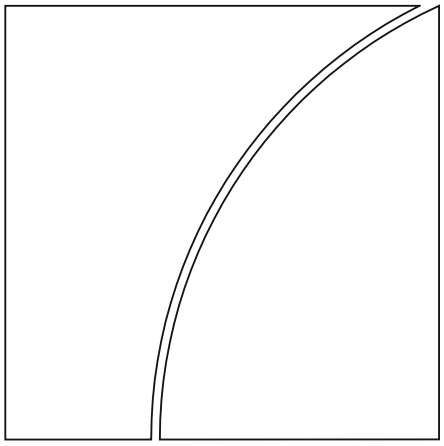
Irving Fisher Committee on Central Bank Statistics

BIS



IFC Bulletin

No 65

New insights from financial statements

August 2025

Contributions in this volume were prepared for the proceedings of workshop on "New insights from financial statements" co-organised by the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain and the Irving Fisher Committee (IFC) in Madrid, Spain, 17 October 2024.

The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the ECCBSO, the Bank of Spain, 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 2025. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1991-7511 (online)

ISBN 978-92-9259-874-7 (online)

New insights from financial statements

IFC Bulletin No 65

August 2025

Proceedings of the workshop on "New insights from financial statements" co-organised by the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain and the Irving Fisher Committee (IFC) in Madrid, Spain, 17 October 2024.

Overview

Leveraging corporate financial statements for policy insights

Dominik Elgg, Expert in data accessibility of public corporate reports of non-financial entities, Department of Data and Statistics, Deutsche Bundesbank

Taejin Park, Head of Financial Markets and Research Support, Monetary and Economic Department (MED), Bank for International Settlements (BIS)

Olivier Sirello, Senior Statistical Analyst, MED, BIS

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

Jose Maria Vidal Pastor, Principal Financial Market Analyst, MED, BIS

1. Improving financial data with innovative techniques and alternative/complementary data sources

Outlier detection optimization using machine learning for improving data quality

Salah Ahl Mbarek, Bank Al-Maghrib (Central Bank of Morocco)

Superstar firms, intangibles, and productivity mismeasurement: evidence from firm-to-firm transactions

Gert Bijnens, National Bank of Belgium, Joep Konings, Nazarbayev University GSB, KU Leuven and CEPR, and Aaron Putseys, KU Leuven

Connecting financial data with intellectual property rights: relevancy and (data) complexity

Julie Callaert, Phoenix Chen, Bart Van Looy, Steven Vanhaverbeke, Jan-Bart Vervenne, and Jesse Wursten, KU Leuven

Uplifting the usability of financial statement data to leverage the quality of the Sectoral Balance Sheet (SBS) estimates

Jittima Dummee, Prossaya Jitapunkul, Natchaya Mahaviriyakul and Pornsawan Rugpenthum, Bank of Thailand

2. Environmental and Social Governance (ESG) and sustainability in financial statements

Green financing and the relationship between banks and non-financial corporations through the lens of balance-sheet interaction

Michael Harrison and Mimoza Shabani, University of East London

Quality over quantity? Using corporations' climate change related disclosure for risk assessment

Friedrich Bähr, Daniela Gasser and Sabine Wukovits, Central Bank of the Republic of Austria

It's in the financials, stupid! But is it certain?

Christian Haas, Ulf Moslener and Sebastian Rink, Frankfurt School of Finance and Management

Quantifying material risks from textual disclosures in financial statements using LLMs and LLM agents – a case study on physical risks

Taejin Park, BIS

3. Harnessing corporate reports to identify risks

The allocation of cash flow by Spanish firms: new evidence on the impact of financial frictions

Martin Farias, Center for Monetary and Financial Studies (CEMFI)

A study on complementary indicators to the ICR for identifying distressed firms

Jiyoung Choi, Sejong Lee, and Hye Won Shin, Bank of Korea

The Banco de España's Central Balance Sheet Data Office Database: a regional perspective

Alejandro Fernández Cerezo, Borja Fernández-Rosillo San Isidro and Natividad Pérez Martín, Bank of Spain

4. Analysis of firms' responses to shocks and policy measures

How resilient companies navigated through Covid-19 pandemics

António Costa and Pedro Cordeiro, Banco de Portugal

The pass-through of cost-shocks to firms' prices and the impact on value added

Fabio Parlapiano, Bank of Italy

Corporate debt structure and heterogeneous monetary policy transmission

Marie Alder, European University Institute, Nuno Coimbra, Bank of France, and Urszula Szczerbowicz, SKEMA Business School

EU regulation on combating late payment in commercial transactions

Markus Federau and Bernd Schwendinger, Vorarlberg University of Applied Science

Irving Fisher Committee on Central Bank Statistics

BIS

Leveraging corporate financial statements for policy insights

Dominik Elgg, Taejin Park, Olivier Sirello, Bruno Tissot and Jose Maria Vidal Pastor[1]

Executive summary

The growing availability of detailed **information in published corporate financial statements provides a key source for policy-relevant insights**. These documents, rooted in formal accounting standards, typically offer a unique perspective on firms' behaviour and vulnerabilities, contributing to monitoring broader economic developments and informing monetary and financial stability decisions.

Meanwhile, the **content of financial statements is continuously expanding**, as firms appear to increasingly disclose additional material in response to evolving regulations, accounting standards and societal expectations. In particular, sustainability issues, such as climate risk and carbon footprint, have become an integral part of corporate reports, even though gathering this information presents important challenges in terms of quality, harmonisation and certification practices.

Central banks have been actively combining information from corporate reports with other data, including of a non-public nature, to enhance their usefulness. Specifically, the growing digitalisation of economies has enabled the **integration of new and alternative data sources** to complement "traditional" financial reports, helping to address quality issues and coverage. These promising opportunities are reinforced by technological innovation, including big data analytics and artificial intelligence (AI), which allows for better access, processing and use of structured and unstructured data, helping to capture previously underutilised information.

As a result, both public and non-public **corporate financial statement data can increasingly be leveraged to support central banks' statistical, analytical and policy tasks**. The greater availability of more comprehensive firm-level information can in particular allow to identify global and regional economic dynamics, facilitate

[1] Respectively, Expert in data accessibility of public corporate reports of non-financial entities, Department of Data and Statistics, Deutsche Bundesbank (Dominik.Elgg@bundesbank.de); Head of Financial Markets and Research Support, Monetary and Economic Department (MED), Bank for International Settlements (BIS) (Taejin.Park@bis.org); Senior Statistical Analyst, MED, BIS (Olivier.Sirello@bis.org); Head of Statistics and Research Support, BIS, and Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC) (Bruno.Tissot@bis.org); Principal Financial Market Analyst, MED, BIS (JoseMaria.VidalPastor@bis.org).

The views expressed here are those of the authors and do not necessarily reflect those of the ECCBSO, the Bank of Spain, the BIS, the Deutsche Bundesbank, the IFC or any of those institutions mentioned.

We thank Martine Druant, Klaus Gerstner, Robert Kirchner, Alfredo Maldonado, Manuel Ortega, Ana Bárbara Pinto and Florian Resch as well as participants at the workshop "New insights from financial statements" (October 2024) for helpful comments and suggestions. We are grateful to Alison Arnot for editorial review and Márcia Cavalinhos for administrative support.

the development of new indicators and forecasts, improve early warning systems and, more broadly, reveal hidden risks, especially during crises such as the Covid-19 pandemic. These elements are essential ingredients that inform monetary and financial stability policies, helping to design and calibrate public interventions and monitor their impact effectively.

Yet **important challenges remain**, in particular regarding the accessibility and interoperability of the vast amount of information contained in corporate reports, including supplementary disclosure such as management's discussion and analysis. Fortunately, important international initiatives are under way to improve the coverage and quality of this information, especially in the area of sustainability disclosures.

Looking forward, **making the most of financial statement data calls for a clear roadmap to enhance their accessibility and adequate sharing as well as their comparability and overall quality**, in at least four main directions. First, enhancing the global statistical infrastructure appears to be a major requirement, particularly by further enhancing global common identifiers, information standards and, broadly, interoperability. Second, stronger collaboration among the various stakeholders involved in the corporate reporting ecosystem can facilitate adequate data-sharing and unlock the full potential of existing information. Third, innovation should be properly pursued to make the most of promising technologies, such as AI, while addressing the important challenges associated with their use. Ultimately, and perhaps more fundamentally, ensuring public trust in corporate disclosures and in turn in their use for official statistics and public policy is essential. From this perspective, central banks can play a key role as "data curators" to safeguard the quality, integrity and authenticity of financial statement data.

