents might be concerned if they come across a synthetic record that exactly matches their own, especially if their combination of attributes is unique.
- Common disclosure risk measures for fully synthetic data include evaluating the **distance** between synthetic and real data records and the proportion of **unique matches**. In order to minimize reputation loss, these unique matches are often removed from the synthetic dataset.
- The synthetic data software employs various **distance-based metrics** to identify synthetic samples that closely resemble real ones. The synthesized datasets have **successfully passed** all of these tests.
- Furthermore, at Banco de España, **unique matches** in the CIR dataset have been identified to ensure that no real individual can be re-identified from the synthetic dataset to help assess and minimize **reputational risk**.

CONFIDENTIALITY ASSESSMENT

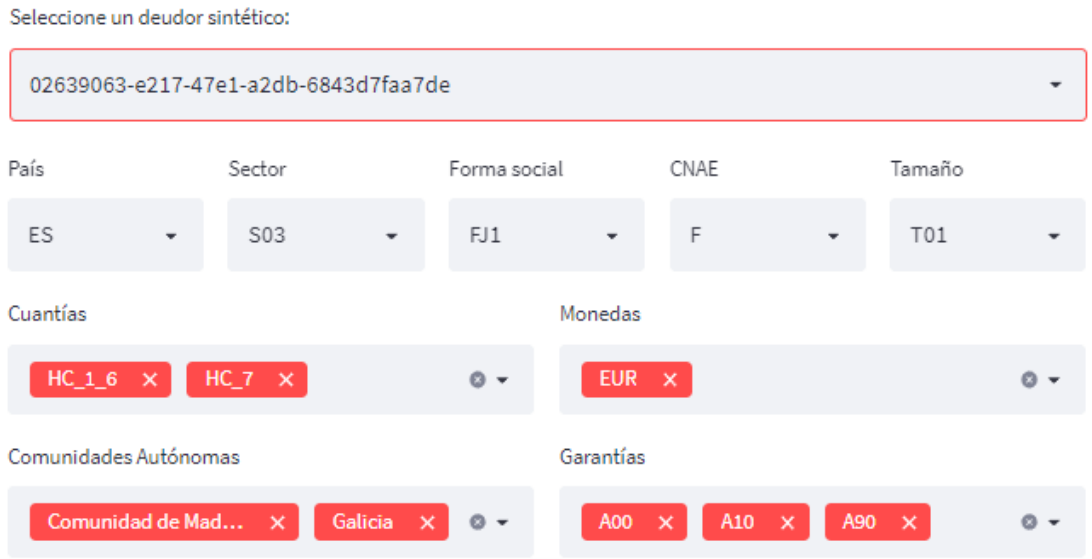
Unique matches between real and synthetic data samples

- The CIR dataset consists of **static debtor variables** (company size, sector of activity, etc) and a variable number of **time-series** describing loans, with variable length and attributes (amounts, currency, region of investment, etc).
- To facilitate distance computation, fixed-length **debtor profiles** were created and compared. Summarized time-series information was incorporated using one-hot encoding.

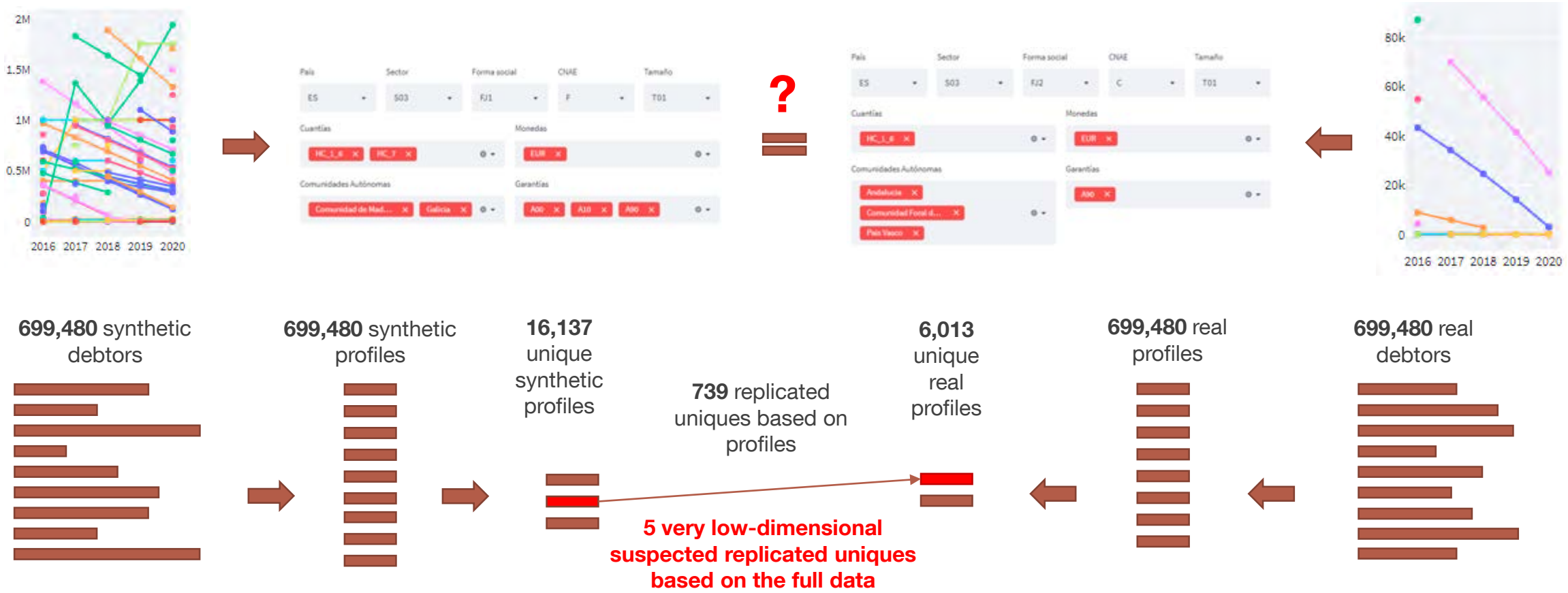
Full debtor's information



Summarized debtor's profile



- Fixed-length low-dimensional **debtor profiles** were created and compared.



- As the **dimension** of time series increases, the probability of exact matches decreases exponentially.
- Unique matches do not necessarily imply disclosure risk, but they have been supressed to **minimize reputational risk**.



INDEX

1. Introduction:
   - Synthetic data pilot with the European Commission
   - Data description and performed synthetization
2. Utility assessment:
   - Utility metrics for fully synthetic data
   - Fit-for-purpose and outcome-specific utility metrics
   - Utility assessment performed at Banco de España
3. Privacy assessment:
   - Disclosure risk in fully synthetic data
   - Unique matches between real and synthetic data samples
4. Summary and conclusions

SUMMARY AND CONCLUSIONS

- A **pilot on synthetic data generation and evaluation** has been conducted at Banco de España in collaboration with the EU Commission and a private synthetic data generation software provider.
- The **main outcomes** of the pilot are the following:
  - The software allows to accurately reproduce **univariate and bivariate distributions** as well as the **correlations** present in the real data and guarantees the **privacy** of individual samples.
  - However, the provided **fit-for-purpose utility metrics** do not guarantee high utility for any analysis, which needs to be evaluated for each dataset and use case. Specific observations:
    - Outlier suppression heavily affects **summary statistics**. Datasets consisting of SMEs only and variables describing ratios instead of absolute values appear to be less sensitive to outlier suppression.
    - **Linear and non-linear relationships** among variables are often not preserved.
    - **Subsequent analyses** of the generated synthetic data have yielded similar conclusions regarding data utility [4].
  - **Merging synthetic datasets** with different, correlated variables is not possible with the technology used.
  - **No significant privacy issues** have been identified. Only a very low number of low-dimensional unique matches have been identified, which should be supressed to minimize reputational risk.

BANCO DE ESPAÑA
Eurosistema

Thank you for your attention!

STATISTICS DEPARTMENT



Irving Fisher Committee on
Central Bank Statistics

BIS

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

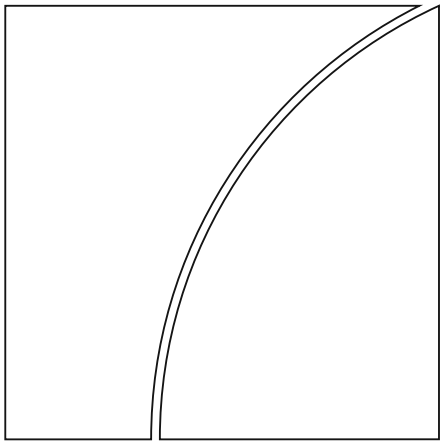
Finding a needle in a haystack: a machine learning framework for anomaly detection in payment systems[1]

Ajit Desai and Jacob Sharples,
Bank of Canada

and

Anneke Kosse,
Bank for International Settlements

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.



BIS Working Papers

No 1188

Finding a needle in a haystack: a machine learning framework for anomaly detection in payment systems

by Ajit Desai, Anneke Kosse and Jacob Sharples

Monetary and Economic Department

May 2024

JEL classification: C45, C55, D83, E42.

Keywords: payment systems, transaction monitoring, anomaly detection, machine learning.

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2024. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

Finding a Needle in a Haystack: A Machine Learning Framework for Anomaly Detection in Payment Systems*

Ajit Desai[1,‡], Anneke Kosse[2] and Jacob Sharples[1]

[1]Bank of Canada
[2]Bank for International Settlements

May 13, 2024

Abstract

We propose a flexible machine learning (ML) framework for real-time transaction monitoring in high-value payment systems (HVPS), which are a central piece of a country's financial infrastructure. This framework can be used by system operators and overseers to detect anomalous transactions, which—if caused by a cyber attack or an operational outage and left undetected—could have serious implications for the HVPS, its participants and the financial system more broadly. Given the substantial volume of payments settled each day and the scarcity of actual anomalous transactions in HVPS, detecting anomalies resembles an attempt to find a needle in a haystack. Therefore, our framework uses a layered approach. In the first layer, a supervised ML algorithm is used to identify and separate 'typical' payments from 'unusual' payments. In the second layer, only the 'unusual' payments are run through an unsupervised ML algorithm for anomaly detection. We test this framework using artificially manipulated transactions and payments data from the Canadian HVPS. The ML algorithm employed in the first layer achieves a detection rate of 93%, marking a significant improvement over commonly-used econometric models. Moreover, the ML algorithm used in the second layer marks the artificially manipulated transactions as nearly twice as suspicious as the original transactions, proving its effectiveness.

Keywords: Payment Systems, Transaction Monitoring, Anomaly Detection, Machine Learning
JEL Codes: C45, C55, D83, E42

*The views expressed in this paper are solely those of the authors and do not necessarily reflect those of the Bank of Canada, the Bank for International Settlements (BIS), the BIS Committee on Payments and Market Infrastructures (CPMI), or its members. We would like to thank Leonard Sabetti for his contribution during earlier stages of the work. We also thank Segun Bewaji, Alessio Brini, Narayan Bulusu, Ricardo De Avillez, Laura Felber, Marc Glowka, Tarush Gupta, Constanza Martinez, and Ellen van der Woerd for their detailed comments. In addition, we thank participants of the following conferences for their comments and suggestions: BIS Research Webinar Series (2022), Economics of Payments XI Conference (2022), DNB Central Bankers Go Data Driven Conference (2022), Canadian Economics Association Annual Conference (2022), RBI Global Conference on Financial Resilience (2023), International Conference on Economic Modeling and Data Science (2023), the Bank of Canada and the Bank for International Settlement's Seminar on Granular Data (2023), the Second CEMLA Regional Conference of Payments and Financial Market Infrastructures (2023), and the AEA Annual Meeting Poster Session (2024). ‡Corresponding author: adesai@bankofcanada.ca.

1 Introduction

High-value payment systems (HVPSs), such as Lynx in Canada, Fedwire in the US, Chaps in the UK, and Target2 in the Eurozone, are vital components of jurisdictions' financial systems. Typically, these are real-time gross settlement (RTGS) systems that process large-value transactions between financial institutions, often requiring settlement by a particular time. As such, the safety and efficiency of HVPSs are key to financial stability and economic growth. If not properly managed, an HVPS can be a source of a shock, such as payments fraud, a cyber attack, market stress, or operational problems (Chapman et al. 2015; BIS-Report 2019; FED-Report 2019; Kotidis and Schreft 2023). Moreover, as HVPSs provide a link between their participating financial institutions, they could become a channel through which shocks are transmitted across domestic or even international financial markets (Kosse and Lu 2022; Kotidis and Schreft 2023).

In particular, cyber attacks pose a growing risk to financial institutions and HVPSs.[1] For instance, in 2016, the Central Bank of Bangladesh (CBB) fell victim to a cyber heist, where hackers attempted to steal nearly one billion dollars from the CBB reserves account at the Federal Reserve Bank of New York (Bukth and Huda 2017). Similarly, cyber attacks on Mexico's interbank payment network and Banco de Chile in 2018 resulted in losses amounting to millions of dollars (Nish and Naumaan 2019).[2] Recently, Kotidis and Schreft (2023) has documented the effects of a cyber attack on a service provider to the banks participating in Fedwire and highlighted the importance of operational resilience. Moreover, simulation-based studies have shown that cyber attacks, even when targeting individual HVPS participants, can have a significant impact on the system in which they participate (Eisenbach et al. 2021; Kosse and Lu 2022; Docherty and Wang 2010). Therefore, real-time transaction monitoring for timely detection of anomalies in HVPSs is a way to enhance cyber resilience and avoid such shocks from having unintended system-wide consequences.

To enhance safety in HVPSs and, more broadly, to limit systemic risk and foster financial stability, various guidelines have been issued and many initiatives have been undertaken globally.[3] For instance, in 2016, the Bank for International Settlements (BIS) Committee on Payments and Market Infrastructures (CPMI) and the International Organisation of Securities Commissions (IOSCO) issued guidance to boost cyber resilience in financial market infrastructures (FMIs) like HVPSs.[4] This guidance stresses the importance of FMIs maintaining effective capabilities to monitor anomalous activity and outlines the tools and processes they should employ for detecting cyber incidents.

Despite these efforts, real-time transaction monitoring for anomaly detection in HVPSs faces several challenges, with the scarcity of anomalies and the absence of pre-identified examples being the primary obstacle. Additionally, the high frequency of payments, the extensive size of payment networks, the complex strategic interactions among participants,[5] and the limited availability of detailed transaction information[6] further complicate the detection process. As a result, there is room for improvement. For instance, HVPSs with capabilities to monitor payment flows mostly rely on rule-based and ad-hoc monitoring approaches. These approaches involve a delay and require prior assumptions about how anomalous payments would

[1] The Bank of Canada's recent financial stability review highlights that the financial sector globally has the largest share of reported cyber attacks. https://www.bankofcanada.ca/2023/05/financial-system-review-2023/.

[2] A comprehensive overview of cyber incidents involving financial institutions can be found at CEIP (2021).

[3] The Federal Reserve published strategies for improving safety in the US payment systems and provides models that help organizations classify fraudulent payment activity (FED-Report 2019). Likewise, the Eurosystem reported the development of a methodology enabling the operator to detect potentially anomalous transactions in their HVPS (TARGET-Report 2019).

[4] See https://www.bis.org/cpmi/publ/d146.pdf.

[5] In certain situations, participants might strategically alter their typical payment behavior due to liquidity constraints or to minimize payment delay costs (Bech and Garratt 2003; Castro et al. 2021). These factors are unrelated to cyber incidents or outages.

[6] Having additional information that provides details about transactions could offer supplementary features to learn and understand typical payment behavior (Glowka 2019; León 2020; Sabetti and Heijmans 2021).

look. Such assumptions may not cover all forms of anomalies, as it is impossible to anticipate all potential scenarios (FED-Report 2019; Arjani et al. 2020; Arjani and Heijmans 2020). Furthermore, anomaly detection tools employed by individual system participants only capture transactions to and from that particular participant, which limits their utility for system-wide transaction monitoring.

To mitigate these challenges, in this paper we propose a flexible, centralized, and layered transaction monitoring framework for pattern recognition and anomaly detection in HVPS, leveraging data-driven and nonlinear machine learning (ML) tools. In this framework, monitoring occurs centrally at the HVPS level, utilizing transaction data from all system participants to comprehend their payment patterns. This approach contrasts with the monitoring tools of individual participants, which focus only on their own transactions. Moreover, the framework consists of multiple layers to overcome the challenges due to the extensive size of usual payments and unavailability of pre-identified anomalies. In the first layer, a supervised ML algorithm is used to classify payments by submission time, effectively screening for typical, or usual, transactions. In the second layer, only the misclassified (unusual) payments are processed through an unsupervised ML algorithm for anomaly detection.[7] Using nonlinear ML-based models in each layer allows us to learn complex patterns from a large amount of historical data, eliminating the need for predetermined rules. Moreover, the proposed ML models in our framework do not require prior assumptions about the data-generating process or the structure of the potential anomalies, allowing for a more generalized detection process. Furthermore, both layers in our framework are independent and have flexible components capable of integrating additional ML models. Therefore, they can serve as robust tools for detecting anomalies within HVPSs.

We test our framework with manipulated transactions and actual-transaction data from Canada's former and current HVPS, the Large Value Transfer System (LVTS), and Lynx, respectively.[8] The results demonstrate that the proposed framework is a promising approach for transaction monitoring and anomaly detection in HVPSs. The gradient-boosting-based ML model used in the first layer outperforms a logistic regression and other ML models by up to 44%, and it exhibits higher out-of-sample accuracy when classifying transactions on non-regular days and artificially manipulated transactions.[9] In particular, our model demonstrates a notable ability to correctly detect 93% of all artificial transactions. Moreover, the isolation forest (IF) model used in the second layer successfully assigns higher scores to manually altered anomalous transactions—on average, twice as high as their original counterparts. Also, we perform various scenario analyses, which demonstrate that the framework is flexible enough to be applied for different payment system designs and to be extended with other features to even further improve its robustness.

To gain insight into the primary features driving the anomalous payments identified by our algorithm, we employ the Shapley value-based SHAP approach (Lundberg and Lee 2017) to interpret the results of the ML models used in both layers. We find that for a given HVPS transaction, the time elapsed since the previous incoming transaction and basic transaction features, like sender-receiver pair and payment amount, effectively predict payment submission patterns. By contrast, the more complex intraday features, such as the time elapsed since specific types of incoming and outgoing transactions, assume a prominent role in identifying and isolating anomalies. We also show that the SHAP approach can help system operators in assessing the nature and severity of anomalies, enabling timely interventions and prompt action.

[7] Our two-step approach is analogous to the two-step process of airport security screening for passengers, in which the first step involves initial machine screening, and only passengers who fail this screening undergo detailed inspections.

[8] We use payments settled between January 2011 and August 2021 in LVTS and from October 2021 to August 2023 in Lynx. Note that we excluded the month of September from the dataset due to the transition from LVTS to Lynx on August 29, 2021.

[9] To prevent seasonal influences from affecting the model, we identify and exclude "special days," or non-regular, days from our training sample. This helps us avoid the model solely identifying payments on such special days as anomalous. These days include Canadian provincial holidays, US national holidays, days with known operational incidents, and the Covid-19 period.