1. Introduction

Attention is increasingly paid to leveraging the vast and diverse information contained in corporate financial statements to support statistical work and produce policy-relevant insights. Central banks have been at the forefront of this trend, reflecting their important roles as both producers of financial data in national statistical systems[2] and users of such information to fulfil their mandates, especially in the areas of monetary and financial stability (Rosolia et al (2021)).

The concept of "financial statements"[3] is often used in a wider sense. **In simple terms, financial statements are formal records that provide accounting information about a firm** over a specific period.[4] These statements mainly include the balance sheet, which shows the firm's financial position at a specific point in time,

[2] For example, central banks are tasked with compiling financial accounts and external statistics in many jurisdictions; see Diz Dias et al (2025) and IFC (2020a).

[3] This concept is usually associated with various synonyms, such as annual accounts, annual reports, corporate reports and financial reports.

[4] Typically on a yearly or, for most publicly listed companies, half-yearly or quarterly basis. For further information, see IAS 1.10–11 regarding the definition of a complete set of financial statements, and, in the case of the European Union (EU), Commission Regulation (EU) 2023/1803 (eur-lex.europa.eu).

and income statement, which details the revenues, expenses, profits or losses over a certain period.[5] This information is either published or kept internally but generally made available to public authorities for specific purposes.

In practice, **financial statements have become increasingly complex and multidimensional**, for four main reasons. First, these records may follow different accounting standards and practices, for example in terms of presentation, classifications and consolidation. Second, they may also differ in terms of disclosure requirements. Large, publicly traded firms generally publish their audited accounts regularly, but this may not be the case for other – usually smaller – firms.[6] Third, financial statement data are often combined with other (non-public) sources, which include administrative, business and credit registers, and statistical surveys. Fourth, there has been a general trend towards providing additional non-financial information along with core financial reporting, such as on sustainability and governance issues. Such supplementary disclosure is typically given in "notes" attached to financial statements, with relevant details often published on a voluntary basis.

Reflecting the above factors, one will typically refer to "financial statement data" as the overall (disclosed or not) information related to the financial position and activities of a firm. **The supply of this information has been growing significantly over time**, especially for non-financial firms.[7] It provides rich insights on various corporate activities, which are particularly valuable to shareholders or creditors (Rejison (2025)) as well as public institutions, such as central banks, supervisors and fiscal authorities. Three major aspects are particularly noteworthy from this perspective.

First, **the granularity of the data has proved particularly valuable for analysing complex social and economic developments** and supporting policymaking with greater agility. In particular, micro-level records can help to better assess the increasing interconnectedness of the global financial system – a need clearly evident since the Great Financial Crisis of 2007–09 (GFC) and the expansion of non-bank financial intermediation (BIS (2025)). More broadly, granular financial information can assist users with a wide range of analyses related to structural changes, such as those observed in the wake of the Covid-19 pandemic in 2020–21 (Jahangir-Abdoelrahman and Tissot (2023)).

Second, the **information provided in financial statements has expanded to cover emerging issues**. A clear example is climate risk, which has become an important policy concern and has led to the incorporation of environmental

[5] Other main items include the liquidity position (cash flow statement) and changes in equity related in particular to retained earnings (statement of shareholders' equity).

[6] For example, in a number of jurisdictions, publicly traded firms also disclose a management's discussion and analysis (MD&A) section as part of their financial statements. This information accounts for a company's business and comments on the changes to financial statements in terms of profitability, cash flows, capitalisation and asset structure. This section has become even more relevant with the new corporate sustainability disclosure requirements, such as in the EU; see Directive (EU) 2022/2464 on corporate sustainability reporting (eur-lex.europa.eu).

[7] Noting that a large part of the financial sector is supervised and reports such data in a separate way to central banks and financial supervisors.

indicators, alongside broader sustainability matters, into financial reporting frameworks. These metrics are increasingly integrated into core financial analysis, reflecting their growing importance in economic and financial decision-making processes.

Third, **innovative techniques allow for better leveraging the financial statement data**. Big data analytics have already proved effective in coping with the expansion of micro data sources over the past few decades, supporting central banks' key functions, including statistical production, economic analyses and forecasts, monetary and financial stability assessments, regulatory monitoring and, more recently, the promotion of sustainable finance (IFC (2024a)). A fundamental aspect from this perspective is that new AI-based tools, specifically large language models (LLMs), can greatly help to tap into unstructured data sources, especially textual information contained in published corporate statements.

Despite these developments, **the overall quality [8] of the information contained in financial statements remains an important challenge**, especially in terms of accuracy, frequency, timeliness and completeness. Certainly, reports published by companies should generally be of higher quality when they are subject to external and independent audits, but there can still be important inaccuracies, especially those arising from incorporating this information in general databases. This is even more challenging when authorities need to compile directly by themselves additional/non-public corporate-level indicators to support their functions. Reflecting the above, central bank statisticians typically have to deal with a very large number of accounting questionnaires and often need to manually debug the data or, worse, discard them due to missing or inconsistent information.

Therefore, **accurate, comprehensive and timely financial statement data are increasingly required to effectively support evidence-based decision-making** across various financial and economic domains. This underscores the critical importance of collaboration between reporters, statisticians, analysts and policy users to ensure the quality of financial statement data, facilitate their integration with other data sources and promote their access to and use by relevant stakeholders.

To address these developments and explore emerging opportunities, the Irving Fisher Committee on Central Bank Statistics (IFC) of the Bank for International Settlements (BIS), the European Committee of Central Balance Sheet Data Offices (ECCBSO) and the Bank of Spain jointly organised **a workshop on "New insights from financial statements"** in late 2024 in Madrid, Spain, building on previous IFC work in this area (see IFC (2014, 2017, 2022, 2024a)). This collaborative initiative brought together experts from various fields to discuss innovations and challenges in extracting valuable insights from financial statement data and address information needs.

[8] In line with previous work conducted by the IFC, the concept of "data quality" is understood here in a broad sense, spanning the various characteristics sought for in official statistics and captured by the Fundamental Principles of Official Statistics (UN (2014)). These include in particular accuracy, integrity, security, transparency and trustworthiness. The concept also covers the user side, for instance to ensure that the data are fit for purpose, easily accessible, findable, traceable and reusable, and that they can eventually be deleted adequately. Lastly, it includes ethical aspects, such as assuring impartiality, objectivity, professional independence and social acceptability (IFC (2021c)).

The overview of this *IFC Bulletin* presents the workshop's key contributions across **four main dimensions**:

- The rich and expanding information contained in financial statement data, including on new sustainability issues (section 2)
- The new prospects offered by alternative data sources and innovative tools and techniques (section 3)
- The opportunities for developing better statistics and policy insights on the economy and the financial system (section 4)
- Proposals for addressing the ongoing challenges and advancing the effective use of financial statement data (section 5).

2. The rich information content of financial statement data

Multiple information dimensions

Financial statement data encompass a wide range of information influenced by three key dimensions: (i) accounting standards; (ii) disclosures practices; and (iii) integration with other sources.

First, financial reporting follows **accounting standards and principles** based on the prevailing practices where the firm is incorporated. Noteworthy examples include the Generally Accepted Accounting Principles (GAAP) in the United States[9] and the International Financial Reporting Standards (IFRS) that are recognised in many jurisdictions globally, including Europe. These standards typically aim to harmonise the accounting rules used for corporate reporting – such as in terms of consolidation, the presentation of the items disclosed by firms as well as the classification of the accounts involved.

Second, financial statements depend on specific **disclosure practices and requirements**. Large and publicly traded firms are usually obliged by regulations to disclose accounting reports to the public, with notable differences across jurisdictions.[10] Smaller or privately held entities may not be required to do so, or they may choose to disclose their accounts selectively, for instance through a simplified presentation. A key reason for this is the evident trade-off between full (public) disclosure – with benefits in terms of accountability and transparency – and reporting burden, which is typically greater for smaller firms with limited administrative

[9] Local versions of GAAP can also be implemented in a number of jurisdictions, such as in Switzerland (so-called Swiss GAAP) and the EU; see the EU "accounting directive", Directive 2013/34/EU (eur-lex.europa.eu), as well as van Hulle and van der Tas (2000)). It should be noted that IFRS has had a strong impact in the development of local versions of GAAP over time.

[10] For instance, parent undertakings with securities traded on an EU-regulated market are obliged to set up consolidated financial statements in accordance with IFRS.