The remainder of the paper is structured as follows. In section 2 we provide further background and explain the key contribution of our paper. Section 3 presents an overview of our proposed framework for classifying and identifying anomalies, including the ML models used in each layer. This is followed in section 4 by a description of the data and transaction features used for the analyses. The results and implications are presented in section 5, and scenario analyses showing the framework's adaptability for different payment system designs are presented in section 6. We conclude in section 7.

2 Background and Contribution to Existing Literature

Changes in timing and patterns of HVPS transactions have historically been studied to understand the impact of economic disruptions, such as the 2008 global financial crisis (Bech and Garratt 2012; Massarenti et al. 2012; Alexandrova-Kabadjova et al. 2015; Zhang 2015), and to investigate liquidity and delay trade-offs faced by HVPS participants (Bech and Garratt 2003; McAndrews and Rajan 2000; Martin and McAndrews 2008; Castro et al. 2021; McMahon et al. 2022). However, given the importance of the availability and smooth functioning of HVPSs for the broader financial system and with cyber threats growing over time, there is a strong case for HVPS transaction monitoring in the context of operational risks, such as cyber events and operational incidents (Chapman et al. 2015; Eisenbach et al. 2021; Kotidis and Schreft 2023).

Only recently, driven by the increasing risk of cyber threats and operational outages, researchers have started developing tools for monitoring transactions. Given the high frequency of payments settled in HVPSs, this work mostly relies on advanced analytical tools using ML. For instance, Glowka (2019) applies clustering techniques on Target2 data to identify payment profiles of banks and Arévalo et al. (2022) use an ML model to identify clusters of anomalous payments in the Salvadorian payment system. Artificial neural network (ANN) models have also been increasingly used to identify anomalous payment flows between participants. For instance, Triepels et al. (2017) employ an autoencoder, an unsupervised ANN methodology that is able to recognize complex patterns, and León et al. (2020) use an ANN model for identifying patterns in participants' payment behavior. Similar models have been set up by Rubio et al. (2020) and Sabetti and Heijmans (2021) to classify and identify anomalous payments in the HVPS of Ecuador and the retail payment system in Canada, respectively.

Two common challenges faced in this research field thus far are the rarity of known anomalies and the large size and complexity of payment networks. In our paper, we develop a framework for anomaly detection that addresses these two challenges. Typically, HVPSs settle thousands of transactions without systematically monitoring for anomalies, resulting in limited or no pre-identified anomalies to learn from. Additionally, the timing of these transactions are driven by multiple factors such as business requirements, liquidity constraints, time sensitivity, and intricate bilateral and multilateral business relationships among participants. Existing work based on clustering-based methods, for instance, requires the identification of the optimal number of clusters, which may be particularly challenging with high-dimensional data and outliers. This, in turn, leads to a slow training process when using large datasets. Likewise, the main limitation of ANN models is that they are hard to train on transaction-level data, require manual aggregation, and are difficult to interpret. In contrast, the framework proposed in this paper is easier to train on high volumes of transaction-level data. By using multiple layers and a combination of supervised and unsupervised tree-based ML models, we mitigate the challenges resulting from the rarity of anomalies and the large volume of payments. Moreover, our framework provides tools for the interpretation of the results to gain insight into the influential factors driving HVPS participants' usual and anomalous payment behavior.

Given the novelties and flexibility of our proposed framework, it contributes to the broader literature

on financial fraud detection and financial stability monitoring. The challenges related to dealing with large volumes of data and identifying rare anomalous events not only apply when monitoring HVPS transactions but are also relevant when identifying credit card fraud, illegal trading activities, financial statement fraud, and systemic financial instability (Ngai et al. 2011; Flood et al. 2016; Ryman-Tubb et al. 2018; James et al. 2023). Our framework can be adapted and applied to this broader set of applications.

3 Framework for Pattern Recognition and Anomaly Detection in HVPSs

Given the large size and complexity of payment networks and the rarity of known anomalies, we propose a framework that consists of two layers,[10] each using different ML models, that can be employed in near real-time (see Figure 1). In the first layer, a supervised ML classifier is used to classify settled HVPS payments based on a target variable.[11] This classifier allows us to separate the full set of HVPS payments into two subsets: correctly classified and misclassified transactions. This separation is crucial: it allows us to systematically filter out a large subset of typical transactions in the first layer such that we can focus on a smaller subset of payments in the second layer. Subsequently, we use this subset of misclassified transactions for anomaly detection with the help of an unsupervised ML algorithm in the second layer.

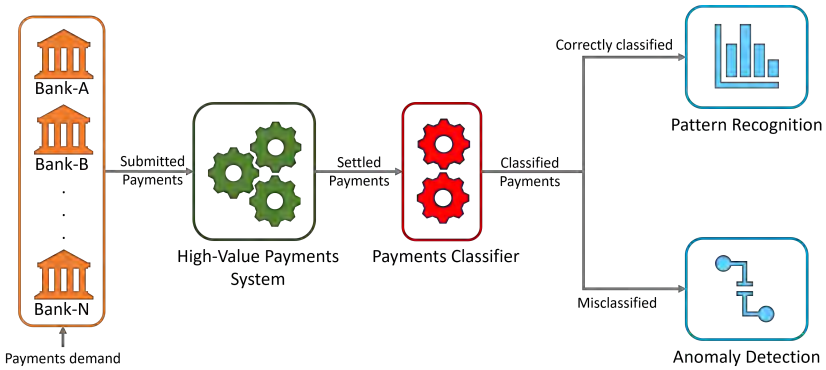


Figure 1: Schematic overview of the layered approach proposed in this paper for transaction monitoring in HVPSs. All settled HVPS payments pass through a payments classifier. Subsequently, the subset of misclassified payments is sent to the anomaly detection model, while the correctly classified payments are used to study the usual payment patterns of the HVPS participants.

The key advantage of using the layered approach, combining supervised and unsupervised ML tools, is its ability to separate typical payments in the first layer. This assists in mitigating difficulties arising from high volume and helps address the challenges due to the rarity of anomalies. Moreover, the first layer is flexible and independent from the second layer, allowing the integration of additional algorithms to enhance robustness. For instance, multiple classification algorithms can be used simultaneously, each with a different classification target to improve the performance of the first layer and ease the subsequent task of anomaly detection (see section 6.1 for an example of a joint classifier). Furthermore, our framework allows the use of the subsets of correctly classified and misclassified payments from the first layer to study the usual and unusual payment patterns of HVPS participants in the second layer. To this end, we employ the Shapley value-based (SHAP) approach to assess the impact of various transaction features on the timing of payments and to gain a deeper understanding of which features are more likely to uncover anomalies.

[10] Having sufficient pre-identified anomalies could have eliminated the need for a layered approach. However, obtaining a sufficiently large set of labeled data covering a diverse set of anomalies is challenging for many applications, including HVPS.

[11] Note that the classifier module in the first layer is adaptable; it can incorporate multiple ML algorithms and handle various target variables. Additionally, it can be expanded to perform regressions with continuous targets.

3.1 Layer 1: Payments Classifier

As a first step, we run all HVPS payments through one or more *payments classifiers*. These classifiers use supervised learning models with categorical targets, which could be binary or multi-class. The selection of target variables can be tailored based on the characteristics of the payment systems. For simplicity, we choose a binary classifier based on the payment submission period, which classifies payments as either morning or afternoon payments. Morning payments are defined as those that are submitted to the HPVS before noon, whereas afternoon payments are those submitted in the afternoon.

The model can be formulated as follows: $X = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^M\}$ represents a set of M predictors (also known as features or independent variables), each being a vector of N data points (samples), and $\mathbf{y}$ is the target variable (or dependent variable), also containing N data points. In our case, $\mathbf{y} = \mathbf{1}, \mathbf{2}$, where 1 represents the label for the set of morning payments and 2 represents the afternoon payments. We denote $\hat{\mathbf{y}}$ as the predicted target (i.e., the submission time of the payment), which can be estimated using the classification model $f : \hat{\mathbf{y}} = f(X)$.

A commonly used econometric model for such a classification is a logistic regression that analyzes the relationship between the target variable and the predictor variables and that generates probabilities for the target variable belonging to one of the two groups. Logistic regression models are generally valued for their simplicity and interpretability. However, their performance may be limited when dealing with large and complex datasets. In particular, challenges arise when there are many continuous and categorical predictors and when dealing with nonlinearity, collinearity, and outliers (Hastie et al. 2009).

In light of these challenges and given the nature of our data, we instead use a decision-tree-based ensemble learning approach. Random forest and gradient boosting models are examples of such an approach and are based on flexible nonlinear ML algorithms with a demonstrated ability to perform well for complex categorical learning tasks. These non-parametric models can effectively handle collinearity, nonlinearity, outliers, and datasets containing both continuous and categorical predictors, making them suitable for our analysis (Friedman 2001; Friedman et al. 2001). We primarily use the Light Gradient Boosted Machine (LightGBM) algorithm. While we explored various other models, including logistic regression, decision trees, random forests, and standard gradient boosting, LightGBM stands out for several reasons. It is an open-source library that allows for a fast and efficient implementation of the gradient boosting algorithm by focusing on boosting examples with larger gradients. It also adds a type of automatic feature selection that has been demonstrated to generate better predictive performance (Ke et al. 2017). Particularly for handling very large datasets like HVPS data, such a scalable and efficient model is useful for the purpose of fast training. See Appendix A for more details on the LightGBM algorithm and the implications of its use.

3.2 Layer 2: Anomaly Detection on the Subset of Missclassified Payments

In the subsequent step, we run the subset of misclassified payments from the first layer through an ML-based anomaly detection model. As there are commonly no pre-identified anomalous transactions in HVPS datasets, we are dealing with unlabelled data. Therefore, we use an unsupervised learning-based IF algorithm (Liu et al. 2008).[12]

The choice of the IF algorithm is motivated by its effectiveness in handling transaction-level data and its capacity to rank transactions according to abnormality levels. Additionally, its simplicity and shared characteristics with the chosen ML model in the first layer enhance its suitability. Like gradient boosting,

[12] We could have used unsupervised learning with the IF model without combining it with the payments classifier model as a first step. However, such a standalone IF model would exhibit relatively low accuracy, likely due to the large size and complexity of the HVPS payments. For more details, see Appendix B.

the IF algorithm uses decision trees (called isolation trees) and is based on the concept of isolating data points that are rare and significantly different from the majority of the data. The IF algorithm has been proven to be fast and memory-efficient, due to which it can handle high-dimensional and large-scale transaction-level data (Liu et al. 2008). Therefore, the IF approach is particularly well-suited for transaction monitoring and anomaly detection in HVPSs.

In short, the IF algorithm first randomly selects a subset of n data points from the set of M features ($X = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^M\}$), each with N sample points. Next, it creates isolation trees by recursively partitioning the data until either the maximum tree depth is reached or there is just one data point left. During each partitioning step, the algorithm randomly selects a feature ($\mathbf{x}^m$ from X) and a split point along that feature to divide the data into two subsets. The algorithm repeats these steps with different sub-samples to generate multiple trees. Finally, it computes the anomaly score for each data point by calculating the average number of splits (or path length) using all isolation trees. This average path length is then transformed into an anomaly score (with lower path lengths leading to higher scores), which indicates the likelihood of that data point being an anomaly. See Appendix B for more details on the IF and the implications of its use.

3.3 Interpretation Tools for Classification and Anomaly Detection Models

We use the Shapley value-based (SHAP) approach (Lundberg and Lee 2017) to interpret the outcomes of the LightGBM and IF models in both Layer 1 and Layer 2.[13] The SHAP approach allows us to examine how the characteristics of HVPS payments influence their submission time and the identification of anomalies. The SHAP values, which we visualize in so-called dependence plots, explain how specific payment features impact participants' usual payment patterns and how they contribute to the detection of anomalies therein.

We also use the SHAP approach to study individual transactions, known as local interpretation. We do so using force plots, which visualize the marginal contribution of transaction features with the help of arrows. These arrows either push the prediction scores up (towards an afternoon prediction) or down (towards a morning prediction). The length of the arrows indicate the magnitude of the SHAP value, which is analogous to the strength of the contribution of that transaction feature to the model's prediction. In addition, we use the SHAP approach to study subsets of payments, known as global interpretation. This approach provides the average impact of transaction features on the model's prediction for that subset of transactions. The local and global interpretation analyses help us gain insight into why a particular transaction or a subset of transactions are identified as unusual. This allows us to understand the influence of specific payment features on the submission time of HVPS payments.

Note that the SHAP approach does not provide causal inference or any optimal statistical criterion (Slack et al. 2020; Molnar 2020). Instead, its purpose is to explain the marginal contribution of each transaction feature to the difference between the models' predictions and the average prediction of the entire training sample. Therefore, its primary use is intuitively explaining the predictions of the models (Lundberg and Lee 2017). Further details on the SHAP approach are provided in Appendix D.

In the following sections, for simplicity, we primarily focus on demonstrating our framework's usefulness with a binary classifier, discussed in section 3, on a transaction sample from LVTS, presented in section 4. After that, in section 6, we showcase the framework's flexibility using a joint classifier in the first layer, and we extend the framework's application to other LVTS sub-samples and a data sample from Lynx.

[13] We use TreeSHAP, a variant of SHAP, which offers computational efficiency in estimating Shapley values for tree-based ML models, such as LightGBM. Although initially designed for supervised learning models, TreeSHAP can be adapted with minor modifications for unsupervised learning-based IF models (Lundberg and Lee 2017).

4 Data

4.1 LVTS Transaction Data

We test our proposed framework using around ten years of historical transaction data from the Canadian HVPS, Large Value Transfer System (LVTS), from January 2012 to August 2021. Our focus is on payments settled during the system's regular business hours, which is from 6 am to 6 pm. As such, we cover approximately 85 million transactions exchanged among 17 participants. The daily average transaction value is approximately 135 billion Can$ over a daily average transaction volume of 35,000 transactions.

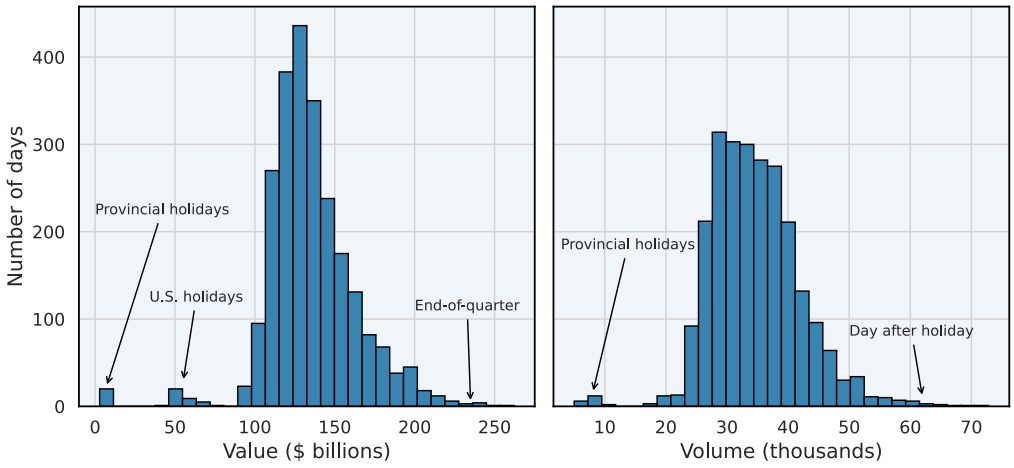
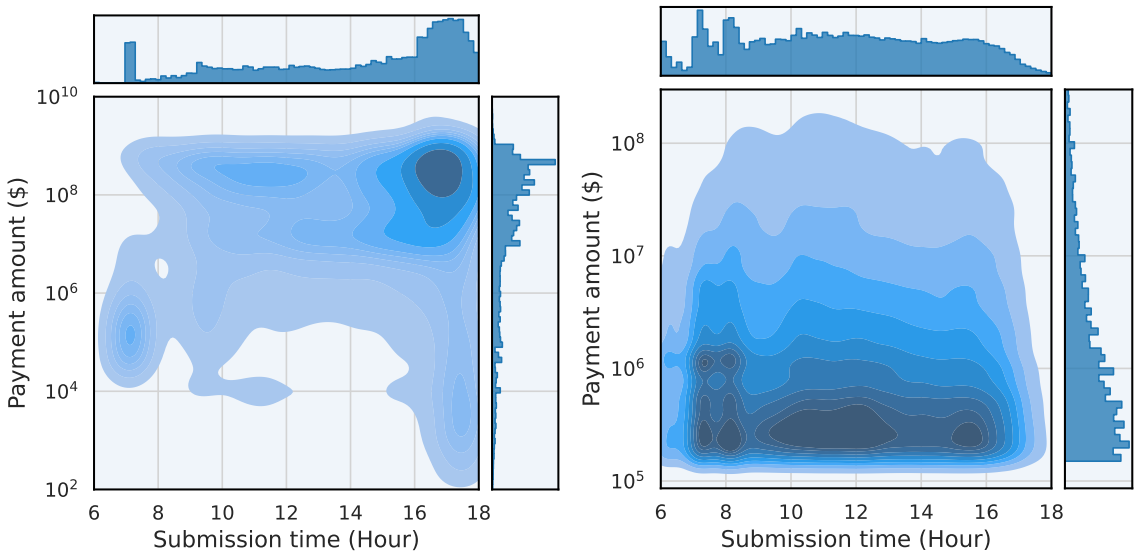


Figure 2: Distribution of daily value (in Can$) and volume (number of transactions) settled in LVTS between January 2012 and August 2021. Arrows highlight special days, such as Canadian provincial holidays, US national holidays, and days after a holiday.

Our sample contains various "special days." These are days when the system is open but only processing a small number of payments, such as Canadian provincial holidays and US national holidays. We call these "partial holidays." Also, there are days when the total value of payments is higher than usual, such as days after national holidays (which we call "post-closure" days) and at the end of the quarter. These special days explain the tails in Figure 2. Other special days observed in our sample are days with operational incidents. On these days, one or more LVTS participants could not send and/or receive payments for a certain amount of time because of technical difficulties. Also note that our sample covers the onset of the Covid-19 pandemic, during which participants changed their payment behavior because of policy changes (Chaudhry et al. 2021). As further discussed below, we use the data from these different types of "special days" to test our models' ability to identify changes in payment patterns and to isolate anomalous transactions.

LVTS settles transactions through two tranches: tranche 1 (LVTS-T1) and tranche 2 (LVTS-T2). The settlement mechanism of LVTS-T1 is similar to a traditional real-time gross settlement (RTGS) system in that transactions are settled using the senders' own collateral. By contrast, LVTS-T2 has components of an RTGS and a deferred net settlement (DNS) system that settles transactions using a joint collateral pool extended by the receiving participants (Arjani and McVanel 2006). Given the fundamental difference between these two LVTS tranches, we divide the dataset into LVTS-T1 and LVTS-T2 subsets. Figure 3 shows how the payment patterns differ between the two tranches. LVTS-T1 is primarily used for high-value payments, which are mainly submitted during the final opening hours of the system. By contrast, LVTS-T2 settles lower-value payments, the submissions of which are spread throughout the day. In our analyses, we

include all LVTS-T1 payments[14] and only those LVTS-T2 that are larger than 150,000 Can$. We do so for reasons of simplicity and to demonstrate the effectiveness of our framework in capturing especially high-value anomalous payments, as these will have the largest impact if left undetected.



(a) LVTS-T1 **(b)** LVTS-T2

Figure 3: Marginal and joint distribution of payment amount and submission time by LVTS tranche. Darker areas in the joint distribution indicate a higher density of payments. Note that the payment amount is based on a log scale.

4.2 Lynx Transaction Data

In August 2021, Lynx replaced LVTS and became Canada's new HVPS. Lynx is an RTGS and offers a new centralized queuing and liquidity savings mechanism (LSM). Also, Lynx's queuing policy is designed to encourage the use of the LSM. This, in turn, is expected to influence participants' payment patterns (Desai et al. 2023). We use Lynx data from October 2021 to August 2023 for our analysis. During this period, Lynx settled an average of 40,000 transactions per day, most of which were settled using the LSM. Note that the majority of this sample coincides with the Covid-19 pandemic. As we do with the LVTS-T2 data, we only focus on those Lynx transactions that are larger than 150,000 Can$.

4.3 Transaction Features

For each payment in our LVTS and Lynx datasets, we extract a set of features that we use in layer 1 for the *payments classifier* to classify payments and in layer 2 for isolating anomalies. These features include basis payment characteristics, as well as more complex indicators, such as the temporal relation between payments and the liquidity position of participants. We believe that these features effectively capture the intricate dynamics of payment patterns. See Table 1 for the comprehensive list of transaction features used in our analyses. These features can be categorized into the following subsets:

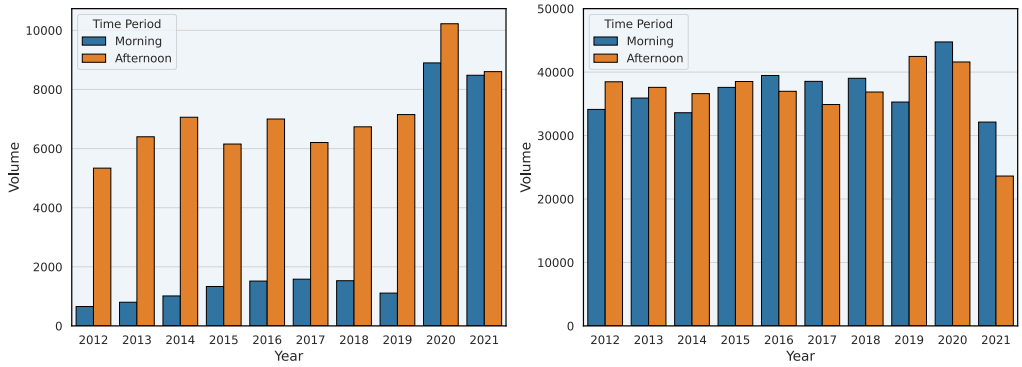
- Basic transaction features, such as sender, receiver, and amount.

[14] In the LVTS-T1 sample we do not include the payments sent by the Bank of Canada.

- Liquidity features, such as total collateral pledged by the sender, sender's credit limits, and system-level liquidity.
- Timestamp features, such as the month of the year, week of the year, and day of the month.
- Intraday features, such as the time elapsed since the last payment was sent to any receiver (or the same receiver), the time elapsed since the last payment was received (from any receiver or the same receiver), and the time elapsed since the start of the current period.

Certain transaction features, such as sender, receiver, and payment type, are categorical, whereas others, such as payment amount, collateral, and liquidity, are continuous of nature. Additionally, certain features, such as collateral, overnight rate, and specific timestamp features, do not change throughout the day. The mixed nature of these features introduces additional complexity during the training of the model, which may affect the performance of traditional models (Hastie et al. 2009). It is also for this reason that we use decision-tree-based models.

We use "period" as the target variable for the payments classifier. Payments submitted between 6 am and noon are classified as morning payments, and those submitted between noon and 6 pm are defined as afternoon payments. Figure 4 shows the yearly number of morning and afternoon payments by LVTS tranche. Before the start of the Covid-19 pandemic, LVTS-T1 payments were mostly submitted in the afternoon. However, during the pandemic, participants started to submit their payments both in the morning and afternoon (Figure 4, left). LVTS-T2 transactions have generally been submitted both in the morning and afternoon over the entire sample period, except for some changes in 2021 (Figure 4, right).



(a) LVTS-T1 **(b)** LVTS-T2

Figure 4: Number of payments submitted in the morning versus afternoon. The LVTS-T1 sample contains all LVTS-T1 payments. The LVTS-T2 sample includes all LVTS-T2 payments with a value greater than 150K Can$.

4.4 Training and Test Data

To test the ML models' performance in layer 1 and layer 2 of our framework, we divide our data samples into training and test sets. The training set is used to estimate the ML models and to tune their parameters. The "special days" are excluded from this set. By contrast, the test set is used for the out-of-sample evaluation of the models and therefore includes all transactions, those from both "normal" and "special" days. By doing so, the payments classifier is trained to classify the training data as accurately as possible and then tested out of sample on a wider range of transactions.

Table 1: List of features extracted for each of the transactions settled in LVTS.

Transaction feature	**Description**
Binary target variable: Period	Submission time: 6am–noon (morning) or noon–6pm (afternoon)
Sender	ID of LVTS participant sending the payment
Receiver	ID of LVTS participant receiving the payment
Amount	Amount of the transaction (in Can$)
Rounding	Rounded (zero cents) vs non-rounded payments
Payment type	Type 1 for interbank and type 2 for client-driven payments
Overnight loan	Whether the payment is for an overnight loan
Loan repayment	Whether the transaction is a repayment of an overnight loan
Settlement balance	Size of overnight settlement balance of sender (in Can$)
Collateral	The collateral pledged by the sender
Liquidity	System-wide available liquidity (for all LVTS participants)
BCL	Sender's bilateral credit limit for LVTS-T2 payments
MCL	Sender's multilateral credit limit for LVTS-T1 and LVTS-T2 payments
Rate	Overnight money market rate on previous day
Settlement time: Month	Month in which the payment was settled
Settlement time: Month-day	Day on which the payment was settled
Settlement time: Week	Week of the year the payment was settled
Settlement time: Week-day	The specified day of the week
Settlement time: Hour-time	Seconds elapsed since the start of the latest hour
Settlement time: Period-time	Seconds elapsed since the start of the period (morning or afternoon)
Multilateral-time-sender	Seconds elapsed since the last payment sent
Multilateral-time-receiver	Seconds elapsed since the last payment received
Multilateral-type-time-sender	Seconds elapsed since last same-type payment sent
Multilateral-type-time-receiver	Seconds elapsed since last same-type payment received
Bilateral-time-sender	Seconds since last payment sent from same sender to same receiver
Bilateral-time-receiver	Seconds since last payment received by same sender from same receiver
Bilateral-type-time-sender	Seconds since last same-type payment from sender to receiver
Bilateral-type-time-receiver	Seconds since last same-type payment received from sender to receiver

As a second testing procedure, we also run the models on a sample of artificial anomalies. These anomalies are generated by manually altering certain transaction features, such as increasing the actual payment amount or changing the original submission time from morning to afternoon. These intentionally designed artificial transactions exhibit significant deviations from typical transaction behavior, allowing us to assess the model's ability to detect such outliers. For further details on the artificial transactions sample, see Appendix C.

5 Results of Proposed Framework Using Canadian LVTS-T1 Data

5.1 Layer 1 - Payments Classification

We estimated the *payments classifier* with LVTS-T1 settlement data from January 2012 to August 2021, using the model outlined in section 3.1 and the training procedure described in section A.1. Subsequently, we tested its out-of-sample performance with artificial anomalies on a sample containing both normal and special days (as described in section 4.4).

Table 2: Out-of-sample model performance using different models*

Model	Accuracy (in %) on clean data[a]	Detection rate (in %) on artificial data[b]
Logistic Regression	67.8	66.4
Decision Tree	92.3	82.6
Gradient Boosting	94.6	76.8
Random Forest	96.7	94.4
Light Gradient Boosting	97.6	92.2

* All models are trained using LVTS-T1 settlement data from January 2012 to August 2021 on clean (normal) days.

[a] Out-of-sample accuracy of models when classifying transactions into morning and afternoon payments on normal or clean days only. Note that the accuracy provides a general measure of how well a binary classification model is performing. It is calculated as the ratio of correctly predicted transactions (from both the morning and afternoon periods) to the total number of transactions in the sample.

[b] Ability to detect artificially manipulated transactions out of sample, calculated as the fraction of manually altered transactions correctly identified by the classifier.

The results demonstrate that the gradient-boosting-based LightGBM model outperforms the logistic regression model by up to 44% when classifying payments into morning and afternoon payments (Table 2). The LightGBM also performs better than other ML models, such as the decision tree, random forest, and gradient boosting. Specifically, it achieves 97.6% accuracy during the out-of-sample testing on the clean sample, with only 2.4% of transactions being misclassified (see column one). Moreover, when testing the model with artificially manipulated anomalous transactions (column two), LightGBM correctly detects 92.2% of payments out of sample, which is slightly lower than the random forest model, which achieved 94.4% accuracy, but better than other models.

The misclassification rate of LightGBM increases when using the datasets that include special days (Figure 5, left). Approximately 10% of payments are misclassified when tested on the dataset that includes partial holidays. This percentage increases to 46% when testing it on the dataset covering the Covid-19

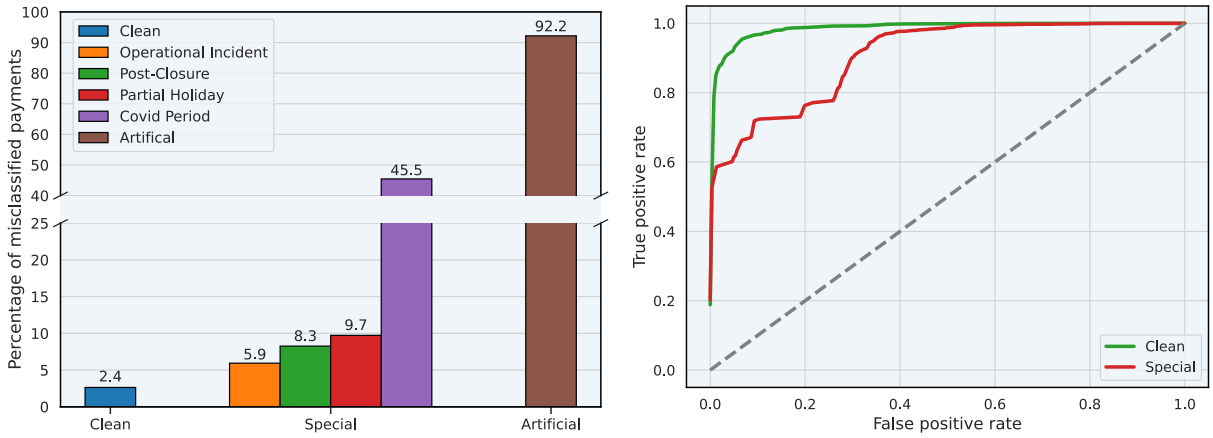


Figure 5: Left: Out-of-sample performance of payments classifier using LightGBM, as measured by the fraction of misclassified payments on the clean (normal) days, special days, and a sample of artificial transactions. Right: Receiver operating characteristic (ROC) curve showing the LightGBM's performance at various threshold values between 0 and 1.

period. This demonstrates that LVTS participants exhibited different payment patterns on these days and during this period, and proves that the model is capable of learning participants' "usual" payment behavior and detecting deviations in it.

An alternative way to assess the model's performance is with the receiver operating characteristic (ROC) curve (Figure 5, right). The fact that the area underneath the green curve is larger than that underneath the red curve reveals that the model performs significantly better when using the clean (normal) data instead of the data with special days. This again demonstrates the model's ability to detect deviations from LVTS participants' typical payment behavior.

5.2 Layer 2 - Anomaly Detection

As proposed in Section 3.2, we subsequently use the set of misclassified transactions to train the isolation forest (IF) algorithm to detect anomalies. We assess the model using artificially manipulated transactions and their corresponding original counterparts (i.e., the real transactions).

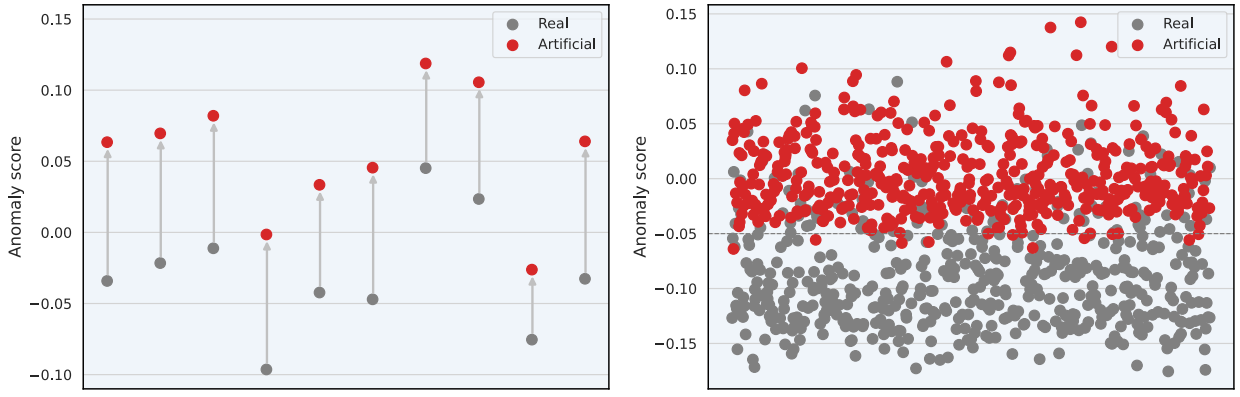


Figure 6: Anomaly detection using the IF model on LVTS-T1 data: Anomaly scores when using the set of transactions before (real, shown in gray) and after (artificial, shown in red) manipulation. A transaction with a high (positive) score is likely more anomalous than a transaction with a low (negative) score.

The anomaly scores generated by the IF algorithm are useful in evaluating the model's performance: transactions with higher and positive anomaly scores indicate greater anomalous behavior compared to those with lower and negative scores. The results presented in Figure 6 show that the IF algorithm successfully assigns higher anomaly scores to the manually generated anomalous transactions than to the real transactions. On average, the model assigns approximately a twice-as-high anomaly score to manually altered transactions than to the original data (see right panel). As such, the IF algorithm proves to be a successful tool to detect unusual transactions and to prioritize further investigation of the most likely suspicious transactions.[15]

5.3 Interpretation of Layer 1 Model Predictions - Payments Classifier

5.3.1 Role of payment features in participants' usual payment behavior

The training results of the payments classifier also allow us to gain further insights into how specific payment features influence the submission time of HVPS payments. The average SHAP values across the training sample for each feature as depicted in Figure 7 show that the time elapsed since the previous payment received and the time elapsed since the start of the period contribute most towards the model's capability to correctly predict payments. Other important predictors are the amount of the payment, the amount of collateral pledged by the payer, and recipient-specific features. The model's reliance on the time elapsed since the previous payment received suggests that incoming payments play an important role in participants' behavior. This aligns with earlier research that demonstrates the presence of substantial payment coordination between HVPS participants (see e.g., McAndrews and Rajan 2000; Becher et al. 2008; Alexandrova-Kabadjova et al. 2023).

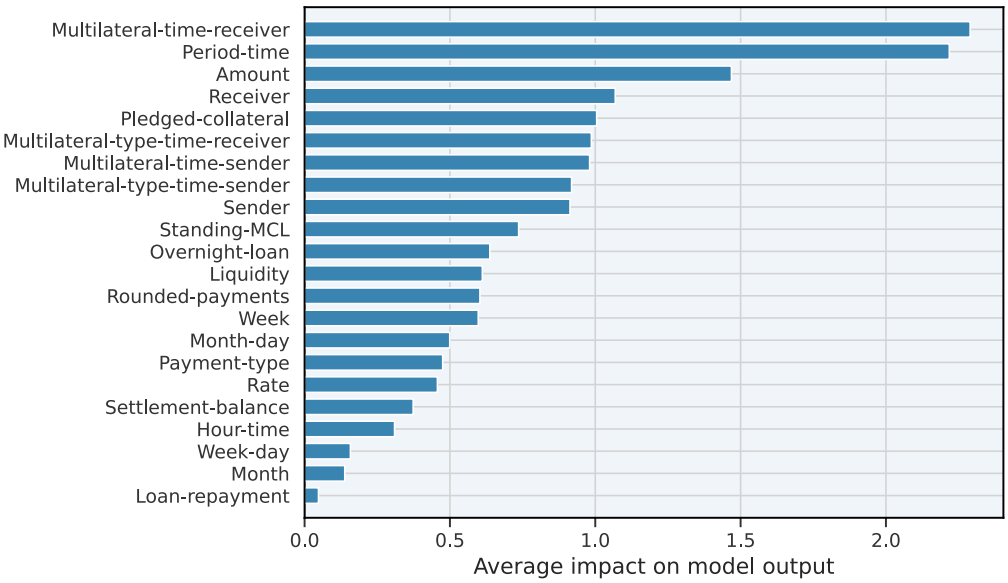


Figure 7: LVTS-T1: SHAP global feature importance, measured by the mean absolute SHAP values for each transaction in the training sample. The features are ranked from high (top) to low (bottom) based on the average impact on the model outcomes.

Figure 8 illustrates how the predictions of the payments classifier are influenced by the amount and type of the payment. Customer-driven payments (type 2, shown in red) are more likely to be classified as morning

[15] Note that the accuracy is reduced when relying solely on the unsupervised IF model without using the preceding supervised classifier model in the first layer . See Appendix B.2 for further details.

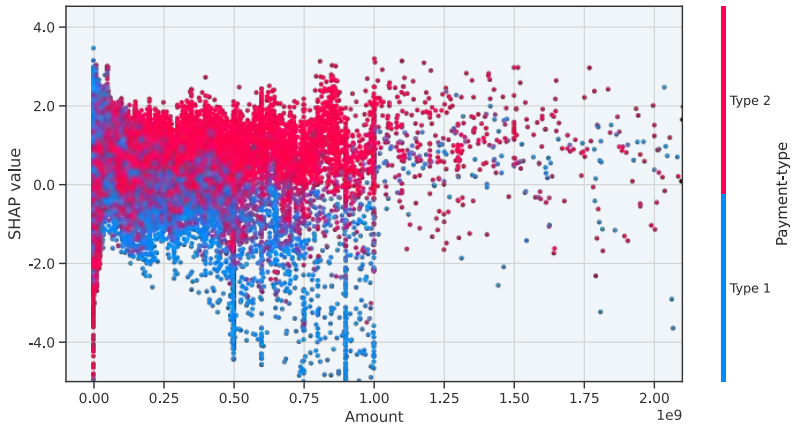


Figure 8: Dependence plot showing the SHAP values for payment amount (in Can$) by payment type for each transaction. Type 1 are inter-bank payments (shown in blue) and type 2 are client-driven payments (shown in red).

payments (negative SHAP value) when small, whereas they are predicted as afternoon payments when large (positive SHAP value). By contrast, interbank payments (type 1, shown in blue) are more commonly submitted in the afternoon when small and sent in the afternoon when large. Interestingly, for payments larger than 1 billion Can$, the payment type no longer strongly contributes to the model's prediction. Most high-value payments of either type tend to be submitted in the afternoon, with the exception of a few large interbank payments being submitted in the morning.