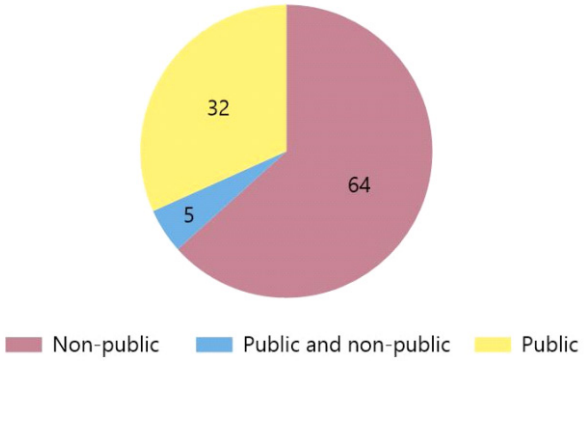
capacities,[11] as well as privacy concerns and peer competition. For example, it is estimated that around two thirds of financial statement data sources in Europe are not publicly disclosed (Graph 1).

Public and non-public financial statement data sources used by European CBSOs

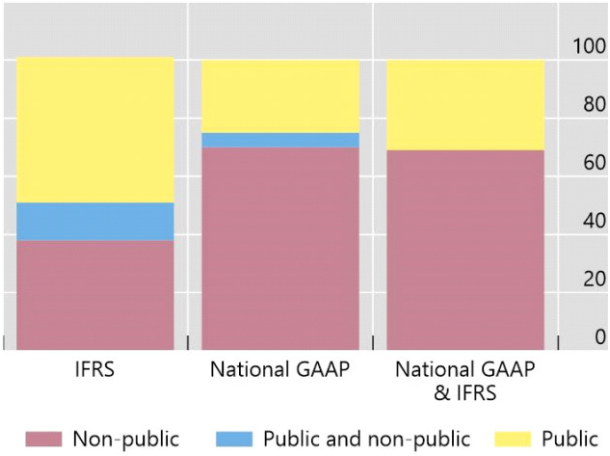
In per cent

Graph 1

A. Types of financial statement data sources



B. Public and non-public sources by accounting standard



CBSOs = central balance sheet data offices; IFRS = international financial reporting standards; GAAP = generally accepted accounting principles.

Source: ECCBSO Report 2021.

Third, authorities typically combine published **financial reports with other (often non-public) sources**, which include administrative, business and credit registers and statistical surveys. These sources depend on various national practices and/or institutional factors related, in particular, to the legal framework governing the collection of firm-level data, their degree of confidentiality and shareability.[12] To this end, a number of jurisdictions have established central balance sheet data offices (CBSOs) that collect, process, disseminate and analyse various types of corporate financial statement data (IFC (2017)).

The value of firm-specific insights

A key factor underlying the richness of the information contained in financial statement data relates to their high granularity. Indeed, micro-level corporate

[11] Reflecting this point, publication in Europe depends for example on whether an entity limits its liability, thereby conditional on the size of the company or whether the entity belongs to a group (Directive 2013/34/EU; see eur-lex.europa.eu).

[12] Historically, these offices, often operated by the statistical department of the central bank in cooperation with the statistical office, were gradually established in the second half of the 20th century, eventually leading to the creation of the ECCBSO (ECCBSO (2021), Ortega (2014)). This group, which mostly but not only comprises EU central banks, aims to improve the analysis of non-financial corporate enterprise data through the exchange of information and joint studies; see eccbso.org.

reports can reveal idiosyncrasies that would otherwise remain undetected by looking at aggregate statistics alone.

First, and obviously, this information represents a **main source to analyse corporate health, behaviour and potential risks**. It provides subtle details for assessing key financial vulnerabilities, such as credit, market or liquidity exposures as well as the expected default probability. One telling example relates to the use by many central banks of financial statement data to assess the credit quality of eligible assets in order to conduct their market operations.

Second, these granular data **allow the analysis of interconnections** between sectors, shedding light on concentration risks and spillover channels in the economy and the financial system – a critical dimension for early identification of vulnerabilities and systemic risk assessment (IFC (2017)). For instance, the careful tracking of economic agents' positions can help to assess corporate exposure to foreign debt and currency mismatches and thereby to better anticipate potential corporate stress before it materialises and reaches systemic dimensions (IFC (2020b)).

Third, **firm-level specificities can in turn provide insights on broader macroeconomic developments**. This aspect proved particularly valuable during the Covid-19 pandemic, as detailed firm-level data helped to precisely decipher its economic impact. For instance, Banco de Portugal used financial statement information to identify the key characteristics that enabled certain firms to weather the crisis more effectively. Similar data also allowed the Bank of Italy to assess how price increases in intermediate goods and energy affected firms' pricing strategies and value added. In other words, a good understanding of corporate-level characteristics helped understand broader and structural macroeconomic developments.

Fourth, the granularity of financial statements also enables a **better understanding of local developments**, complementing the usual set of macroeconomic statistics that are mostly presented at the national level. For example, firm-level financial data collected by the Bank of Spain's CBSO have been combined with locational information from other sources to produce detailed regional insights (Fernández Cerezo et al (2024)).

Fifth, these data also represent a **useful window into future economic developments.** In particular, information on corporate debt structures, risk exposures, investment plans and profit projections provides a forward-looking perspective that ideally complements the traditionally backward-looking nature of macroeconomic statistics. For instance, domestic companies allocating a higher proportion of cash flow to debt repayment (instead of investments, cash savings and dividends) are found to experience larger financial frictions (Farias (2024)). This suggests that monitoring cash flow allocation could help to anticipate episodes of financial stress.

Filling new data needs, especially to tackle sustainability issues

The **content of financial statement data is constantly evolving**, as are the requirements to address information gaps, depending on national circumstances, regulations, standards and stakeholders' needs. But, perhaps more significantly, these

changes have increasingly blurred the boundary between "financial" and "non-financial" information in the context of corporate reporting.

A key example relates to so-called "sustainability reporting", which integrates environmental, social and governance (ESG) information in corporate disclosure.[13] This development is essentially driven by regulatory measures, evolving global reporting standards and growing public interest in sustainability issues in recent years (eg impact of climate change, social responsibilities, diversity and inclusion). Increased appetite by firms for ESG disclosure is also influenced by organisational, reputational and legitimacy strategies, for example to show environmental friendliness and sustainability progress (Arkoh et al (2024)). As a result, once considered peripheral to reporting frameworks traditionally centred on "core" financial performance indicators, the ESG dimension has now become an important component of comprehensive firm-level assessments, for instance in terms of risk identification, investment analysis and attractiveness.

The above changes reflect a **fundamental shift in how markets conceptualise corporate performance and value creation**. For instance, climate-related risk has become a major concern, representing a requirement or a strong incentive to report the carbon footprint and associated mitigation measures (CMFB (2023a), IFC (2024b)). Yet the absence of widely agreed standards has led to the proliferation of a variety of practices, raising important challenges in terms of quality – as firms may selectively disclose data with little degree of assurance, consistency and comparability – and disinformation – such as the well-known issue of "greenwashing" (IFC (2025a)).

Against this backdrop, **enhanced sustainability reporting has been supported by key regulatory measures in recent years**, particularly in Europe, with the Corporate Sustainability Reporting Directive (CSRD) and the respective European Sustainability Reporting Standards.[14] In addition, there is a growing recognition globally that financial statements should incorporate climate-related data in alignment with International Sustainability Standards Board (ISSB) standards (IFRS Foundation (2023)).

Increasing the amount and variety of sustainability indicators disclosed at the level of individual firms would also support **addressing pressing information gaps at the macro level**. For instance, over 100 key ESG metrics derived from corporate financial statements have been identified that could better support green finance monitoring and central banks' policy decisions (IFC (2021b)). The integration of these data is therefore not merely a reporting exercise, as it has profound implications for informing public decision-making with meaningful macro analyses. For instance, research conducted at the University of East London shows that green finance can significantly influence national patterns in capital flows, reshaping traditional lending relationships between banks and the non-financial sector.

[13] Anecdotally, the "non-financial reporting directive" (2014/95/EU) in the EU has been replaced by the CSRD, highlighting the central place of sustainability issues in the broader context of financial reporting.

[14] See EU Delegated Regulation 2023/2772 (eur-lex.europa.eu) and ESMA (2024).

3. Prospects offered by alternative data sources and innovative techniques

New information sources

While corporate financial reports can provide a robust foundation for firm-level analysis, they **can be usefully complemented by statistical and alternative data sources that are increasingly available with the ongoing information revolution**. Two fundamental contributions are worth mentioning here: (i) delivering deeper insights on corporate activities that are not explicitly captured by "traditional" financial statements; and (ii) enhancing statistical coverage and quality.