5.3.2 Role of Payment Features in the Prediction of Individual Payments

Figure 9 shows the SHAP force plots for one particular transaction, with $(f(x))$ being the prediction score. A negative score indicates the morning period, while a positive score indicates the afternoon. The upper plot is based on the actual features of the transaction, whereas the bottom plot was generated after manipulating its period-time feature (from 7750 to 100 seconds). The actual submission period (morning) is correctly predicted by the model, as indicated by the negative (-4.25) prediction score. The bottom panel, however, shows that the model successfully misclassifies and predict the manipulated transaction as an afternoon payment (with a higher and positive score of 3.02). Moreover, the large red "period-time" arrow pushes the model towards an afternoon prediction, which indicates that it is this period-time feature that makes the model predict it as an afternoon payment.

Similarly, Figure 10 shows the force plots comparing an actual transaction (upper panel) with one where we manually lowered its amount (from 11.8 thousand to 5.2 billion) (bottom panel). The model correctly classifies the actual transaction as a morning transaction (with a negative score of -10.34) and misclassifies the manipulated one as an afternoon transaction (with a positive score of 0.21). The payment amount is indeed the most prominent feature driving this misclassification, as illustrated with the relatively large right-pushing arrow.

These examples indicate that the payments classifier model has successfully learned the typical payment behavior of LVTS-T1 participants and that it is able to identify transactions that deviate from this typical behavior. Moreover, the SHAP values provide a useful tool to better understand why the model is predicting a particular payment as such. This not only provides insight into the drivers behind LVTS-T1 participants' true behavior, but also serves as a robustness check of the model's specification.

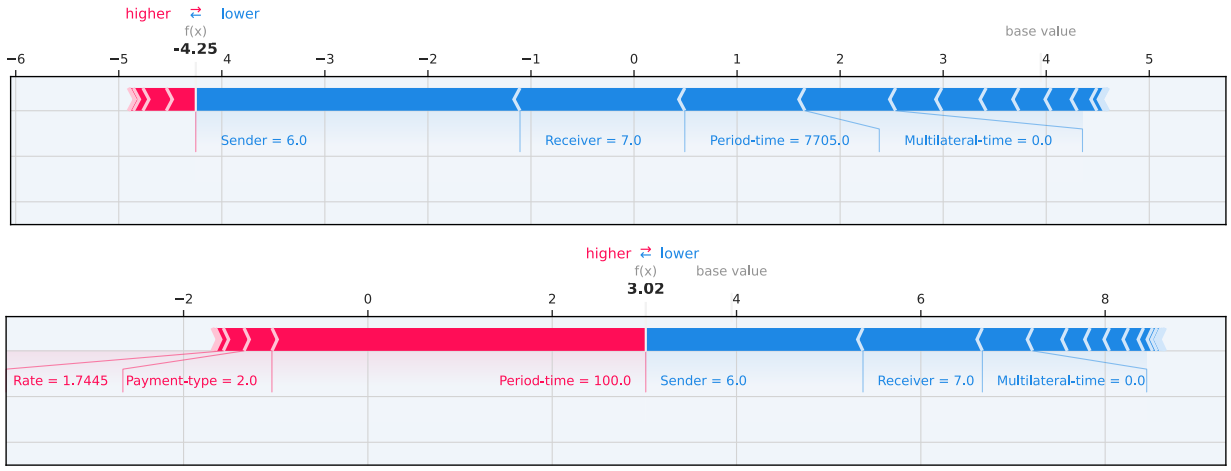


Figure 9: LVTS-T1: SHAP force plots showing the marginal contribution of each predictor for a chosen transaction. The red arrows are positive SHAP values (pushing the prediction towards an "afternoon" prediction), and the blue arrows are negative SHAP values (pushing the prediction towards the morning period). $f(x)$ is the prediction score generated by the model for that transaction, and the base value is the average of all predictions over the entire sample. The feature values shown in red and blue are the predictor values for that transaction. The original transaction was settled in the morning and correctly classified as such by the model (see negative prediction score in the upper panel). After manipulating the period-time feature, the transaction is misclassified as an afternoon payment (see positive prediction score in the bottom panel), which is mainly driven by the amended period-time.

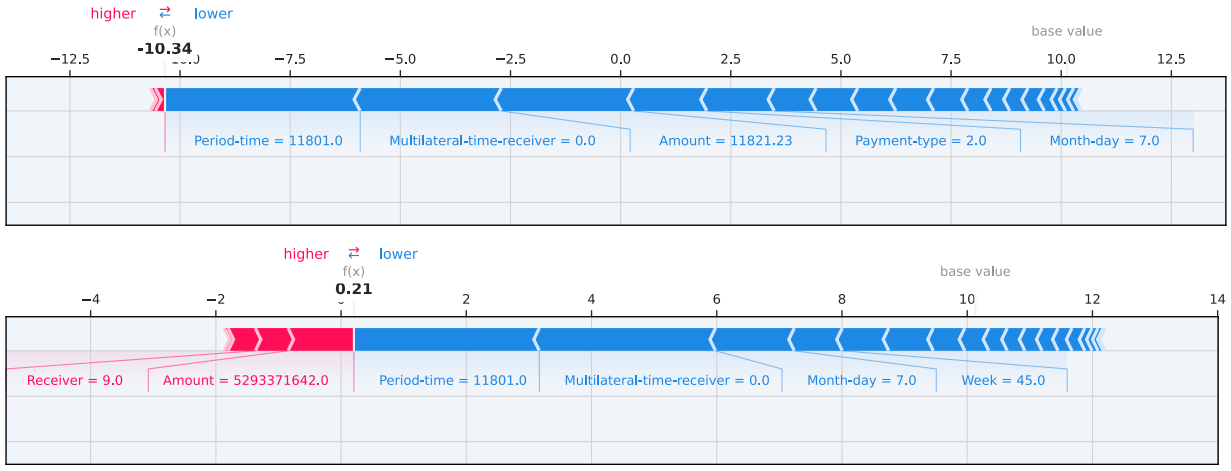


Figure 10: LVTS-T1: SHAP force plots showing the feature contribution for a chosen transaction. The original transaction was settled in the morning and correctly classified as such (see negative prediction score in upper panel). After manipulating the payment value, the model classifies the transaction as an afternoon payment (see positive prediction score at the bottom). The payment value is the dominant feature driving the misclassification.

5.4 Interpretation of Layer-2 Model Predictions - Anomaly Detection

From Figure 11 it is evident that the intraday time features contribute the most to the identification of anomalous payments. This is slightly different from the features contributing to the outcomes of the payments classifier in the first layer (see Figure 7). This suggests that the more complex intraday time features, such as the time passed since the last incoming payment of the same kind, the time elapsed since the last outgoing payment, and the time since since the last incoming payment of any kind, are crucial for identifying and isolating anomalies from the subset of misclassified transactions. Other features that the IF model heavily draws on when identifying anomalies include the payment value, the credit position of the sender, and the liquidity available in the system.

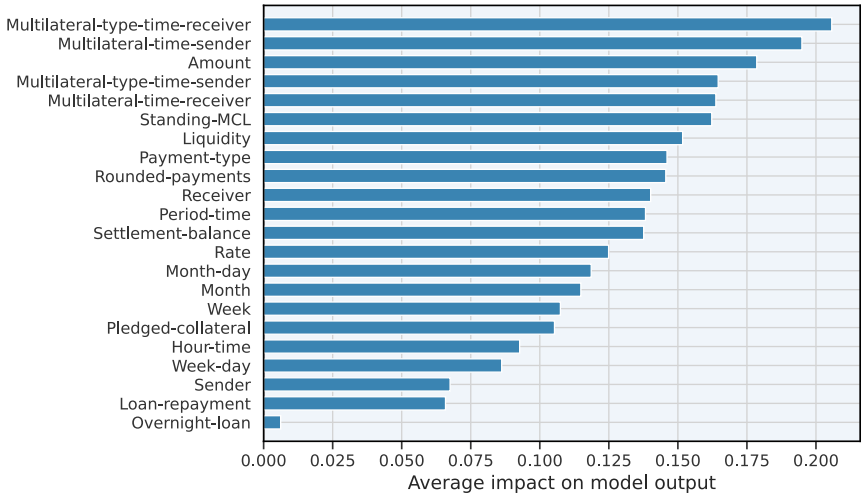


Figure 11: Anomaly detection model using LVTS-T1 data: SHAP global feature importance measured as mean absolute SHAP values for each transaction in the misclassified subset of transactions. The features are ranked from high (top) to low (bottom), based on the average impact on the model's output.

6 Results of Alternative Applications

6.1 Model Training and Evaluation Using Joint Payment Classifiers

To demonstrate the flexibility of our two-layered framework, we introduced a second payments classifier module that operates parallel to the original classifier. This effectively creates a joint classifier. The second classifier is trained using the payment type (interbank vs client-driven) as the target variable,[16] while the first classifier still focuses on the submission period (morning vs afternoon). As such, the joint classifier classifies payments based on both the payment timing and the type of payment. It can be formulated as follows: let $\mathbf{y_1}$ be the first target variable (submission time) and $\mathbf{y_2}$ the second target variable (payment type). We can denote $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ as the predicted targets, which can be estimated using the classification models $f_1 : \hat{\mathbf{y}}_1 = f_1(X)$ (classifier 1) and $f_2 : \hat{\mathbf{y}}_2 = f_2(X)$ (classifier 2).

We train and test the joint-classifier model on the LVTS-T1 dataset using the same procedures as above and as visualised in Figure 12: all transactions that are misclassified based on either one of the target variables are automatically sent to the anomaly detection module. Subsequently, only those correctly classified based

[16] See Appendix A.2 for more details on the second payments classifier type.

on both dimensions are used for pattern recognition. The results (see Appendix A.2) demonstrate how using classifiers that capture multiple payment features at the same time can enhance the model's accuracy and robustness. In our tests, the joint classifier was able to correctly catch 96% of the manipulated transactions, which is higher than using either the original (92%) or the alternative classifier (66%) independently. The use of multiple payments classifiers can also improve the performance of the anomaly detection model in the second layer. On average, the IF model assigns a 13% higher anomaly score to manually altered transactions when used in conjunction with the joint-classifier model, compared to the standalone classifier-1 model.

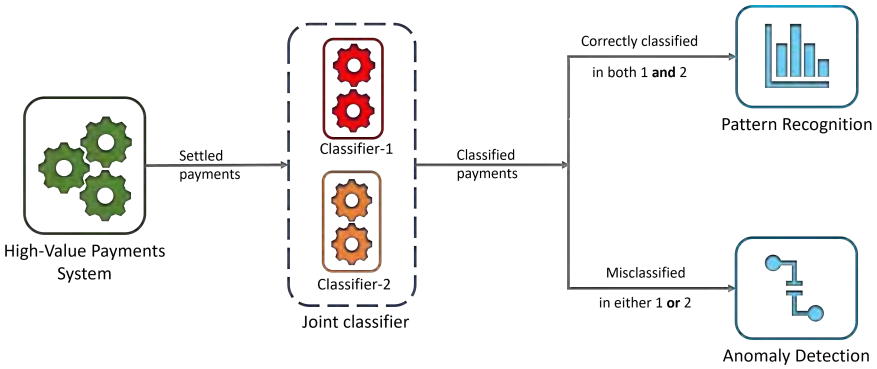


Figure 12: Schematic overview of alternative layered approach for pattern recognition and anomaly detection using a joint classifier. The payments classifier 1 is trained with the submission time (morning vs afternoon) as the target variable, while classifier 2 is trained with the payment type (interbank vs client-driven) being the target variable.

6.2 Model Training and Evaluation Using LVTS-T2 Data

To assess the extent to which the proposed single-classifier model can be applied to payment systems or settlement mechanisms other than LVTS-T1, we trained and tested it on the LVTS-T2 dataset, using the data and procedure outlined in sections 4.1 and A.1. Overall, the model is able to detect 96.4% of the artificially manipulated LVTS-T2 transactions (Figure 13, left panel). These results indicate that the binary-payments-classifier model is able to handle artificial transactions pretty well when applied to another type of settlement mechanism. However, its ability to capture more subtle deviations in payment patterns, such as those on special days, is reduced when testing it beyond the LVTS-T1 data: the misclassification rate using the clean sample (14%) is only slightly better than when using the special days samples (15% to 19%). In particular, the number of misclassified payments using the clean sample is higher when using the LVTS-T2 data (14%) than when using the LVTS-T1 set (2.4%). This could be attributed to the sheer volume and variety of payments processed through LVTS-T2. So although our proposed classifier can be applied to payment systems beyond that of LVTS-T1, the set of transaction features used for the classification would need to be tailored to generate the same level of performance.

The time elapsed since the start of the period is the most important transaction feature that the model uses to classify payments, followed by the time elapsed since the last incoming payment from the same counterpart. This "bilateral" transaction feature is more influential than the "multilateral" transaction features that were prominent when using the LVTS-T1 sample. This highlights that bilateral relationships are important in LVTS-T2, with participants attaching higher value to coordinating payments on a bilateral than on a multilateral basis (as in LVTS-T1). This is in line with what one would expect, as the risk management in LVTS-T2 is heavily based on bilateral credit limits (BCL) (see Arjani and McVanel (2006)).

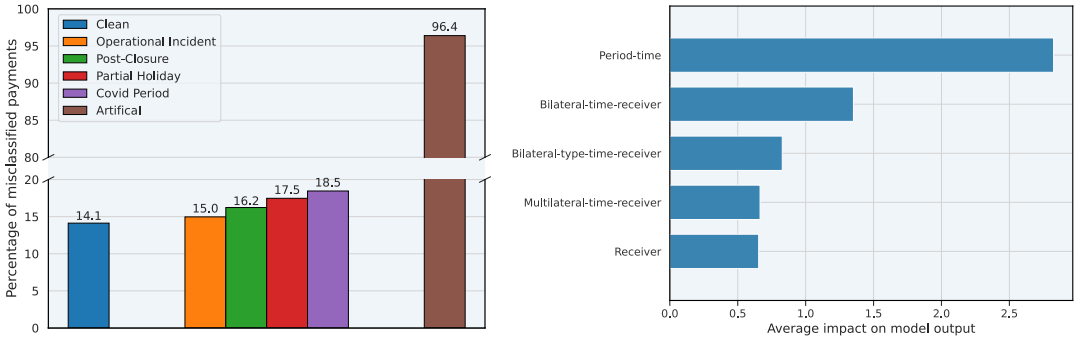


Figure 13: Results using LVTS-T2 data. Left panel: Out-of-sample performance of payments classifier using LightGBM, as measured by the fraction of misclassified payments on the clean (normal) days, special days, and a sample of artificial transactions. Right: Key features contributing to the LightGBM model's predictions based on mean absolute SHAP values.

6.3 Model Training and Evaluation on Lynx

We also trained and tested the single-classifier model on the Lynx dataset. When interpreting the results and comparing these with the results of the LVTS datasets, it is important to keep in mind that (i) Lynx is an RTGS system with an LSM, so of a different design than both LVTS-T1 and LVTS-T2, (ii) the size of the sample is much smaller and only ranges from October 2021 to August 2023, and (iii) the Lynx data contains all HVPS transactions in Canada, i.e., all transactions that used to be sent through either LVTS-T1 or LVTS-T2 (Desai et al. 2023).

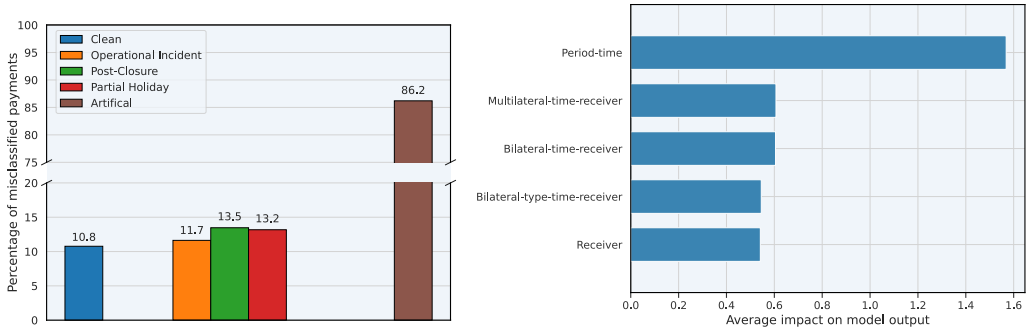


Figure 14: Results using Lynx data. Left panel: Out-of-sample performance of the payments classifier using LightGBM, as measured by the fraction of misclassified payments on the clean (normal) days, special days, and a sample of artificial transactions. Right: Key features contributing to the LightGBM model's predictions based on mean absolute SHAP values.

Overall, our LightGBM model is able to detect 86.2% of the artificially manipulated transactions when applying it to the Lynx data. This is pretty good, especially given the newness of the system and the small size of the dataset, which make it more challenging to effectively capture usual payments behavior. This may also explain why the payments misclassification rate is only slightly lower when using the normal sample than when including the special days (Figure 14, left).

The right panel of Figure 14 shows the key features that the Lynx payments classifier relies on during its training. Like the LVTS-T2 sample, "period-time" is the most prominent feature. However, in Lynx, not only the "bilateral" timing features but also the "multilateral" timing features are key predictors that the payments classifier relies on when classifying payments into morning and afternoon payments. This is in line with the RTGS nature of Lynx (Desai et al. 2023).

7 Conclusions and Practical Implementation

In this paper we propose a novel ML framework for real-time transaction monitoring in HVPSs. We use advanced data-driven ML algorithms to learn the "typical" payment patterns of HVPS participants and to isolate anomalous transactions when deviations are spotted. The key strength of our framework is that it is able to address the challenging task of identifying anomalies in large and high dimensional sets of payments data, particularly when anomalies are not known a priori. The framework derives its strength from its layered approach. In the first layer, we train a supervised ML-based payments classifier to efficiently screen typical transactions. This payments classifier then streamlines the subsequent anomaly detection task in the second layer, for which we use an unsupervised ML-based IF model. We show that our framework can be applied for different types of HVPSs. Also, we demonstrate that it can be extended with additional algorithms, such as with joint classifiers, to further enhance its robustness.

Due to its flexibility, the ML framework proposed in this paper can be used for different applications in finance. Payment system operators and overseers may implement it to detect cyber attacks or operational outages, which, if left undetected, could have serious implications for the HVPS, its participants, and the financial system more broadly. Moreover, the framework could be used to detect early signs of financial stress at individual institutions or the unusual behavior of their clients. While trained on HVPS data, our framework could also be applied to other types of payment systems that generate large and highly dimensional datasets. For example, banks and other payment service providers could use it for anti-money laundering transaction screening to identify suspicious behavior and financial crime.

From a practical standpoint, the initial training of the ML models in our framework may demand a substantial time investment and computational resources. The models may also require regular retraining and updates to ensure that they remain attuned to evolving payment patterns. That said, once in place and up to date, they allow for real-time (fast) and computationally efficient (utilizing fewer resources) monitoring and predictions. Moreover, the Shapley value-based (SHAP) tools can assist operators in their regular evaluation of the framework and provide them with an initial indication of where flagged anomalies may be coming from or what they are hinting at, thus aiding a timely identification and resolution.