Regarding the first contribution, the **integration of various information sources can help bridge gaps in conventional financial reporting**. A telling case relates to the measurement of intangible assets such as data, software and other research and development investments which are typically difficult to estimate, despite their significant contribution to productivity and economic growth especially in knowledge-intensive sectors (Corrado et al (2025)). One solution is to combine transaction-level data on such assets with the more standard financial metrics derived from financial statements. Another option, explored at Katholieke Universiteit Leuven, is to integrate information on firm-level intellectual property rights. The university's findings suggest that linking data sets on patent registrations, trademarks and other intellectual property indicators with financial reports can provide useful insights on the contribution of innovation to economic outcomes, depending on sectors and firm types. Finally, a project at the National Bank of Belgium and Katholieke Universiteit Leuven combined annual company financial accounts with transaction level data. The project underscored the significant role of intangible assets in productivity growth and the importance of capturing these investments for a more accurate understanding of economic dynamics.

Turning to the second contribution, **the integration of more diverse data sources can be instrumental in enhancing the coverage and quality of statistical indicators** (IFC (2025b)). For instance, a project at the Bank of Thailand has been combining data from financial statements with statistical sources, such as monetary and financial statistics and international investment position data, to create more robust sectoral balance sheet estimates. This integration was able to improve the coverage, granularity and timeliness of the statistics produced.

New data techniques

Developments in big data analytics, AI and machine learning (ML) have allowed expanding the scope and depth of insights that can be extracted from financial statement data, along three dimensions: making data more accessible, making them more useful and improving their quality.

The first important dimension relates to the **accessibility of information**, which can be greatly improved by AI.[15] Perhaps the most significant development observed in recent years involves natural language processing techniques that can extract structured insights from unstructured textual disclosures (IFC (2025c)). Hence, one can more easily unlock qualitative information that was previously available in financial reports but was underused – including critical discussions of firm-level risks, opportunities and strategic plans. For instance, a promising avenue explored at the BIS involves the application of new AI-based tools, especially LLMs, to support text analytics and systematically process qualitative discussions contained in financial reports. In a similar vein, a recent project conducted at the Frankfurt School of Finance & Management examined how ML models can generate sustainability data from published reports, offering insights into corporate environmental performance even when explicit ESG disclosures are limited or inconsistent.

Second, **innovative analytical capabilities can improve the usefulness of financial statement data to support analysis and policy work**. For instance, LLMs can be particularly valuable for identifying emerging risks that lack established quantitative metrics but can thereby be more systematically monitored across large populations of firms. Central banks and regulators are indeed increasingly deploying such tools to conduct sophisticated analysis and derive insights that would be impossible to obtain through more conventional methods (IFC (2021a)). Moreover, as LLMs continue to advance rapidly, their ability to capture context, detect subtle signals and extract targeted information from complex financial narratives will certainly help to make better use of financial statements. Here also, a revealing example relates to ESG data, for which the new techniques can support financial supervisors and investors in better assessing the associated risks, bridging the gap between qualitative sustainability disclosures and quantitative risk assessment exercises (BISIH (2024)).

Third, **innovation can help address the important quality issues often characterising financial statement data**. In fact, new scalable data management techniques and sophisticated modelling tools appear well suited for dealing with reports disclosed by a large number of firms or across diverse accounting frameworks. For instance, one project at the Bank Al-Maghrib aims to integrate advanced ML techniques to enhance the accuracy, scalability and efficiency of anomaly detection systems. By automatically identifying potential quality issues across large data sets, this approach enables human expertise to focus on the most problematic cases, improving both the efficiency and the reliability of data processing. Such a hybrid approach, where human expertise is augmented rather than replaced by computational methods, can also be valuable in financial supervision (Bell et al (2025)).

[15] One important related development relates to digital reporting, such as "marking up" specific data points within a corporate report, which can offer significant benefits for the processing and analysis of granular data. This calls in particular for expanding the use of AI in financial statements or incorporating it into ESG reporting.

4. Leveraging financial statement information to support official statistics and policymaking

The insights derived from corporate reports have proved to be particularly useful for public authorities in general. **This is particularly the case regarding the wide range of activities performed by central banks**, including the compilation of statistics, the conduct of economic analysis, the designing of policy actions and their impact monitoring.

First, **financial statement data can greatly help central banks in their role as producers of official statistics**. A prime example is climate change statistics, for which central banks and other statistical authorities have been actively developing new sets of indicators to better inform decision-making on the imminent risks posed by environmental hazards (UNECE (2025)). In particular, linking data from corporate reports with other sources, such as administrative records, geospatial information and credit registers, appears to be instrumental to accurately map physical and transition risks and estimate banks' exposure to the most vulnerable firms (Aurouet et al (2023)). More broadly, granular data, such as firm-level reports, can help developing new (mostly experimental) statistics, allowing to "zoom in" on specific developments without losing sight of the broader macro perspective (Israël and Tissot (2021)). Such enhanced "statistical agility" can play an important role during periods of uncertainty, such as financial, geopolitical or health crises (de Beer and Tissot (2020), IFC (2025b)).

Second, the **granularity of financial statement data can support a wide range of central banks' analytical tasks**, which often require in-depth understanding of economic and financial developments. For one, corporate reports can feed into various economic analyses[16] They can also offer rich insights for macroeconomic modelling, which is certainly an important input for the conduct of their policies. Another key advantage is that firm-level data can document the behavioural changes of firms, an information that is often needed to accurately assess economic agents' reactions to shocks and have a better grasp of underlying developments (IFC (2024a)). In addition, sophisticated techniques can increasingly be deployed to make better use of the wealth of granular data available and substantially improve analysis. For instance, a recent project at the Bank of Korea analysed how the identification of distressed firms can be enhanced by examining the financial statements of listed companies through multiple lenses. It shows that several alternative indicators, such as return on operating assets, can offer robust insights on corporate distress than conventional metrics, such as the "traditional" interest coverage ratio.

Third, **financial statement data are increasingly being used for designing central banks' policies** (IFC (2024a)). For example, in the area of financial stability, firm-level inputs can be key for enabling central banks to detect early warning signals and take timely preventive measures before vulnerabilities escalate into systemic risk. Illustrating this point, the global community of central banks and financial supervisors has set up the BIS-hosted International Data Hub to compile, store and analyse confidential credit, funding and balance sheet data from a set of large global financial

[16] See Davis (2024) for an example of those related to firm-level price dynamics during the Covid-19 pandemic.

institutions complemented by granular information on their exposures to key counterparties (including non-financial firms) across national financial systems, sectors and markets (Bese Goksu and Tissot (2018)).

Finally, financial statement data can enhance the **effectiveness of policy decisions, by helping to better monitor and evaluate their impacts**, not only on macroeconomic aggregates but also on the distribution of those aggregates. For instance, recent research at the Bank of France shows how corporate debt structures can influence monetary policy transmission mechanisms. This reflects the fact that a tightening in bond liquidity tends to increase interest rate spreads and reduce issuance activity, while higher policy rates typically lead more directly to lower business investment overall (Alder et al (2023)). Such a granular understanding of transmission channels can enable a more precise calibration of monetary interventions. Similarly, financial statement data can provide important insights for the implementation of effective fiscal and regulatory policies, as shown in a recent study conducted at the Vorarlberg University of Applied Sciences analysing the impact of late payment regulations on European firms. Another fiscal example relates to the way specific types of firms might respond differently to public direct transfers versus subsidised credit guarantees.

5. Enhancing the information provided by financial statements: the way forward

While financial statement data can provide essential insights, **important challenges are associated with their use, especially for policy support.** Addressing these challenges properly calls for pursuing efforts in three main areas, namely (i) data quality, (ii) data access and sharing and (iii) responsible innovation, and various action points can be considered in this context looking forward (Table 1).

Improving data quality

One key challenge posed by firm-level data is quality, especially in terms of both accuracy and comparability. The vast and growing universe of financial statement data does not always follow clear, rigorous and robust quality standards which are used, for example, for producing official statistical information. As a result, the integration of these data, especially new, "alternative indicators", with "more traditional" statistics requires further processing, management and checking (IFC (2025b)). Another issue relates to the scarce comparability and consistency of some types of corporate reports across jurisdictions and over time, for example sustainability reports. This is notably explained by the lack of globally harmonised standards and mandatory (public) disclosures.

Yet **improving the quality of corporate reports is often a prerequisite to developing meaningful analysis** in increasingly complex global markets, where policy needs require granular insights that are sufficiently harmonised across jurisdictions. An obvious reason is that more comparable reports can reduce instances of information asymmetry among market participants, fostering better investment decisions and facilitating effective policy monitoring.