Despite the numerous advantages of our ML framework for HVPS transaction monitoring, there is room for improvement, which can be explored in future research. In particular, further work may consider some of the challenges that have surfaced in our work. First, the number of transaction features automatically increases with the number of HVPS participants and payment types, which increases the complexity of model training and the predictions. This scalability challenge could be addressed by applying feature selection techniques or by exploring alternative approaches that handle large sets of features. Second, the classification of a large number of anomalous transactions could prevent payment system operators from reviewing them and taking action in a timely manner. This may be addressed by improving the accuracy of the model through feature engineering and additional training. Finally, it could be challenging to accurately classify edge cases along the classification boundary, such as, in our case, payments close to noon. This could be addressed by using a regression model in the first layer of the framework.

References

Alexandrova-Kabadjova, B., A. Badev, S. B. Bastos, E. Benos, F. Cepeda-Lopez, R. Garratt, R. Heijmans, A. Kosse, A. Martin, T. Nellen, T. Nilsson, J. Paulick, A. Pustelnikov, F. Rivadeneyra, M. R. do Coutto Bastos, and S. Testi (2023). Intraday liquidity around the world. *BIS Working Paper* (No 1089). doi: www.bis.org/publ/work1089.htm.

Alexandrova-Kabadjova, B., A. Serguieva, R. Heijmans, and L. Garcia-Ochoa (2015). Direct participants' behavior through the lens of transactional analysis: The case of SPEI. In *Proceedings of the International Conference on Social Modeling and Simulation*, pp. 205–215.

Arjani, N. and R. Heijmans (2020). Is there anybody out there? Detecting operational outages from large value transfer system transaction data. *Journal of Financial Market Infrastructures 8*(4). doi: 10.21314/JFMI.2019.118.

Arjani, N. and D. McVanel (2006). A primer on Canada's large value transfer system. Technical report, Bank of Canada. https://www.bankofcanada.ca/wp-content/uploads/2010/05/lvts_neville.pdf.

Arjani, N., L. Sabetti, and F. Li (2020). Monitoring intraday liquidity risks in a real-time gross settlement system. *Journal of Financial Market Infrastructures 9*(3). doi: 10.21314/JFMI.2021.012.

Arévalo, F., P. Barucca, I.-E. Téllez-León, W. Rodríguez, G. Gage, and R. Morales (2022). Identifying clusters of anomalous payments in the Salvadorian payment system. *Latin American Journal of Central Banking 3*(1). https://www.sciencedirect.com/science/article/pii/S2666143822000059.

Bech, M. L. and R. Garratt (2003, April). The intraday liquidity management game. *Journal of Economic Theory 109*(2), 198–219.

Bech, M. L. and R. J. Garratt (2012). Illiquidity in the interbank payment system following wide-scale disruptions. *Journal of Money, Credit and Banking 44*(5), 903–929.

Becher, C., M. Galbiati, and M. Tudela (2008). The timing and funding of CHAPS Sterling payments. *FRBNY Economic Policy Review 14*(2), 113–133.

BIS-Report (2019). Reducing the risk of wholesale payments fraud related to endpoint security: A toolkit. Technical report, Bank for International Settlements. https://www.bis.org/cpmi/publ/d188.pdf.

Bukth, T. and S. S. Huda (2017). *The soft threat: The story of the Bangladesh bank reserve heist*. SAGE Publications: SAGE Business Cases Originals.

Castro, P. S., A. Desai, H. Du, R. Garratt, and F. Rivadeneyra (2021). Estimating policy functions in payments systems using reinforcement learning. Technical report, Bank of Canada Working Paper 2021-7.

CEIP (2021). Timeline of cyber incidents involving financial institutions. Technical report, Carnegie Endowment for International Peace. https://carnegieendowment.org/specialprojects/protectingfinancialstability/timeline#MexicanBankTheft2018.

Chapman, J., J. Chiu, S. Jafri, and H. Pérez Saiz (2015). Public policy objectives and the next generation of CPA systems: An analytical framework. Technical report, Bank of Canada.

Chaudhry, A., A. Kosse, and K. Sondergard (2021). Behaviour in the Canadian large-value payment system: Covid-19 vs. the global financial crisis. Technical report, Bank of Canada.

Chen, T., T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, et al. (2015). Xgboost: Extreme gradient boosting. *R package version 0.4-2 1*(4), 1–4.

Desai, A., Z. Lu, H. Rodrigo, J. Sharples, P. Tian, and N. Zhang (2023). From LVTS to Lynx: Quantitative assessment of payment system transition in Canada. *Journal of Payments Strategy & Systems 17*(3), 291–314.

Docherty, P. and G. Wang (2010). Using synthetic data to evaluate the impact of RTGS on systemic risk in the Australian payments system. *Journal of Financial Stability 6*(2), 103–117.

Eisenbach, T. M., A. Kovner, and M. J. Lee (2021). Cyber risk and the US financial system: A pre-mortem analysis. *Journal of Financial Economics 145*(3), 802–826.

FED-Report (2019). Federal Reserve fraudclassifier model. Technical report, Federal Reserve. https://fedpaymentsimprovement.org/.

Flood, M. D., V. L. Lemieux, M. Varga, and B. W. Wong (2016). The application of visual analytics to financial stability monitoring. *Journal of Financial Stability 27*, 180–197.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics. New York: Springer. doi: 10.1007/978-0-387-84858-7.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics 29*(5), 1189–1232. doi: 10.1214/aos/1013203451.

Glowka, M. (2019). Profiling banks: How to use cluster analysis with payment system data. *Journal of Financial Market Infrastructures 8*(2), 21–45. doi: 10.21314/JFMI.2019.118.

Hastie, T., R. Tibshirani, and J. H. Friedman (2009). *The elements of statistical learning: Data mining, inference, and prediction*, Volume 2. New York, Springer.

James, R., H. Leung, and A. Prokhorov (2023). A machine learning attack on illegal trading. *Journal of Banking & Finance 148*, 106735.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, pp. 3146–3154. doi: https://lightgbm.readthedocs.io/en/latest/.

Kosse, A. and Z. Lu (2022). Transmission of cyber risk through Canadian wholesale payments system. Technical report, Bank of Canada Working Paper. https://doi.org/10.34989/swp-2022-23.

Kotidis, A. and S. Schreft (2023). The propagation of cyberattacks through the financial system: Evidence from an actual event. Technical report, Federal Reserve System Working Paper 2022-25.

León, C. (2020). Detecting anomalous payments networks: A dimensionality-reduction approach. *Latin American Journal of Central Banking 1*(1-4), 100001.

León, C., P. Barucca, O. Acero, G. Gage, and F. Ortega (2020). Pattern recognition of financial institutions' payment behavior. *Latin American Journal of Central Banking 1*(1-4), 100011.

LightGBM-Documentation (2022). LightGBM's official documentation. https://lightgbm.readthedocs.io/en/latest/.

Liu, F. T., K. M. Ting, and Z.-H. Zhou (2008). Isolation forest. In *2008 eighth IEEE international conference on data mining*, pp. 413–422. IEEE.

Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence 2*(1), 2522–5839. doi: 10.1038/s42256-019-0138-9.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*, pp. 4765–4774. Curran Associates, Inc.

Martin, A. and J. McAndrews (2008). Liquidity-saving mechanisms. *Journal of Monetary Economics 55*(3), 554–567.

Massarenti, M., S. Petriconi, and J. Lindner (2012). Intraday patterns and timing of TARGET2 interbank payments. *Journal of Financial Market Infrastructures 1*(2), 3–24.

McAndrews, J. and S. Rajan (2000). The timing and funding of Fedwire funds transfers. *Economic Policy Review 6*(2).

McMahon, C., D. McGillivray, A. Desai, F. Rivadeneyra, J.-P. Lam, T. Lo, D. Marsden, and V. Skavysh (2022). Improving the efficiency of payments systems using quantum computing. *arXiv preprint arXiv:2209.15392*.

Molnar, C. (2020). *Interpretable machine learning*. Lulu.com.

Ngai, E. W., Y. Hu, Y. H. Wong, Y. Chen, and X. Sun (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems 50*(3), 559–569.

Nish, A. and S. Naumaan (2019). *The cyber threat landscape: Confronting challenges to the financial system*. Carnegie Endowment for International Peace. https://carnegieendowment.org/files/02_19_Nish_Naumaan_Fin_Threats_final.pdf.

Osborne, M. J. and A. Rubinstein (1994). *A course in game theory*. MIT Press.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12*, 2825–2830.

Rubio, J., P. Barucca, G. Gage, J. Arroyo, and R. Morales-Resendiz (2020). Classifying payment patterns with artificial neural networks: An autoencoder approach. *Latin American Journal of Central Banking 1*(1-4), 100013.

Ryman-Tubb, N. F., P. Krause, and W. Garn (2018). How artificial intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence 76*, 130–157.

Sabetti, L. and R. Heijmans (2021). Shallow or deep? Training an autoencoder to detect anomalous flows in a retail payment system. *Latin American Journal of Central Banking 2*(2), 100031.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games 2*(28), 307–317.

Slack, D., S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 180–186.

TARGET-Report (2019). TARGET annual report. Technical report, European Central Bank. https://www.ecb.europa.eu/pub/pdf/targetar/ecb.targetar2019.en.pdf.

Triepels, R., H. Daniels, and R. Heijmans (2017). Anomaly detection in real-time gross settlement systems. In *Proceedings of the 19th International Conference on Enterprise Information Systems - Volume 3: ICEIS*, pp. 433–441. INSTICC: SciTePress.

Zhang, N. (2015). Changes in payment timing in Canada's large value transfer system. Technical report, Bank of Canada Working Paper. https://www.bankofcanada.ca/2015/06/working-paper-2015-20/.

A Gradient Boosting Machines

Gradient boosting (GB) is a decision tree (DT)-based ensemble learning approach. It is a general technique of boosting in which a sequence of weak learners (e.g., DTs) are built on a repeatedly modified version of the training set. The data modification at each boosting interaction consists of applying weights to each of the training samples, and for successive iterations, the sample weights are modified. Basically, the next learner is fit on the residual of the previous learner (Friedman 2001; Friedman et al. 2001).

Let $X = \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^M$ represent M features, each with N training samples. The target variable y_i denotes the true label for the i-th instance in binary classification (class 1 or 2). Similarly, let p_i represent the predicted probability of class 1 for the i-th instance. Then the objective function L can be defined as

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right].$$

To perform gradient boosting, we need to compute the gradient of the objective function with respect to the predicted probabilities p_i, which is given by

$$\frac{\partial L}{\partial p_i} = -\frac{1}{N}\left(\frac{y_i}{p_i} - \frac{(1-y_i)}{(1-p_i)}\right).$$

To train this model, during each iteration of the gradient boosting process, a weak learner (a DT) is trained using a subset of the dataset at the start to minimize the gradient of the objective function. By computing the difference between predicted values and the actual targets, pseudo-residuals are derived. Subsequently, another weak learner is trained on these pseudo-residuals from the previous step, aiming to minimize prediction errors. This iterative process continues, with each new learner focusing on the combined model's errors from preceding iterations. The predicted probabilities generated by each new weak learner are incrementally added to the predictions from prior iterations using a specified step size (learning rate) to update the model. This iterative procedure persists, often running for a predefined number of boosting rounds or until a designated stopping criterion is reached. Note that the final prediction for binary classification is often obtained by applying a threshold (e.g., 0.5) to the predicted probabilities. If p_i is greater than the threshold, the example is classified as class 1; otherwise, it is classified as class 2.[17]

Tree-based GB is a popular ML algorithm. However, it is computationally expensive in terms of efficiency and scalability when the feature dimension M is high and the sample size N is large. To tackle this problem, Ke et al. (2017) propose an improved and efficient implementation of GB called LightGBM, which (i) uses a histogram-based approach for tree construction to reduce memory usage and speed up training, (ii) employs gradient-based optimization to excludes a significant proportion of data instances with small gradients and speeds up the training process, and (iii) uses an exclusive predictors bundling technique to bundle mutually exclusive features to reduce the number of features. Using these techniques in the LighGBM model, they achieve a 20 times speedup with almost the same accuracy on multiple public datasets compared to the GB model. Moreover, in many cases LighGBM has been shown to perform similar to or, in some cases better than, other boosting algorithms such as XGBoost (Chen et al. 2015). In this paper we use the open-source LightGBM Python-package available at LightGBM-Documentation (2022).

[17] Note that our procedure involves two simplifying assumptions. First, we reduce the granularity of information in the dependent variable by categorizing payments as either morning or afternoon transactions. Second, we set arbitrary thresholds to classify predicted values into morning or afternoon payments, irrespective of the model's confidence level. For instance, even if a payment is identified with a 51% probability of being an afternoon transaction, we classify it as such. We intend to relax some of these simplifications in future work.

A.1 Gradient Boosting Classifier Model Training

The payments classifier is trained using transaction features as predictors and the payments submission period (morning or afternoon) as the target variable. To train a classifier model like LightGBM, a sequence of weak learners (e.g., DTs) is built, with each learner trained on a modified version of the training set. In boosting, the next learner fits on the residual of the previous one, and this process involves applying weights to each training sample. These sample weights are modified in successive iterations (Friedman et al. 2001).

For model training, tuning, and out-of-sample performance evaluation, we use the standard k-fold cross-validation technique. This is an iterative re-sampling procedure that uses a small number of subsets (k) of the randomly shuffled dataset for in-sample training and out-of-sample testing. This procedure is commonly used to evaluate ML models (Hastie et al. 2009). Figure 15 provides a schematic overview of the k-fold cross-validation procedure, using k set at 5 for illustrative purposes. The dataset is initially shuffled and then divided into two subsets: a training set (for in-sample model training, tuning, and validation) and a test set (for out-of-sample model prediction and evaluation). Subsequently, the in-sample training set is further divided into five folds, each of which is split into a training subset and a validation subset. In each iteration, a unique validation subset (highlighted in blue) is chosen, while the remaining data forms the training subset. A model is estimated on the training subset, evaluated on the validation set, and the score is retained. The scores gathered from each of the k folds contribute to model parameter selection. Finally, the model's performance is tested on the out-of-sample test set (highlighted in orange).

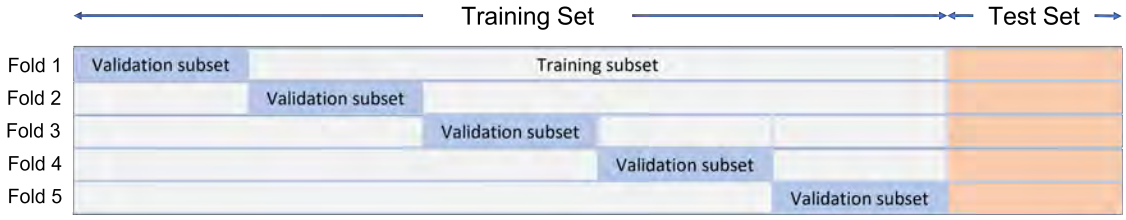


Figure 15: A schematic overview of a k=5-fold cross-validation procedure for the training and tuning of the payments classifier model. First, the dataset is divided into a training set and a test set. Next, for each fold, the training set is split into a training subset (highlighted in light gray) and a validation subset (highlighted in blue). Finally, the out-of-sample test set (highlighted in orange) is used for the prediction.

A.2 Model Training and Evaluation Using Alternative Payments Classifiers

In this section, we train an alternate payments classifier using the payment type as the target variable (rather than the payment period). As laid out in section 4, our dataset contains type 1 (interbank) and type 2 (client-driven) payments. We test this alternative payments classifier using the same transaction features and the same LVTS-T1 sample.

Overall, the share of misclassified payments is larger when using the alternative payments classifier compared to the original one, which classified payments based on their submission time. Additionally, the alternative classifier shows smaller differences in misclassification rates between the clean (approximately 10%) and special-days samples (averaging about 14%). Furthermore, the misclassification rate on the artificial sample is around 80%, which is also lower than that of the original classifier. This suggests that the original classifier is more effective in correctly classifying payments for the LVTS-T1 dataset. Consequently, it can be inferred that the payments classifier model is effective but requires a fine-tuning of its parameters, particularly the target variable for the specific system.

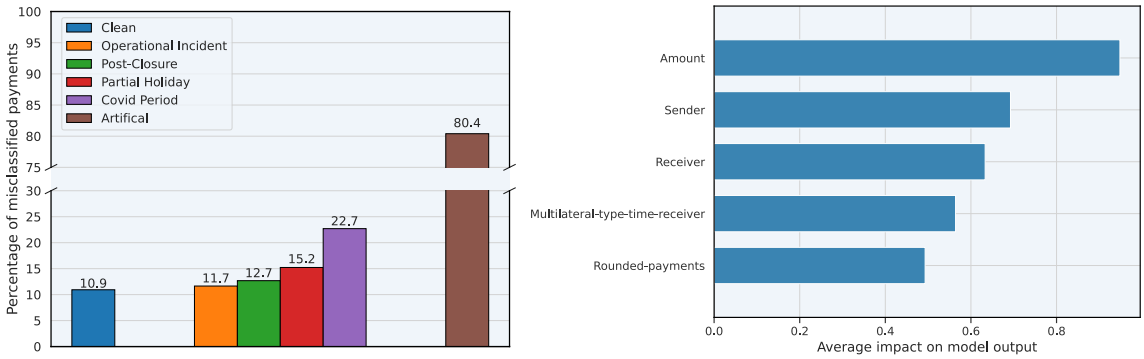


Figure 16: Alternate classifier: (left) Out-of-sample performance of payments classifier measured as the misclassification rate on the clean and special-days transactions. (Right) A few key features were identified and ranked using SHAP global feature importance—measured as mean absolute Shapley values for each instance in the training sample.

The alternative classifier merely learns participants' behavior based on some of the same key features as the original classifier, such as the payment value, sender, receiver, and multilateral-time (see Figure 16, right panel). However, there is also a notable difference; the alternative classification based on the type of payment is also driven by whether the transaction is rounded or not. This reflects the fact that the model correctly captured the fact that most type 1 (interbank) payments are rounded.

B Isolation Forest

The isolation forest (IF) is an unsupervised anomaly detection algorithm introduced by Liu et al. (2008). Unlike other modeling approaches that focus on profiling normal instances to identify anomalies, the IF focuses on directly isolating anomalous instances. It is designed with using the concept that anomalies are rare and distinct points in the feature space. The IF algorithm relies on decision trees known as isolation trees (iTrees). The number of partitions required to isolate a point can be interpreted as the length of the path from the root to a terminating node within the tree.

The IF algorithms can be explained using the following steps (see original paper for complete algorithm (Liu et al. 2008): 1. From a given training dataset, a random sub-sample is selected and assigned to an iTree. 2. By selecting random subsets of features from the chosen sub-sample of data, the iTrees are grown using a random threshold at each split (any value in the range of minimum and maximum values of the selected features). 3. During each split, if the value of a chosen data point is less than the selected threshold, it goes to the left branch otherwise it goes to the right. This process is repeated until each data point is completely isolated or until the maximum (predefined) depth is reached. 4. The above steps are repeated to construct many iTrees by choosing different random subsets of features and sub-samples of data. 5. The anomaly score is then assigned to each of the data points based on the depth of the tree required to isolate that data point.

To elaborate further on the construction of an iTree in steps 2 and 3, the algorithm recursively divides the feature dataset by randomly selecting an attribute and a split value. This division continues until either the node contains only one instance or all data at the node share the same values. Once an iTree is fully grown, each point in the training dataset is isolated at one of the external nodes. Intuitively, data points that are more anomalous are typically easier to isolate in this process, resulting in smaller path lengths in the tree. The path length of each data point in the training set is defined as the number of edges that the point traverses from the root node to reach an external node (final node). Consequently, the anomaly score

for each data point is computed based on the depth of the tree required to isolate that specific point.