Fortunately, **a number of solutions can further enhance the quality of information** contained in financial statement data. A first option is the use of innovative data techniques, such as big data analytics and AI, which can greatly support large-scale data processing and anomaly detection (see Table 1). A second action may be to engage external stakeholders to help companies better understand and prioritise certain issues, such as sustainability. A third solution involves adopting and adequately enforcing regulatory measures to support better harmonisation in terms of disclosure practices. A case in point relates to ESG indicators, where the rapid increase in data availability has not been matched by improvements in quality, standardisation and comparability across firms and reporting regimes – particularly regarding the type of assurance engagement conducted on sustainability disclosures, as underscored by the Central Bank of the Republic of Austria. Fortunately, recent initiatives such as the European CSRD are expected to improve these assurance levels and external auditing, enabling users to differentiate between levels of external verification when conducting analyses.[17]

Strengthening data accessibility and sharing

Obviously, tension exists between the need for analysing firm-level financial statement data that can be confidential and the practical limitations to accessing and sharing them. These barriers may prevent **making full use of the richness of the granular corporate information which often is not published**. Finding ways to tap into these data while also safeguarding their sensitivity is thus an essential prerequisite to unlocking their full potential, combining them with complementary data sources (IFC (2025b)).

A first step is to improve general data accessibility and timeliness. Public availability of financial statement data is often restricted to national publication portals or registers. Authorities could take concrete measures to address this issue, as seen with the introduction of the European Single Access Point initiative to broaden the accessibility of published corporate reports at the European level (IFC (2025e)). Other, non-published information collected for policy purposes (such as the one compiled by CBSOs) could also be made more widely available, for instance through stronger partnerships with academic researchers.[18] Relatedly, these efforts can be complemented by the use of innovative techniques, such as privacy-enhancing technologies, which enable sharing the data without disclosing sensitive information related to individual units (Araujo et al (2025)). Finally, the timeliness of financial statement data remains an issue. Limiting the reporting lags could be an option, such

[17] In order to simplify sustainability reporting requirements, the European Commission published in early 2025 a dedicated package of "Omnibus proposals". They aim to narrow entity scope, promote voluntary reporting and simplify a number of requirements in terms of data accessibility introduced by the CSRD. However, it remains to be seen which consequences the revised Directive and the respective European Sustainability Reporting Standards will have on future corporate data in the EU.

[18] Cf the initiatives pursued by INEXDA – the International Network for Exchanging Experience on Statistical Handling of Granular Data – comprising a group of central banks, national statistical offices and international organisations (IFC (2018)).

as in the case of the EU accounting directive allowing for publication delays only up to a year.

A second step is to further strengthen the global statistical infrastructure, especially in terms of adoption of global identifiers, interoperability and information standards for microdata (IFC (2023, 2025e)). Fortunately, a number of international initiatives are under way on these three fronts. Regarding the development of global and unique identifiers, key international regulators have been endorsing the Legal Entity Identifier (LEI) in the aftermath of the GFC as a global common code to uniquely identify legal entities involved in financial transactions (FSB (2024)).[19] Its widespread adoption worldwide represents a fundamental achievement to facilitate the linking of corporate reports with other financial data sets. Turning to reporting formats, interoperability among information standards can be instrumental to further integrate firm-level reports – often using eXtensible Business Reporting Language (XBRL) – with statistical data which rely on standards such as the Statistical Data and Metadata eXchange (SDMX) and the Data Documentation Initiative (DDI) (UNECE (2024a)). Concrete initiatives relate to the development of converters, such as the one between XBRL and SDMX to facilitate comparability and information-sharing, although these solutions might increase the overall complexity. Finally, the need for greater interoperability emphasises the importance of undertaking further efforts to standardise micro data at the international level – a key objective pursued under the third phase of the Data Gaps Initiative endorsed by the G20 (IMF et al (2023)). Related to this, SDMX has evolved to handle micro and geospatial data with its version 3.0, unlocking new possibilities to handle both disaggregated, firm-level records and aggregated statistical data consistently (Nikoloutsos and Sirello (2023)).

In a third step, the proliferation of data sources and techniques, combined with the increasing complexity of the financial system, calls for **further strengthening and structuring of the exchange of information between central banks, regulators, other public authorities and stakeholders**. Central banks' attention has in particular focused on expanding the sharing of micro-level data, for example by setting up remote execution systems[20] and on-site infrastructures to facilitate user access to granular information in a secure environment (IFC (2025c)). Further, coordination could take the form of data-sharing agreements between central banks and other public authorities within national statistical systems. A concrete and successful initiative in this regard has been the creation of Large Case Units in a number of jurisdictions. These operate as dedicated organisational units that track the complex activities of multinational enterprises by using a wide range of data sources, including corporate statements, tax reports and statistical surveys (CFMB (2024), UNECE (2024b)). In any case, central banks' experience with sharing granular information underscores the importance of carefully balancing benefits and risks, especially regarding the need to safeguard sensitive data – a fundamental aspect that calls for sound governance practices, robust IT security measures and professional independence.

[19] Global common identifiers feature several benefits for statistical purposes, especially to link data sources. Yet adoption may be voluntary, hampering their usefulness (CMFB (2023b), IFC (2025b)).

[20] These systems allow users to perform calculations, typically on confidential micro data sets, without directly accessing the underlying data.

Supporting responsible innovation

The **last important challenge concerns the use of AI and big data techniques** to support the use of financial statement data for various analytical and statistical activities. In this perspective, central banks' experience suggests that specific attention should be put on the implementation and governance of these innovative techniques (IFC (2025d)).

Regarding **implementation**, there are important deployment barriers often related to computational requirements required for dealing with the vast amount of financial statement data as well as to high initial and maintenance costs (IFC (2020c)). A growing issue is the adoption of cloud services, which can help contain infrastructure costs but may also expose to third-party dependencies and legal challenges, especially when the cloud providers are located in foreign jurisdictions. Another major challenge is related to skills shortages, in particular due to the relative scarcity of data scientists and engineers, in turn calling for upskilling strategies internally and strengthening the organisation's attractiveness externally.

Turning to **governance** aspects, innovation may also raise various questions, especially in terms of security, quality management, risks and ethical considerations when dealing with firm-level data. For example, one challenge while using LLMs to extract meaningful insights from textual documents is often the lack of adequate quality frameworks, constraining the detection and quantification of possible data, algorithmic or cultural biases. Generative AI solutions also pose new types of risks and threats, for example by contaminating high-quality data with unverified or unofficial information – an issue that can ultimately mislead investors, policymakers and the public at large.

Against this background, **maintaining the integrity and credibility of and trust in corporate disclosures and in turn in the use of financial statement data for official statistics and public policy is essential** (IFC (2025d)). Fortunately, central banks' long-standing expertise in data governance and management makes them well rounded to play the role of "data curators" in the broader data ecosystem, including by safeguarding and securing the quality of financial information (Križman and Tissot (2022)).

A roadmap for enhancing the use of financial statements for supporting central bank policies	Table 1

Key focus area	Suggested measures
Accessibility	▪ Leverage innovative data techniques (eg to automate data extraction from reports, structure unstructured text format through AI) and alternative sources of information ▪ Develop single access points (eg integrated publication portals or repositories) ▪ Enable search and filtering capabilities to make better use of financial statement databases ▪ Strengthen partnerships with academia and private sector
Data-sharing	▪ Support adequate data-sharing across interested stakeholders, including central banks, statistical offices and academia (eg through memoranda of understanding covering aspects such as confidentiality, quality and security) ▪ Promote the use of global unique identifiers to unequivocally identify entities (eg LEI) ▪ Establish research data centres or secure online environments to access sensitive data, including by researchers ▪ Develop better access to administrative sources, including without appropriate legal gateways ▪ Ensure proper data licensing and compliance (eg copyright)
Data quality	▪ Increase comparability through adequate harmonisation of reporting standards and disclosure requirements ▪ Leverage new techniques to support large-scale editing processes (ie anomaly detection) ▪ Deploy data quality management approaches (eg cross-checking between databases/institutions) to improve data completeness and accuracy ▪ Promote consistent data formats and standards interoperability (eg XBRL-SDMX converters) ▪ Establish ESG assurance levels for auditors and materiality assessment exercises
Transparency	▪ Improve existing databases (eg uncertainty measures and ratio definitions) ▪ Increase transparency in the use of new data sources and AI models

References

Alder, M, N Coimbra and U Szczerbowicz (2023): "Corporate debt structure and heterogeneous monetary policy transmission", *Banque de France Working Paper*, no 933, December.