In this paper, we employ scikit-learn's Python library to use the IF (Pedregosa et al. 2011). In this implementation, the anomaly score of an input sample is computed as the mean anomaly score of the trees in the forest. The lower anomaly score represents the more abnormal data point. The measure of normality of an observation given a tree is the depth of the leaf containing this observation, which is equivalent to the number of splittings required to isolate this point.

B.1 Isolation Forest Model Training

The IF algorithm is trained using a set of transaction features from the subset of misclassified transactions identified in the first layer. In this step, several binary decision trees are constructed during the training phase. This involves using a random sub-sample of the data and selecting a feature subset at random. After the training, each data point (i.e., transaction) is assigned an *anomaly score* based on the aggregation of the depth of the trees required to classify that transaction. For instance, if a smaller tree (i.e., a tree with fewer splits) is needed to segregate the transaction, then a transaction gets a higher anomaly score than those that require deeper trees (i.e., with more splits). This is illusrated in Figure 17 where the • transaction can be isolated from the other transactions (•) with fewer splits than the • transaction can.

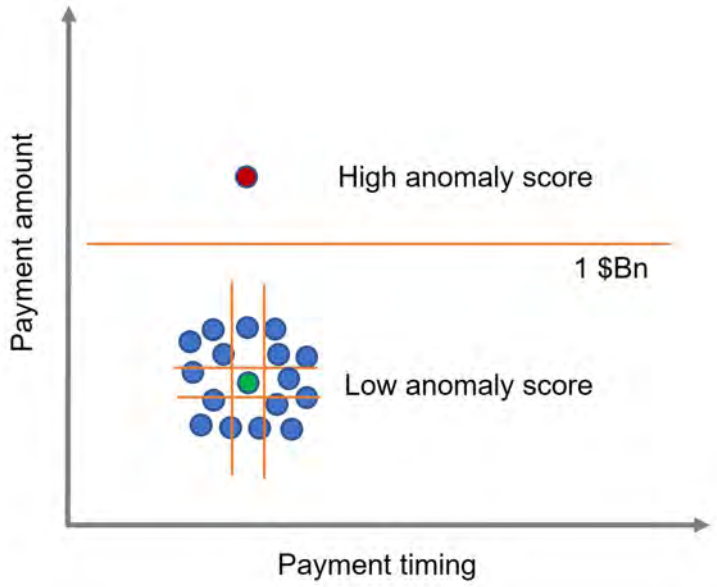


Figure 17: Stylized example of anomaly detection using the IF algorithm. Although both the • and • transactions are submitted at the same time, the • transaction is larger than 1 billion Can$, whereas the • and all • transactions are smaller than that. Consequently, the • data point can be isolated from the • points with fewer splits than the • data point can. The model will therefore assign a higher anomaly score to the • transaction than to the • transaction.

Parameter selection of unsupervised models is challenging due to the lack of target variables. Therefore, we use an ad-hoc approach for tuning the IF model. We fine-tune the model to maximize the average anomaly score on a predefined fraction of the dataset. Different sets of model parameters are selected, and for each set, we calculate the average anomaly score for the designated subset (consisting of transactions with high anomaly scores). Eventually, we adopt the model with the parameters that yield the highest average anomaly score within that subset. As illustrated in Figure 17, we would select the model attributing the highest anomaly score to the • data point.

B.2 Performance of the Standalone Isolation Forest Model

As an alternative to our two-layered approach, we directly trained the unsupervised learning-based IF model on the LVTS-T1 training data, eliminating the preceding payments classifier model. The results indicate that the average anomaly scores for artificially manipulated transactions are lower when employing the standalone IF model (i.e., when not integrated into the two-layer framework), compared to its integration within our framework. This suggests that the IF model within the framework may be more effective in isolating anomalies. This improved performance is likely attributed to the increased complexity resulting from the sheer volume of daily typical transactions within HVPSs. The effective screening of these typical transactions by the first layer streamlines subsequent anomaly detection in the second layer. This underscores the necessity of a two-layer approach and the significance of an efficient payments classifier in the initial layer.

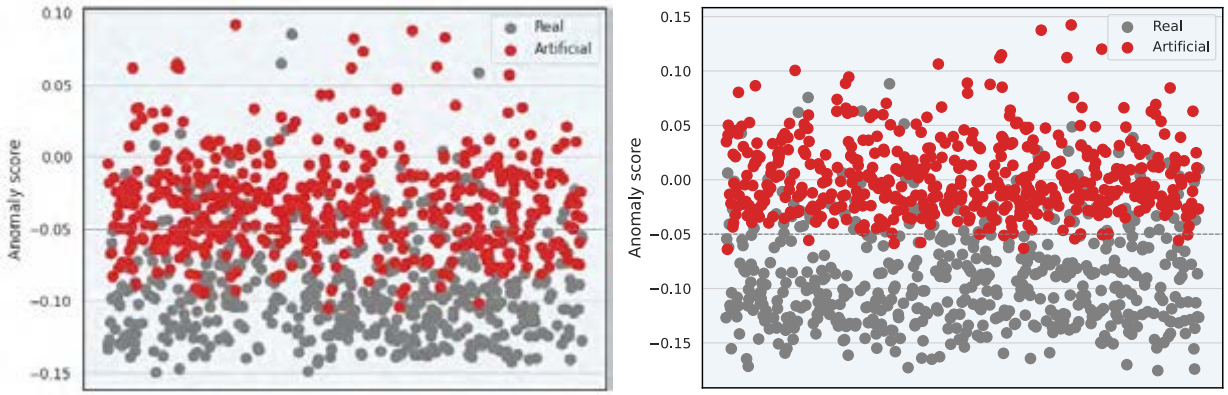


Figure 18: Performance comparison of the IF model (left) when trained independently on the entire dataset without the first layer and (right) as part of the framework in the second layer, using only misclassified samples from the first layer. The out-of-sample anomaly score on the set of transactions before (real) and after (artificial) manipulation. A transaction with a high (+ve) score is likely more anomalous than a transaction with a low (-ve) score.

C Artificial Anomalous Transactions Sample

The artificial anomalous transactions in our dataset are crafted by manipulating specific transaction attributes. For instance, we experiment with altering payment amounts or shifting payment submission times across different periods. Specifically, we employ the following steps to attain the sample of artificial transactions:

- We select 500 random payments from the out-of-sample subset.
- In this subset of payments, the amount featured is altered according to a uniform distribution of 5 to 10 billion Can$.
- Similarly, the period-time feature is altered according to a uniform distribution between 0 and 1000 seconds. This will generate more payments during the early parts of the periods.

Altering payment amounts to higher values and adjusting the period-time to lower values results in generating high-value payments sent earlier in the periods, which is rare in reality. The distribution of real and artificially manipulated transaction samples is depicted in Figure 19, which includes histograms comparing real and artificially generated subsets of transactions. The left side of the figure illustrates the impact of altering payment amounts, while the right side shows the effects of manipulating transaction

submission times for LVTS-T1 payments. These artificial transactions are intentionally designed to exhibit substantial deviations from typical transaction behavior. We employ these extreme cases to assess whether our model has the capability to detect such outliers. Note that for the purpose of this test case, only two features are altered. However, it is possible to simultaneously modify more features.

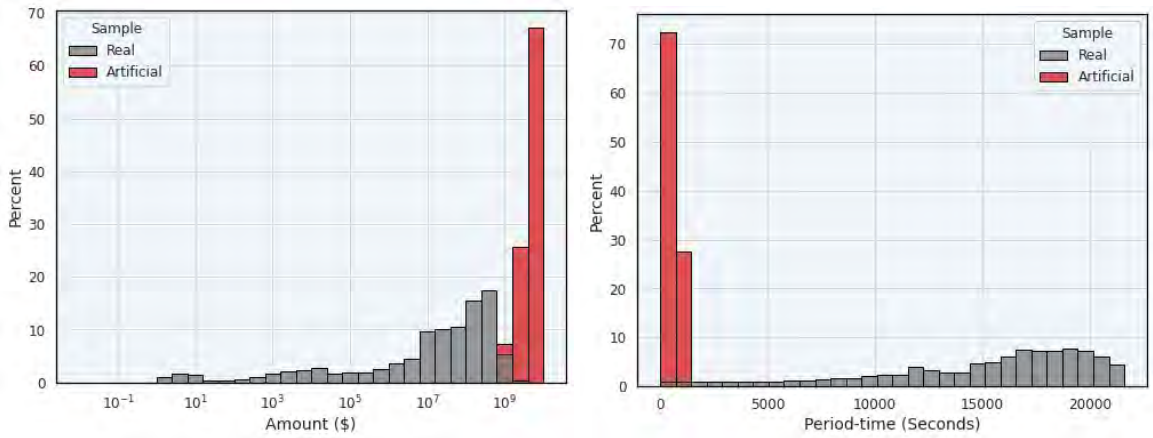


Figure 19: Histograms to compare artificial transaction samples (red) and their real counterparts (gray). These samples are generated by manipulating payment amounts (on the left) and submission periods (on the right) for LVTS-T1 payments.

D The Shapley Values and SHAP for Model Interpretation

The Shapley values is a method from coalitional game theory that provides a way to fairly distribute the *payout* among the *players* by computing the average marginal contribution of each player across all possible coalitions (Shapley 1953; Osborne and Rubinstein 1994).

For a coalitional game, (N, v), where N is a finite set of players indexed by i and v is the utility function or payoff function, the Shapley value can be obtained by this theorem, which satisfies the symmetry, dummy, and additivity axioms (Osborne and Rubinstein 1994):

$$\phi_i(N,v) = \underbrace{\frac{1}{N!}\sum_{S\subseteq N\setminus\{i\}}}_{\text{average over all } S} \underbrace{|S|!\,\Big(|N|-|S|-1\Big)!}_{\text{possible coalitions}} \underbrace{\Big[v(S\cup\{i\})-v(S)\Big]}_{\text{marginal value}}.$$

At a high level, the above equation can be split into three parts. The last part of the equation (the marginal value) gives the marginal contribution of an individual player i, when added to the coalition S that does not have i. The middle part shows how to compute different possible ways in which we could have formed the coalitions. Then, we take an average of possible ways that we could have done the marginal value calculation. The SHAP proposed by Lundberg et al. 2020 uses the Shapley values to explain the model predictions in terms of the marginal contribution of each predictor. The SHAP specifies the explanation of model $\mathscr{F}$ as a linear model of coalitions:

$$\mathscr{F}(S) = \phi_0 + \sum_{i=1}^{M} \phi_i S_i, \tag{1}$$

where $S \in \{0,1\}^M$ is the coalition vector with maximum M coalitions and ϕ_i the Shapley value for ith player.

In S the entry 1 means the corresponding player is present and 0 means the player is absent.

To illustrate, consider the binary classification problem as a *game*. Then the Shapley values can be used to fairly distribute the *payout* (= the prediction) among the *players* (= the predictors). Note: for the computation of the Shapley values in the SHAP, the zero means the corresponding predictor is absent. In that case, the absent predictors' value is replaced by a random value from its sample (Lundberg et al. 2020; Molnar 2020). The procedure is further illustrated as follows:

1. Consider a binary classification (class 0 or 1) problem with three predictors (Figure 20).
2. The average prediction (or base value) of the model is 0.5 (i.e., a 50-50 chance of being in either class). For the current transaction, our model predicts class 1.
3. By computing the Shapley values for all possible coalitions among three predictors, we can explain the difference between the actual prediction (1.0) and the base value (0.5) in terms of each predictor's marginal contribution.
4. In our example, predictor 1 increases the chances of class 1 by 20 percentage points, predictor 2 pushes it down by 10 points, and predictor 3 contributes +40 points. Thus, together they increase the prediction by 50 points from the average prediction, leading to the final prediction of class 1.

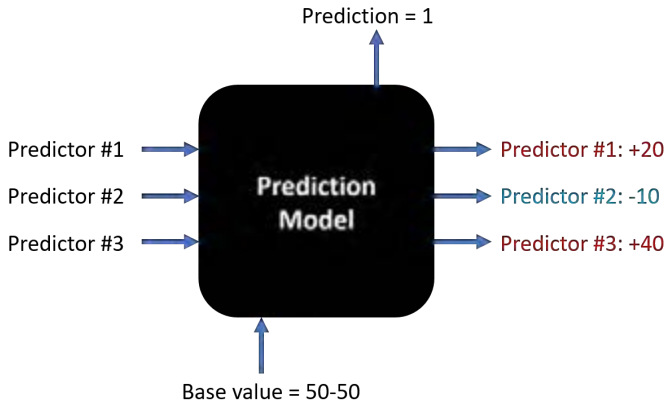


Figure 20: The SHAP explainer provides the marginal contribution of each predictor.

The SHAP values tell us which predictor contributes the most in the current instance of the prediction, that is, a local interpretation. Similarly, by using the Shapley values for each instance in the sample, we can get the average contribution of each predictor for that sample. That would give us a global interpretation of the model in terms of its feature importance. However, it is important to remember that these are only for the chosen model, and they do not explain the causality.

The SHAP package developed by Lundberg and Lee 2017 provides various tools to visualize the Shapley values computed for various ML models commonly used for predictions. For instance, the feature importance plots and summary plots (Figure 7 and Figure 11) are useful for global model interpretations. The force plots (Figure 9 and Figure 10) are useful for local interpretation, that is, at each instance of prediction. Also, the dependence plots (Figure 8) could be valuable for understanding the relationships between given predictors and the targets. The SHAP, although a powerful model-agnostic ad-hoc tool developed based on theoretical foundations for model interpretability, has some shortcomings, and it should be used with caution (Molnar 2020; Slack et al. 2020).

Previous volumes in this series

1187 May 2024	Nothing to hide? Gender and age differences in willingness to share data	Olivier Armantier, Sebastian Doerr, Jon Frost, Andreas Fuster and Kelly Shue
1186 May 2024	Unconditional convergence in the Mexican manufacturing sector (1988-2018)	Alex Rivadeneira
1185 May 2024	Allocative efficiency and the productivity slowdown	Lin Shao and Rongsheng Tang
1184 May 2024	Determinants of currency choice in cross-border bank loans	Lorenz Emter, Peter McQuade, Swapan-Kumar Pradhan and Martin Schmitz
1183 May 2024	Why DeFi lending? Evidence from Aave V2	Giulio Cornelli, Leonardo Gambacorta, Rodney Garratt and Alessio Reghezza
1182 April 2024	Reserve requirements as a financial stability instrument	Carlos Cantú, Rocio Gondo and Berenice Martinez
1181 April 2024	Synthetic controls with machine learning: application on the effect of labour deregulation on worker productivity in Brazil	Douglas K G Araujo
1180 April 2024	Incentive-Compatible Unemployment Reinsurance for the Euro Area	Alexander Karaivanov, Benoit Mojon, Luiz Pereira da Silva, Albert Pierres Tejada and Robert M Townsend
1179 April 2024	The impact of artificial intelligence on output and inflation	Iñaki Aldasoro, Sebastian Doerr, Leonardo Gambacorta and Daniel Rees
1178 April 2024	Finternet: the financial system for the future	Agustín Carstens and Nandan Nilekani
1177 March 2024	Pre-publication revisions of bank financial statements: a novel way to monitor banks?	Andre Guettler, Mahvish Naeem, Lars Norden and Bernardus F Nazar Van Doornik
1176 March 2024	The effect of Covid pension withdrawals and the Universal Guaranteed Pension on the income of future retirees in Chile	Carlos Madeira
1175 March 2024	Unmitigated disasters? Risk- sharing and macroeconomic recovery in a large international panel	Goetz von Peter, Sebastian von Dahlen, and Sweta Saxena

All volumes are available on our website www.bis.org.

BANK OF CANADA
BANQUE DU CANADA

BANK FOR INTERNATIONAL SETTLEMENTS

Machine Learning Framework for Pattern Recognition and Anomaly Detection in Payment Systems

BIS working paper link: https://www.bis.org/publ/work1188.htm

Ajit Desai[1], Anneke Kosse[2], and Jacob Sharples[1]

[1]Bank of Canada, [2]Bank for International Settlements

BoC-IFC Satellite Seminar on Granular Data: New Horizons and Challenges for Central Banks

The opinions here are of the authors and do not necessarily reflect the ones of the Bank of Canada or the Bank for International Settlements

Motivation

- Payment systems are core to the national financial infrastructures
  - Lynx in Canada, Fedwire in US, and Target2 in Europe
- Early identification of changes in payment patterns could help ensure their safety
  - Deviation from usual patterns could hint at **cyber event**, market stress, systemic stress, payments fraud, money laundering, terrorism financing, or an operational issue
  - **Cyber attacks** are an increasing concern: In 2016, hackers attempted to steal billion dollars from the Bangladesh's reserves account at the New York FED, in 2018, similar attacks on Mexico's and Banco de Chile's payment system resulted in millions of dollars in losses
  - If went undetected the effects would be amplified through financial system and could be catastrophic

Challenges



- Problem of finding needles in a haystack
  - Majority of payments are usual behavior
  - Extensive size of payments network
  - Complex strategic interactions
  - Limited information (or features)

Granular data

- Use of granular data and advance analytical tools is recommended
  - 2019 BIS Report: highlights the importance of using **granular data** & analytical tools for wholesale payments fraud detection and prevention
  - 2019 FED Report: emphasizes the necessity of and outlines steps to enhance payments system safety (Fraud-Classifier model)
  - 2019 TARGET Report: underlines the threat posed by fraud and recommends **advanced approaches** to detect fraudulent transactions
- Currently, anomaly detection occurs at the individual participant's level, but **centralized monitoring** has the potential to enhance these individual efforts