Araujo, D, A Cap, I Mattei, R Schmidt, O Sirello and B Tissot (2025): "Data science in central banking: unlocking the potential of data", *IFC Bulletin*, no 64, May.

Arkoh, P, A Costantini and F Scarpa (2024): "Determinants of sustainability reporting: a systematic literature review", *Corporate Social Responsibility and Environmental Management*, vol 31, no 3, pp 1578–97.

Aurouet, D, A De Sanctis, D Giudice, 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*, International Monetary Fund, Chapter 4, April, pp 59–77.

Bank for International Settlements (BIS) (2025): "Financial conditions in a changing global financial system", *Annual Economic Report 2025*, Chapter II, June.

Bank for International Settlements Innovation Hub (BISIH) (2024): *Project Gaia: enabling climate risk analysis using generative AI*, March.

Bell, S, B Gadanecz, L Gambacorta, F Perez-Cruz and V Shreeti (2025): "Artificial intelligence and human capital: challenges for central banks", *BIS Bulletin*, no 100, April.

Bese Goksu, E and B Tissot (2018): "Monitoring systemic institutions for the analysis of micro-macro linkages and network effects", *Journal of Mathematics and Statistical Science*, vol 4, no 4, pp 129–36.

Committee on Monetary, Financial and Balance of Payments Statistics (CFMB) (2023a): "Final report of the Task Force on the statistics on sustainable finance and climate related risks", *Working document*, April.

——— (2023b): "CMFB opinion on the Legal Entity Identifier (LEI) as unique identifier of financial and non-financial companies for statistical purposes", *CFMB Opinion*, May.

——— (2024): *CMFB vision on globalisation and statistics*, April.

Corrado, C, J Haskel, M Iommi, C Jona-Lasinio and F Bontadini (2025): "Data, intangible capital, and productivity" in S Basu, L Elridge, J Haltiwanger and E Strassner (eds), *Technology, productivity and economic growth*, University of Chicago Press, Chapter 7, July.

Davis, L (2024): "Profits and markups during the post-Covid-19 inflation shock in the US economy: a firm-level lens", *European Journal of Economic Policies: Intervention*, vol 21, no 2, pp 309–30.

de Beer, B and B Tissot (2020): "Implications of Covid-19 for official statistics: a central banking perspective", *IFC Working Papers*, no 20, November.

Deutsche Bundesbank (2014): "Consolidated financial statement statistics as a contribution to the extended corporate analysis: approach and initial results", *Deutsche Bundesbank Monthly Report*, July.

Diz Dias, J, M C Gómez Ramos, E Martín de Andrés, F Pastoris, M Schmitz, O Sirello and B Tissot (2025): "External statistics in a fragmented and uncertain world: addressing novel analytical challenges", *Statistical Journal of the IAOS*, vol 41, no 2, May.

European Committee of Central Balance Sheet Data Offices (ECCBSO) (2021): *ECCBSO Report 2021*, December.

European Securities and Markets Authority (ESMA) (2024): "Off to a good start: first application of ESRS by large issuers", *Public Statement*, no ESMA32-992851010-1597.

Fernández Cerezo, A, B Fernández-Rosillo San Isidro and N Pérez Martín (2024): "The Banco de España's Central Balance Sheet Data Office database: a regional perspective", *Banco de España Occasional Papers*, no 2429, September.

Financial Stability Board (FSB) (2024): *Implementation of the legal entity identifier: progress report*, October.

IFRS Foundation (2023): *IFRS S1 General Sustainability Disclosures and IFRS S2 Climate-related disclosures*, June.

International Monetary Fund (IMF), Inter-Agency Group on Economic and Financial Statistics and Financial Stability Board Secretariat (2023): *G20 DGI-3 Workplan – people planet economy,* March.

Irving Fisher Committee on Central Bank Statistics (IFC) (2014): "Integrated management of micro-databases", *IFC Bulletin*, no 37, January.

——— (2017): "Uses of central balance sheet data offices' information", *IFC Bulletin*, no 45, October.

——— (2018): "INEXDA – the granular data network", *IFC Working Papers*, no 18, October.

——— (2020a): "Using financial accounts", *IFC Bulletin*, no 51, February.

——— (2020b): "Bridging measurement challenges and analytical needs of external statistics: evolution or revolution?", *IFC Bulletin*, no 52, October.

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

——— (2021a): "Issues in data governance", *IFC Bulletin*, no 54, July.

——— (2021b): "Sustainable finance data for central banks", *IFC Report*, no 14, December.

——— (2021c): "Machine learning in central banking", *IFC Bulletin*, no 57, November.

——— (2022): "Statistics for sustainable finance", *IFC Bulletin*, no 56, May.

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

——— (2024a): "Granular data: new horizons and challenges", *IFC Bulletin*, no 61, July.

——— (2024b): "Empowering carbon accounting: from data to action", *IFC Report*, no 16, October.

——— (2025a): "Addressing climate change data needs: the central banks' contribution", *IFC Bulletin*, no 63, March.

——— (2025b): "Data integration: opportunities and challenges for central banks", *IFC Guidance Note*, no 6, July.

——— (2025c): "Data science in central banking: enhancing the access to and sharing of data", *IFC Bulletin*, no 64, May.

——— (2025d): "Governance and implementation of artificial intelligence in central banks", *IFC Report*, no 18, April.

——— (2025e): "SDMX adoption and use of open source tools", *IFC Report*, no 17, February.

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

Jahangir-Abdoelrahman, S and B Tissot (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.

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

Ortega, M (2014): "ECCBSO: a tool for the analysis of the non-financial sector of the economy (from individual to consolidated accounts)", *IFC Bulletin*, no 37, January.

Rejison, P (2025): "Identifying the importance of financial statements in strategic decision making", *SSRN*, January.

Rosolia, A, S Stapel-Weber and B Tissot (2021): "New developments in central bank statistics around the world", *Statistical Journal of the IAOS*, vol 37, no 4, pp 1055–60.

United Nations (UN) (2014): "Resolution adopted by the General Assembly on 29 January 2014: 68/261 – Fundamental Principles of Official Statistics", 29 January.

United Nations Economic Commission for Europe (UNECE) (2024a): *Data governance framework for statistical interoperability (DAFI)*, High-Level Group for the Modernisation of Official Statistics, March.

——— (2024b): "Large Cases Units in EU and EFTA countries", Group of Experts on National Accounts, ECE/CES/GE.20/2024/26, April.

——— (2025): *Data for climate action: how national statistical offices can contribute*, January.

Van Hulle, K and L van der Tas (2000): "European Union – individual accounts" and "European Union – group accounts", in D Ordelheide (ed), *Transnational Accounting*, 2nd edition, vol 1, Palgrave Macmillan London, pp 771–877 and 879–951.

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Outlier detection optimization using machine learning for improving data quality[1]

Salah Ahl Mbarek,
Bank Al-Maghrib (Central Bank of Morocco)

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

بنك المغرب
BANK AL-MAGHRIB

Topic

Outlier Detection Optimization using Machine Learning for improving Data Quality

Abstract

This paper explores the implementation of advanced machine learning techniques to optimize outlier detection for enhancing data quality within the financial sector at Bank Al Maghrib. The primary focus was on developing and integrating robust algorithms capable of identifying anomalies in extensive datasets of annual financial statements of Moroccan companies. We began by thoroughly examining existing methodologies and their limitations, which informed the selection of cutting-edge techniques such as Isolation Forest, Mahalanobis Distance, DBSCAN, HDBSCAN, and Adversarially Learned Anomaly Detection (ALAD), along with state-of-the-art reinforcement methods like ML-consensus achieved through Stacking and Meta-learning.

The selected algorithms were meticulously analyzed and integrated into an ensemble model to leverage their individual strengths, thereby improving overall anomaly detection accuracy. The implementation phase involved extensive preprocessing and feature engineering to ensure data integrity and relevance. Rigorous testing and validation procedures confirmed the model's efficacy and reliability, demonstrating significant improvements in outlier detection. These improvements contribute to enhanced financial data quality, ensuring more accurate and reliable insights for decision-making.

Additionally, the solution's deployment included a user-friendly interface for analysts to interact with the model, visualize results, and make informed decisions. Comprehensive documentation and training sessions ensured a smooth transition and knowledge transfer to the in-house team, guaranteeing the solution's sustainability.

The paper concludes with several promising avenues for future work, including algorithm refinement, real-time anomaly detection, scalability and performance optimization, integration with other financial systems, expansion to other domains, and continuous learning and adaptation. These perspectives aim to enhance the model's capabilities and ensure its continued relevance in addressing evolving challenges in data quality management.