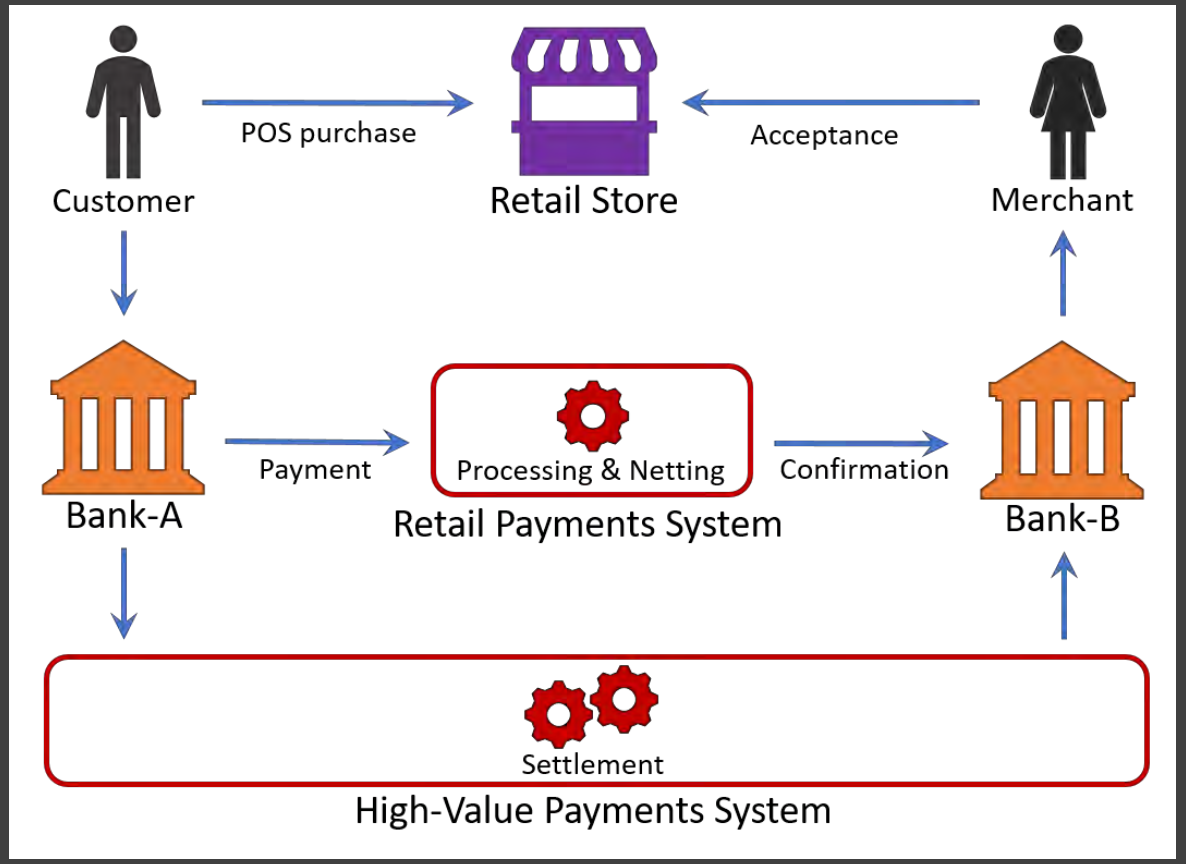
Machine learning

- Tripels et al. (2017): Anomaly detection in payments systems
- Glowka, M. (2019): Profiling banks using cluster analysis with payment system
- Sabetti and Heijmans (2020): Detecting anomalous flows in the Canadian ACSS
- Leon et al. (2020): Pattern recognition of financial institutions' payment behavior
- Arevalo et al. (2022): Clustering of anomalous payments in Salvadorian HVPS
- **Desai et al. (2024): Finding a needle in a haystack: a machine learning framework for anomaly detection in payment systems**

Outline

- Canadian payments systems
- Granular payments data
- Machine learning framework
- Results and discussion
- Conclusions

Payments systems



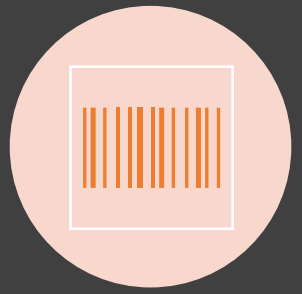
Payments data



Timely:
available immediately after the settlement



Precise:
no measurement or sampling error



High-frequency:
transaction level or daily aggregates

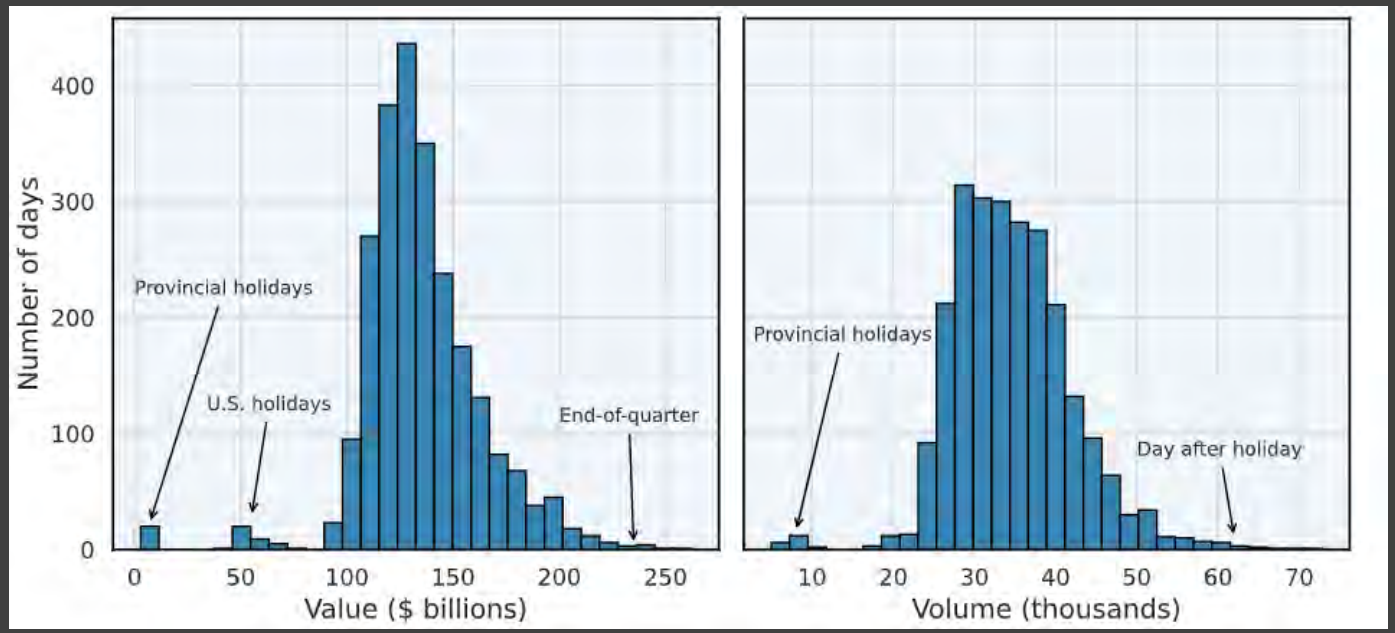


Broad:
15+ years, many financial activities across Canada

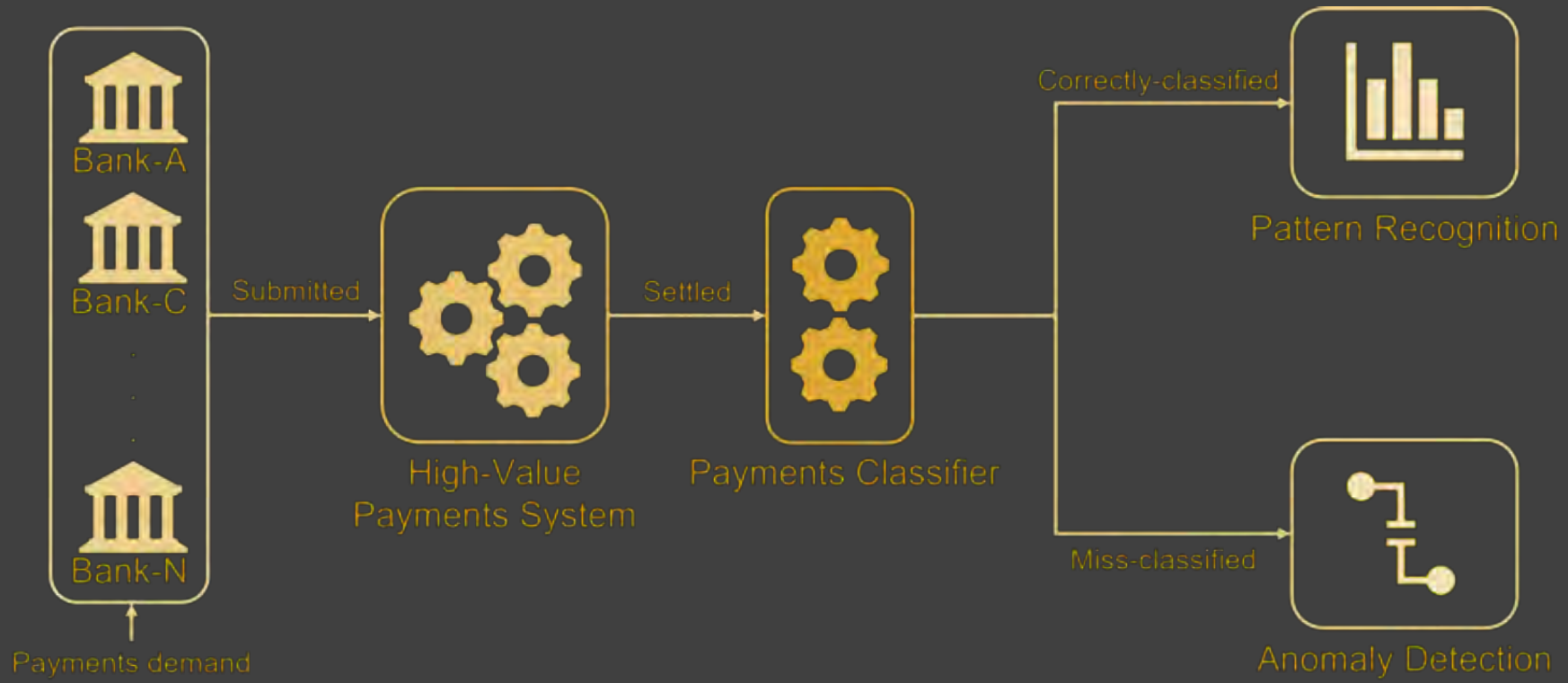
Granular data sample from Canadian HVPS

Anonymized transaction-level settlement data for period: 2012-2021

- Transaction features, liquidity features, timestamp features, intraday timing features
- Transactions on special days are removed and only used for testing

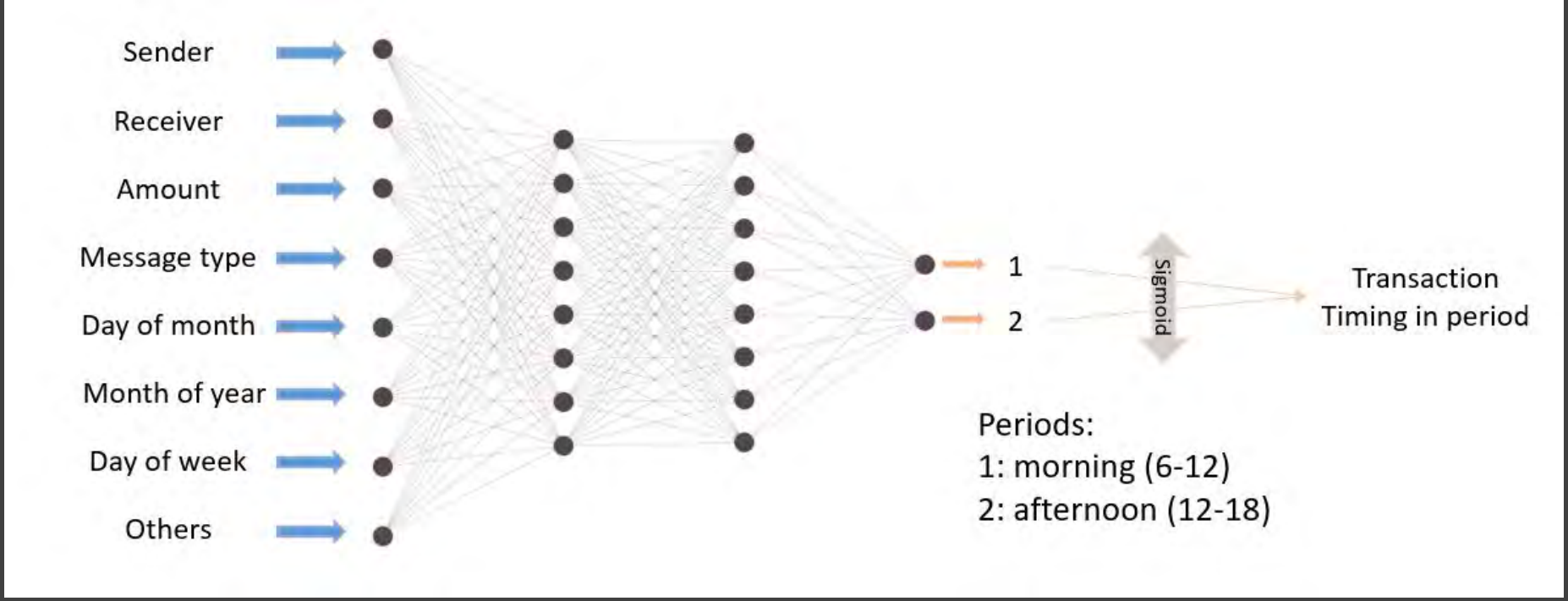


Layered framework: two-step process



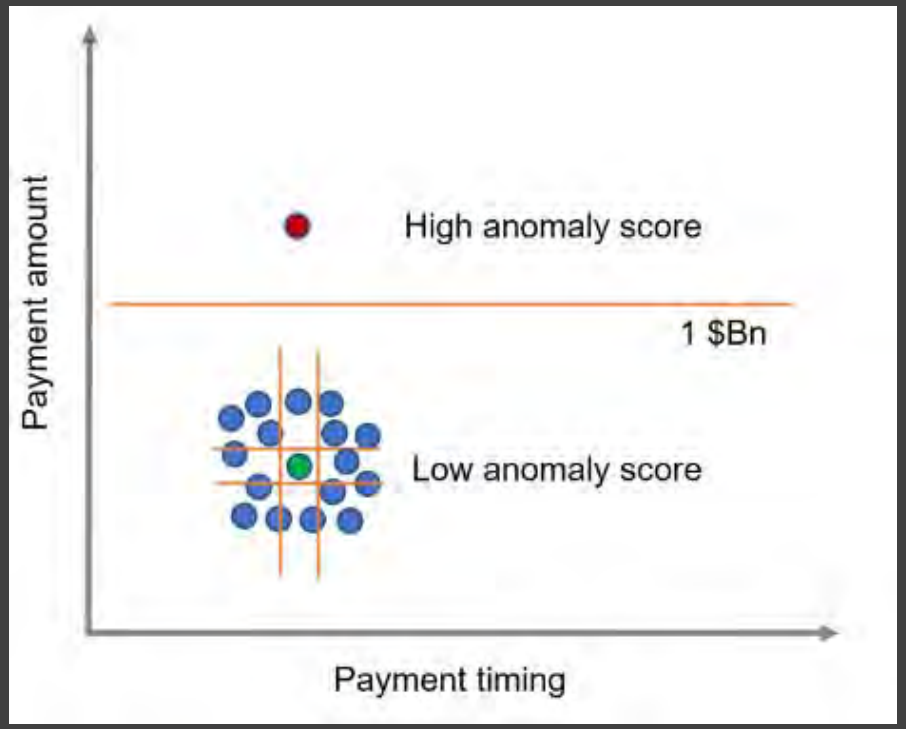
Method: payments classification

Any **classifiers** could be used to classify payments, ML models performed best



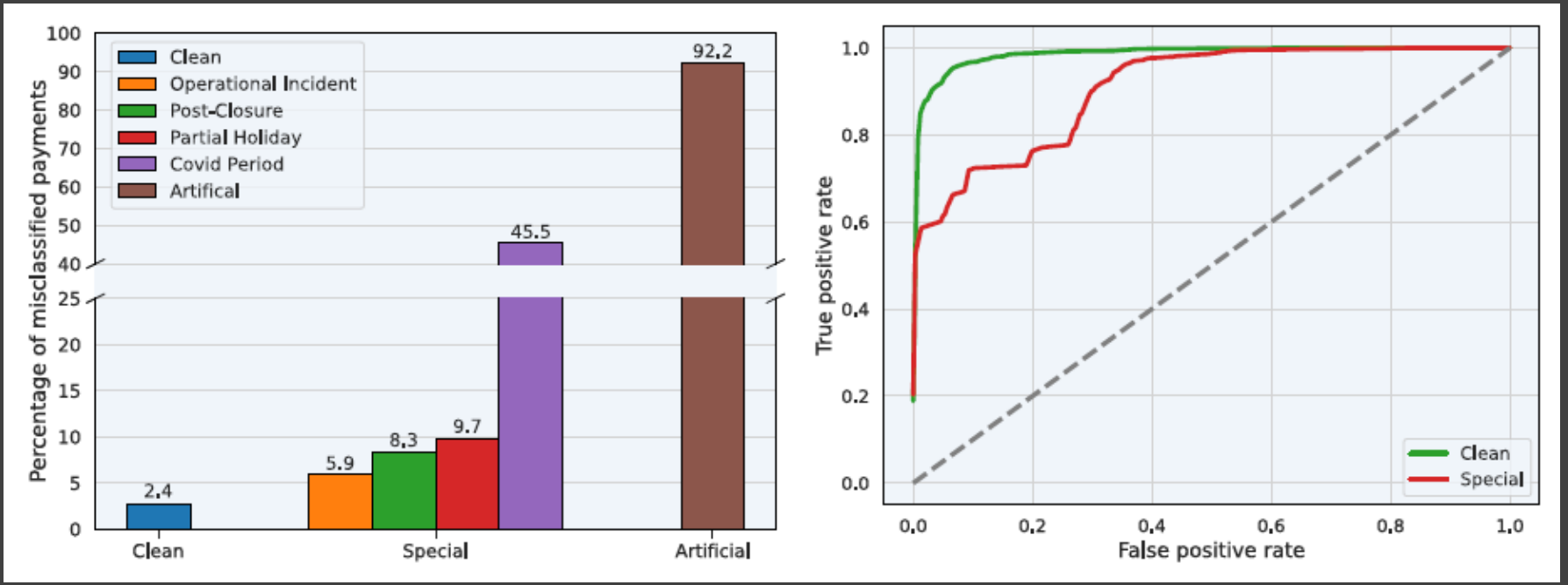
Method: anomaly detection

Isolation forest model can be used for scoring transactions and anomaly detection



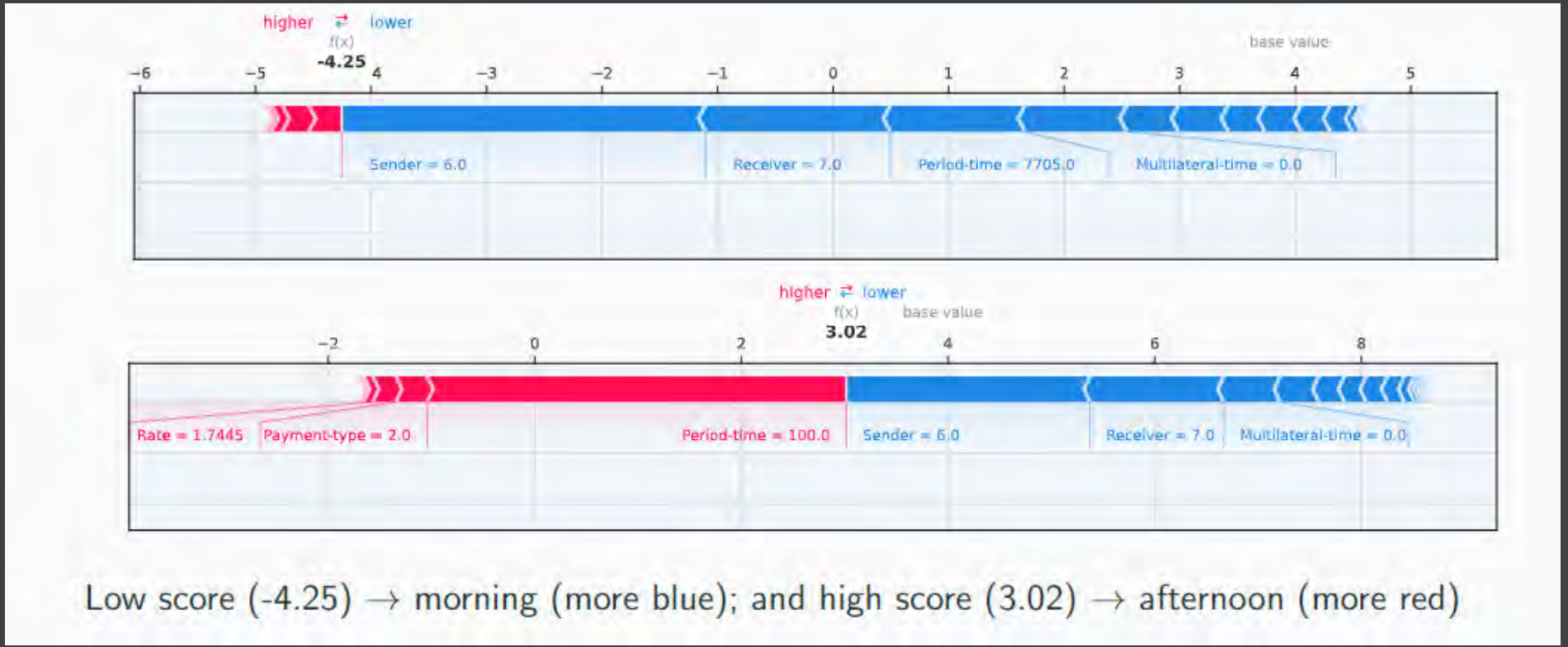
Results: payments classifier performance

Model performance: Normal days (clean) data for training and mix of clean and special days (holidays, post-closure and operational incident) for testing



Results: evaluation on artificial transactions

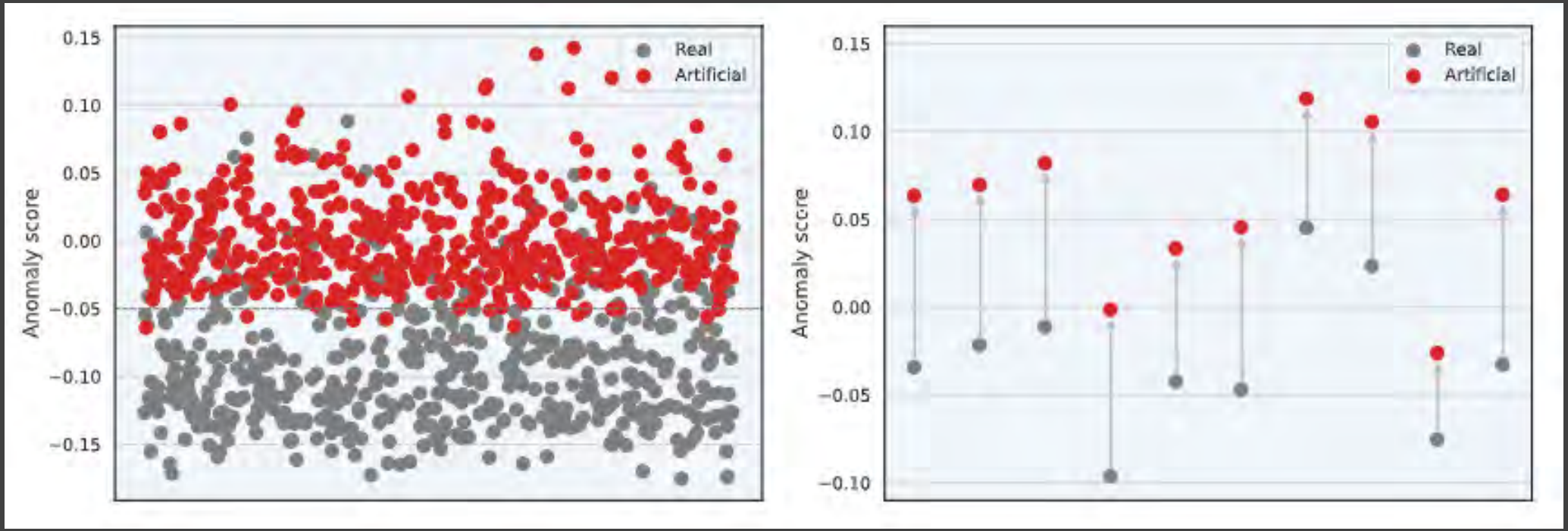
(Top) actual transaction, (bottom) with artificially manipulated period-time



Low score (-4.25) → morning (more blue); and high score (3.02) → afternoon (more red)

Results: isolation forest performance

Anomaly scores for transactions before (real) and after (artificial) manipulation.



Takeaways

- Layered approach helps focus on a smaller subset for anomaly detection—simplifying the problem
- Our approach provides a systematic way to handle anomalies—especially in a large dataset without (or only a few) known anomalies
- This framework is extensible to incorporate more algorithms for performing similar tasks simultaneously to improve the robustness
- It is also general to be used for different types of payment systems

BANK OF CANADA
BANQUE DU CANADA

BANK FOR
INTERNATIONAL
SETTLEMENTS

Thank you!

The opinions here are of the authors and do not necessarily reflect the ones of the Bank of Canada or the Bank for International Settlements

IFC Satellite Seminar on "Granular data: new horizons and challenges for central banks"

Data and analytics risk management framework for central banking[1]

Maryam Haghighi and Vivianne Taillefer,
Bank of Canada

[1] This contribution was prepared for the IFC Satellite Seminar held at the ISI 64th World Statistics Congress, co-organised with the Bank of Canada in Ottawa, Canada, on 15 July 2023. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Canada, the BIS, the IFC or the other central banks and institutions represented at the event.

Data and analytics risk management framework for central banking

By:
Maryam Haghighi and
Vivianne Taillefer

Bank of Canada

The views expressed in this presentation are those of the speaker and do not necessarily reflect the official views of Bank of Canada





Outline

- Introduction
- Data & Analytics
  - Definitions
  - Changing Landscape
- Context
  - Benefits and Risks
  - Legislation, regulations and other initiatives
  - Research & News
- Data and Analytics Risk Management
  - Pillars
  - Responsible AI
- Next Steps
- Conclusion

Introduction

Data and analytics have changed drastically in the past years





















What's new?

Data

- More velocity, granularity, and volume
- Comes from various sources



Analytics

- Increased use of AI
- Ethics of advanced analytics
- Disruptive nature of generative AI

Indeed, it seems like we have entered a new era…



Lawyer apologizes for fake court citations from ChatGPT

By Ramishah Maruf, CNN
Updated 3:28 PM EDT, Sun May 28, 2023



INNOVATIONS

ChatGPT took their jobs. Now they walk dogs and fix air conditioners.

Technology used to automate dirty and repetitive jobs. Now, artificial intelligence chatbots are coming after high-paid ones.

By Pranshu Verma and Gerrit De Vynck
June 2, 2023 at 6:00 a.m. EDT



Disinformation Researchers Raise Alarms About A.I. Chatbots

Researchers used ChatGPT to produce clean, convincing text that repeated conspiracy theories and misleading narratives.



Science News *from research organizations*

Study finds ChatGTP outperforms physicians in providing high-quality, empathetic advice to patient questions

Date: April 28, 2023

Source: University of California - San Diego



MOTHERBOARD
TECH BY VICE

A Judge Just Used ChatGPT to Make a Court Decision

The case is the first time a court has admitted to using the AI text generator's answers in a legal ruling.

By James Rose
NEW YORK, US

February 3, 2023, 12:23pm Share Tweet Snap

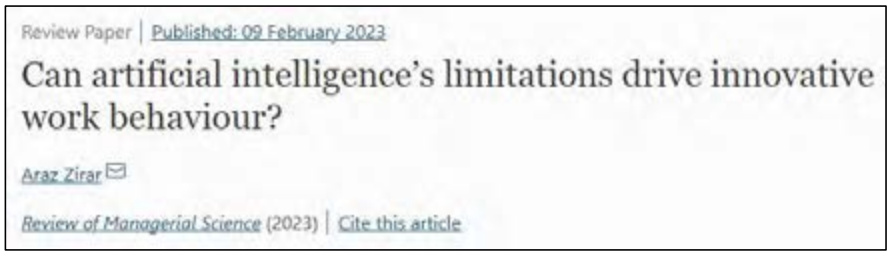
Which is reflected in research…



International Journal of Information Management
Volume 71, August 2023, 102642

Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy ☆

Yogesh K. Dwivedi, Nir Kshetri, Laurie Hughes, Emma Louise Slade, Anand Jeyaraj, Arpan Kumar Kar, Abdullah M. Baabdullah, Alex Koohang, Vishnupriya Raghavan, Manju Ahuja, Hanaa Albanna, Mousa Ahmad Albashrawi, Adil S. Al-Busaidi, Janarthanan Balakrishnan, Yves Barlette, Sriparna Basu, Indranil Bose, Laurence Brooks, Dimitrios Buhalis, Lemuria Carter, ...Ryan Wright

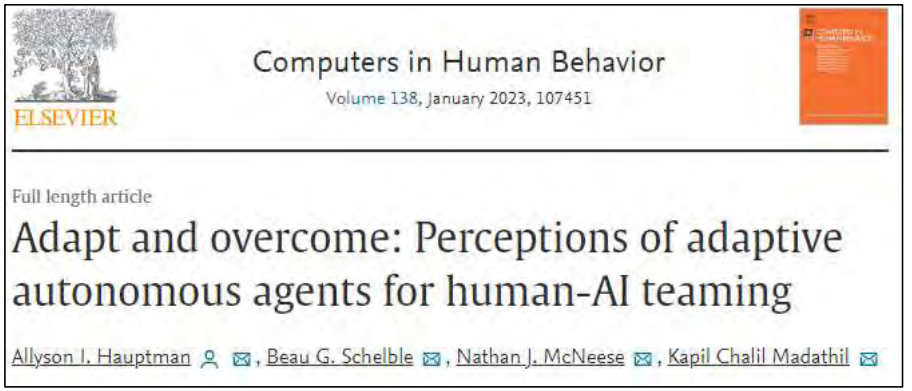


Review Paper | Published: 09 February 2023

Can artificial intelligence's limitations drive innovative work behaviour?

Araz Zirar

Review of Managerial Science (2023) | Cite this article



Computers in Human Behavior
Volume 138, January 2023, 107451

Full length article

Adapt and overcome: Perceptions of adaptive autonomous agents for human-AI teaming

Allyson I. Hauptman, Beau G. Schelble, Nathan J. McNeese, Kapil Chalil Madathil



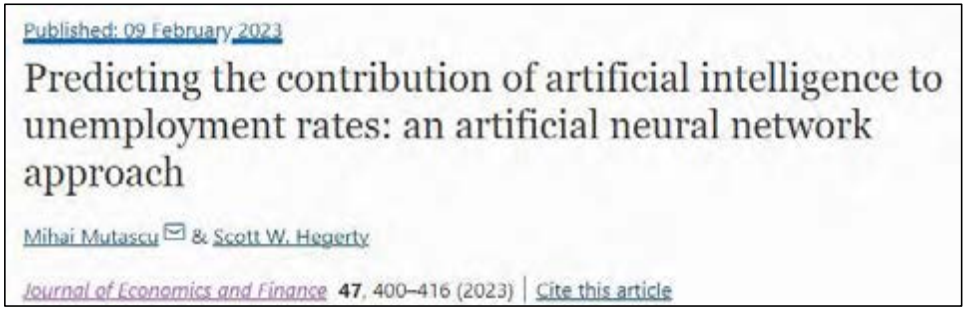
EDITOR'S CHOICE

Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care

Sendhil Mullainathan and Ziad Obermeyer

The Quarterly Journal of Economics, Volume 137, Issue 2, May 2022, Pages 679–727, https://doi.org/10.1093/qje/qjab046

Published: 03 December 2021



Published: 09 February 2023

Predicting the contribution of artificial intelligence to unemployment rates: an artificial neural network approach

Mihai Mutascu & Scott W. Hegerty

Journal of Economics and Finance **47**, 400–416 (2023) | Cite this article

Change brings…

OPPORTUNITIES



and



RISKS



Reduction of menial tasks

Innovation

Responsiveness

Oversight

Data Exploration

Quality

Ethics

Efficiency

Productivity

Clarity

Rights

Responsibility

Agility

Research

Scalable

Law

Decrease in repetitive tasks

Safer workplace

Risks


Opportunity Risk
Unskilled Workforce
Privacy
Accuracy
Psychological Health
Quality
Security
Explainability
Marginal Groups
Productivity
Bias
Ethics
Cyber Threats
Rights
Deception
Copyright
Safety
Law
Less entry level jobs
Data Accessibility

In summary…



Data and analytics risks as we know them today are only the tip of the iceberg.

In the coming years, our exposure to an increasing number of unknown risks related to various data types and artificial intelligence are to grow significantly.

Leveraging Modern Data and Analytics in a central bank context



Public Landscape

Increasingly, institutions are elaborating initiatives for AI systems and non-traditional/alternative data usage.

Government of Canada

- Artificial Intelligence and Data Act (AIDA)
- TBS Automated Decision-Making Directive

GOV.UK

- A Pro-Innovation Approach to Regulating AI

- **Artificial Intelligence Act** (AIA)

OSFI BSIF

- FIFAI (Forum) on Canadian Perspective on Responsible AI
- **Revisions to E-23 on Model Risk Management**

THE WHITE HOUSE

- Blueprint for an AI Bill of Rights

EUROPEAN CENTRAL BANK
EUROSYSTEM

- Provided advice on the AIA

Data and Analytics Risk Management

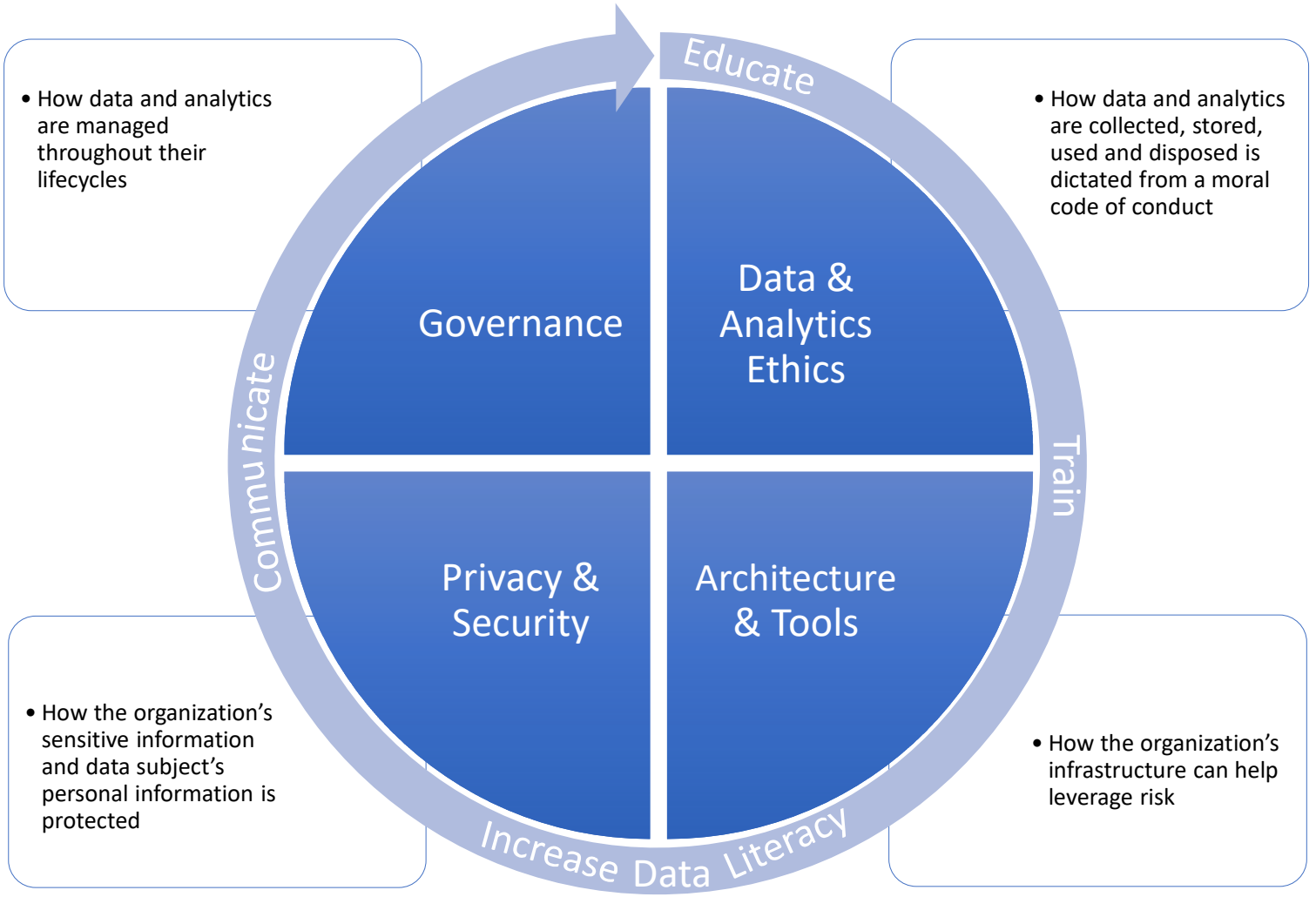
Data and Analytics Risk Management leverages the traditional risk management framework:

- Risk Identification
- Risk Assessment
- Risk Mitigation
- Risk Monitoring
- Risk Reporting



Pillars

Managing those risks entails reviewing current processes and capabilities alongst these four pillars while enabling employees to remain up-to-date in their data and analytics knowledge.



GOVERNANCE

- Review different initiatives including Data Governance and Model Risk

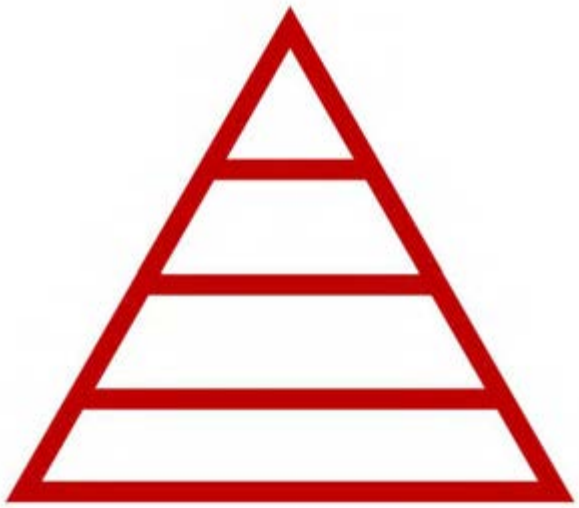
- Create tools for data and models onboarding, management and use (For example, see the Algorithmic Impact Assessment tool)

PRIVACY AND SECURITY



- Assess data classification processes
- Research and use advanced techniques for privacy enhancement
- Build an awareness and educational program around AI systems and tools

ARCHITECTURE AND TOOLS



- Embed risk management checkpoint within the infrastructure (e.g. privacy by design)
- Enable data users to become self-sufficient with low-code tools
- Establish a dataset certification process

Data and Analytics Ethics



- Review risk management processes
- Establish Responsible AI principles and form an Advisory committee on Responsible AI
- Create educational tools for using AI