By harnessing the power of advanced algorithms and ensuring their practical applicability within Bank Al Maghrib, this report lays a robust foundation for future advancements in the field. The insights and results not only contribute to the existing body of knowledge but also provide a practical framework for ongoing innovation and improvement in data quality management.

Keywords : Ensemble Model, Isolation Forest, Mahalanobis Distance, Generative Adversarial Networks (GANs), Adversarially Learned Anomaly Detection (ALAD), Meta-Learning.

Acronymes

	Acronym	Meaning
1	BKAM	Bank Al Maghrib
2	CSE	Casablanca Stock Exchange
3	ML	Machine Learning
4	DL	Deep Learning
5	SVM	Support Vector Machine
6	CRISP-DM	Cross-Industry Standard Process for Data Mining
7	EDA	Exploratory Data Analysis
8	GDPR	General Data Protection Regulation
9	DBSCAN	Density-Based Spatial Clustering of Applications with Noise
10	HDBSCAN	Hierarchical DBSCAN
11	ALAD	Adversarially Learned Anomaly Detection
12	UI	User Interface
13	UML	Unified Modeling Language
14	BDD	Behavioral-Driven Development
15	SEI	Software Engineering Institute

List of Figures

1.1 Logo of the Organization 14
1.2 Organizational Chart 15
1.3 Key Axes of the Data and Statistics Strategy 20

2.1 CRISP-DM Chart 31
2.2 Gantt Diagram 32

4.1 Use Case Diagram 65
4.2 Class Diagram 66
4.3 Sequence Diagram 67
4.4 Activity Diagram 68
4.5 Workflow 70
4.6 Execution Time Comparison 74
4.7 Accuracy Comparison 76
4.8 Precision Comparison 76
4.9 Recall Comparison 76
4.10 F1-Score Comparison 76
4.11 Execution Time Comparison 77
4.12 Accuracy Comparison 78
4.13 Precision Comparison 78
4.14 Recall Comparison 78
4.15 F1-Score Comparison 78
4.16 Area Under the Curve Variation for Models 79
4.17 Training Time Comparison of Deep Learning-Based Methods 80
4.18 Accuracy Comparison 80
4.19 Precision Comparison 80
4.20 Recall Comparison 81
4.21 F1-Score Comparison 81
4.22 AUC Comparison of Deep Learning-Based Methods 81
4.23 Loss Function Evolution of Autoencoders for varying number of data entries 82
4.24 Accuracy Variation by Model 85
4.25 Prediction Time Variation by Model 85

5.1 Development Tools 88
5.2 Programming Language 88
5.3 Application Development Tools 88
5.4 Model Creation Tools 88
5.5 Testing Tools: Jest and Cucumber 88
5.6 Version Control: Git and GitHub 89
5.7 Form Interface 90
5.8 App passing unitary and integration tests 92

5.9 App passing Behavioral Driven tests....93
5.10 User documentation detailing installation and usage instructions. 94
5.11 Developer documentation showcasing application architecture and codebase details. 95

List of Algorithms

1 Synthetic Dataset Generation 73
2 Stacking for Anomaly Detection 84

Contents

Abstract **3**

Acronymes **4**

General introduction **10**

1 Context **13**
1.0 Introduction to chapter 13
1.1 Presentation of the hosting organization 14
1.2 Framework of Monetary Policy 15
1.3 General context 19
1.4 Thematic 21
1.5 Conclusion 23

2 Problem Analysis **24**
2.0 Introduction to chapter 24
2.1 Problem Statement 25
2.2 Critique of Existing Methods 26
2.3 Solution Proposal 27
2.4 Requirements Specification 28
2.5 Methodology of work – CRISP-DM 30
2.6 Project Planning 31
2.7 Conclusion 32

3 Theoretical Background **33**
3.0 Introduction to chapter 33
3.1 Background History 35
3.2 Distance-Based Approaches 37
3.3 Density-Based Approaches 41
3.4 Machine Learning-Based Approaches 46
3.5 Deep Learning-Based Approaches 50
3.6 ML-Consensus 53
3.7 Data Synthesis 59
3.8 Conclusion 61

4 Solution Engineering **63**
4.0 Introduction to chapter 63
4.1 Solution Architecture 64

4.2 Data Synthesis 71
4.3 Solution Analysis 73
4.4 Optimisations and/or Alternatives 83
4.5 Conclusion 85

5 Implementation of the Solution 87
5.0 Introduction to chapter 87
5.1 Used Technologies 88
5.2 The desktop App 89
5.3 Quality Assurance and Testing 91
5.4 Documentation 93
5.5 Conclusion 96

Conclusions and Perspectives 98

Bibliography 102

General introduction

Background

In the modern financial landscape, data accuracy and integrity are paramount, especially for institutions like the Central Bank of Morocco (Bank Al Maghrib) that rely on precise financial data to inform policy decisions and maintain economic stability. Financial data quality directly impacts the credibility and reliability of financial statements, risk assessments, and economic forecasts. One significant challenge in maintaining high data quality is the presence of outliers—data points that deviate significantly from other observations. These outliers can result from errors in data entry, fraud, or unusual but legitimate transactions, and can distort analyses, leading to incorrect conclusions and potentially harmful financial decisions.

In Morocco, the annual financial statements of companies provide critical information for economic analysis and policy-making. Ensuring the accuracy of these financial statements is essential for regulatory oversight, investor confidence, and overall economic health. The traditional process of identifying and correcting outliers in these financial statements is labor-intensive and subject to human error. Manual re-evaluation of financial figures is not only time-consuming but also prone to inconsistencies. As the volume and complexity of financial data continue to grow, there is a pressing need for an automated, robust, and accurate method for outlier detection.

Problem Statement

The Central Bank of Morocco faces significant challenges in maintaining the accuracy and integrity of financial data reported by Moroccan companies. The current manual methods for re-evaluating financial statements to identify and correct outliers are inefficient and error-prone. These methods cannot keep pace with the increasing volume and complexity of data, leading to potential inaccuracies in financial reporting and analysis. This thesis addresses the need for an automated approach to outlier detection that can efficiently and accurately identify anomalies in financial data, thereby improving the overall data quality.

Objectives

This thesis aims to develop and optimize machine learning models for the automated detection of outliers in the annual financial statements of Moroccan companies. The primary objectives of the study are:

- **To Review Existing Outlier Detection Techniques**

  - Conduct a comprehensive review of current outlier detection methods, including statistical approaches and machine learning techniques.
  - Identify the strengths and limitations of these methods in the context of financial data.
- **To Develop Machine Learning Models for Outlier Detection**
  - Implement and evaluate various machine learning algorithms such as Isolation Forest, and other relevant models.
  - Compare the performance of these models in terms of accuracy, efficiency, and scalability.
- **To Optimize the Performance of Machine Learning Models**
  - Apply optimization techniques such as hyperparameter tuning, ensemble methods, and feature engineering to enhance model performance.
  - Evaluate the impact of these optimizations on the accuracy and reliability of outlier detection.
- **To Integrate the Optimized Models into the Central Bank's Data Processing Pipeline**
  - Develop a practical framework or internal tool (Software) for implementing the optimized outlier detection models within the existing data infrastructure at the Central Bank of Morocco.
  - Ensure the solution is scalable, efficient, and capable of handling large volumes of financial data.

Importance of Study

Enhancing the process of outlier detection in financial data has far-reaching implications for the Central Bank of Morocco and the broader financial ecosystem. Improved data quality leads to better decision-making, more accurate financial analyses, and more effective regulatory oversight. By automating and perfecting the re-evaluation of financial data, this study contributes to the reliability and transparency of financial reporting, thereby supporting the overall stability and integrity of the Moroccan financial system.

Automated outlier detection will not only reduce the workload on analysts but also provide a consistent and reliable method for identifying anomalies. This, in turn, enhances the credibility of financial statements and boosts confidence among investors, regulators, and other stakeholders.

Structure of the Report

The thesis is structured as follows:

- **Chapter 1: Context** This chapter Provides an overview of the hosting organization, Bank Al Maghrib, and the monetary policy framework in Morocco.
- **Chapter 2: Problem Analysis** Discusses the problem statement, current methodologies, their limitations, and the proposed solution along with detailed specifications and work methodology.
- **Chapter 3: Theoretical Background** Reviews the existing literature and theories related to outlier detection and cutting-edge techniques in the field.

- **Chapter 4: Solution Engineering** Details the data collection and synthesis, preprocessing, model development, optimization techniques, and evaluation metrics.
- **Chapter 5: Implementation of the Solution** Describes the practical implementation of the models, including user interfaces, model training and testing, quality assurance, documentation, and deployment.

Finally, the thesis concludes with a summary of findings, implications, and suggestions for future work, followed by a comprehensive bibliography and annexes.

By the end of this study, we aim to provide a comprehensive solution for the automated detection of outliers in financial data, enhancing the Central Bank of Morocco's capability to ensure data quality and make informed decisions based on accurate financial information.

Chapter 1

Context

This chapter sets the foundation for the entire study by providing a comprehensive overview of Bank Al Maghrib. Understanding the organizational structure, functions, and missions of Bank Al Maghrib is crucial, as it is the central authority responsible for monetary policy and financial stability in Morocco. We will explore the general description of the institution, its roles, and its organizational hierarchy to understand how it operates and how decisions are made.

Furthermore, the chapter delves into the framework of monetary policy in Morocco, which includes an analysis of the financial market, the Moroccan financial system, and the monetary policy mechanisms. This detailed context is essential for comprehending the environment in which the study is conducted and highlights the importance of maintaining high data quality in financial reporting. By understanding the intricacies of monetary policy and the financial system, we can appreciate the critical role that accurate financial data plays in informing policy decisions and ensuring economic stability.

This chapter's content is directly linked to the problem analysis in the following chapter, where we will discuss the challenges associated with data quality and outlier detection in financial datasets. By establishing the context, we provide a background that underscores the relevance and necessity of our study, ensuring that the reader comprehends the foundational elements that drive the need for improved outlier detection methods.

1.1 Presentation of the hosting organization

1.1.1 Overview



Figure 1.1 – Logo of the Organization

Bank Al Maghrib, the Central Bank of Morocco, is a pivotal institution in the Moroccan financial system. Established in 1959, its primary mission is to ensure price stability and the soundness of the financial system. As a central authority, it regulates and oversees the banking sector, implements monetary policy, and manages the country's foreign exchange reserves. The bank also plays a critical role in formulating and executing policies that promote sustainable economic growth and financial inclusion.

1.1.2 Functions and Missions

Bank Al Maghrib's functions and missions are diverse and encompass several key areas:

Monetary Policy Implementation

The bank designs and implements monetary policy to maintain price stability and control inflation. It uses various instruments such as interest rate adjustments, open market operations, and reserve requirements to influence money supply and demand.

Financial Stability

Ensuring the stability of the financial system is a core mandate. The bank monitors and assesses systemic risks, supervises financial institutions, and enforces prudential regulations to prevent financial crises.

Banking Supervision and Regulation

Bank Al Maghrib oversees the banking sector to ensure it operates efficiently and safely. This involves licensing new banks, conducting regular inspections, and ensuring compliance with regulatory standards.

Foreign Exchange Management

The bank manages Morocco's foreign exchange reserves, conducts foreign exchange operations, and maintains the stability of the national currency, the Moroccan Dirham.

Economic Research and Statistics

Conducting economic research and providing reliable statistical data are essential functions. The bank analyzes economic trends, publishes reports, and offers insights to inform policy decisions and guide market participants.

Consumer Protection

Protecting the interests of consumers and promoting financial literacy are also important missions. The bank ensures transparency in financial services, addresses consumer complaints, and educates the public about financial products and services.

Payment Systems Oversight

Bank Al Maghrib oversees the payment systems to ensure their efficiency, security, and reliability. It regulates and monitors payment instruments and infrastructures to facilitate smooth and secure transactions.

1.1.3 Organizational Chart

The organizational chart of Bank Al-Maghrib provides an overview of the hierarchical structure and roles within the central bank. It illustrates the relationships and reporting lines among different departments and units, highlighting key positions and their responsibilities in the bank's operations and decision-making processes.

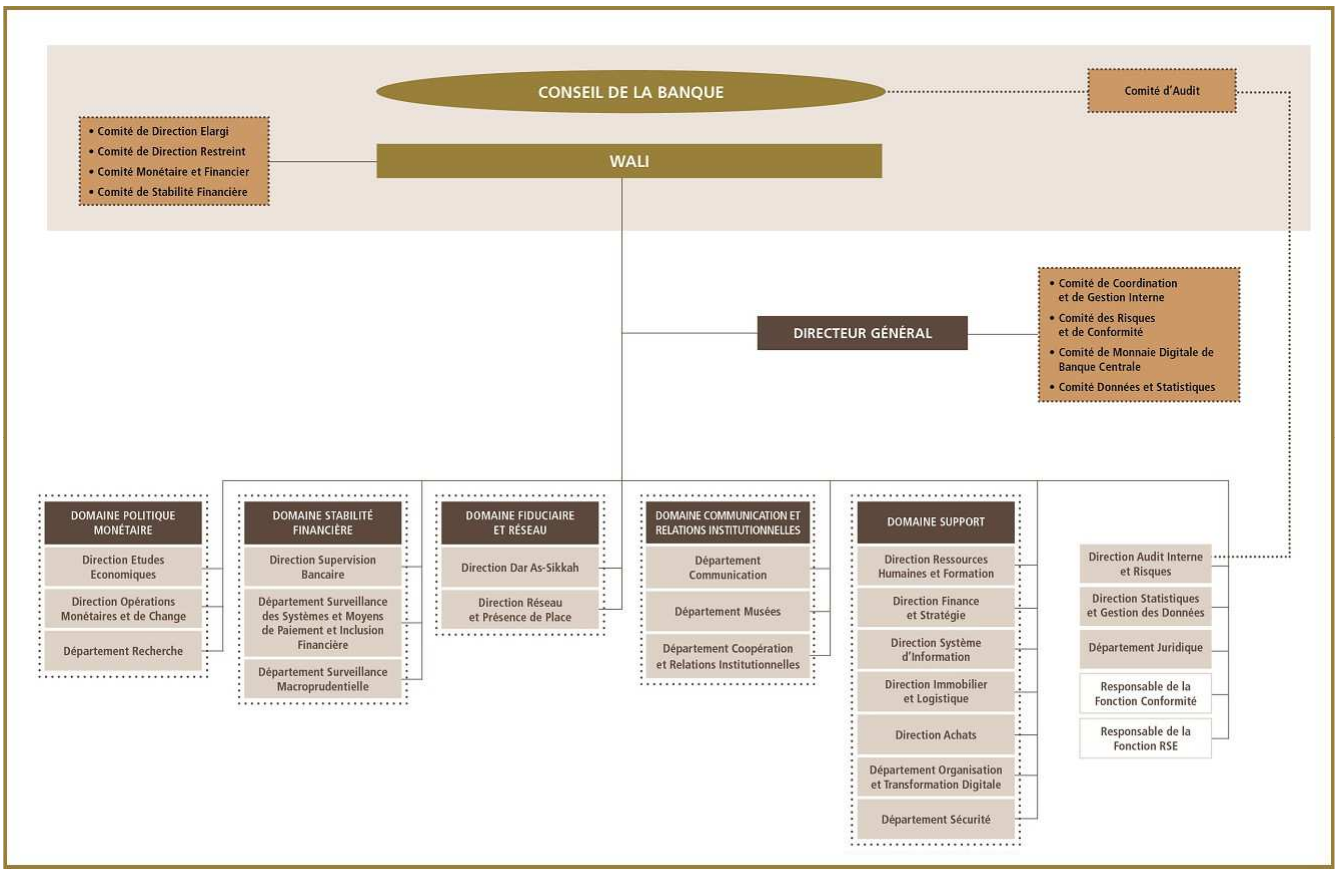


Figure 1.2 – Organizational Chart

This chart serves as a visual representation of the bank's organizational hierarchy, demonstrating how various departments collaborate and contribute to fulfilling the bank's mandates and objectives.

1.2 Framework of Monetary Policy

1.2.1 Financial Market

The financial market in Morocco plays a crucial role in the overall economic framework of the country. It encompasses a variety of sectors, including the stock market, bond market, and foreign exchange market. The primary stock exchange, the Casablanca Stock Exchange (CSE) – Bourse

de Casablanca, is one of the largest in Africa and serves as a vital platform for raising capital for companies. The bond market, meanwhile, provides a mechanism for the government and corporations to borrow funds through the issuance of debt securities.

The foreign exchange market in Morocco, regulated by Bank Al Maghrib, facilitates the exchange of the Moroccan Dirham with other currencies, thus supporting international trade and investment. This market is integral to maintaining the stability of the Dirham and managing the country's foreign exchange reserves. Moreover, the financial market in Morocco is characterized by a growing number of financial instruments and services, which are designed to meet the diverse needs of investors and borrowers.

Bank Al Maghrib plays a key role in overseeing and regulating these markets to ensure their stability, transparency, and efficiency. The central bank implements various policies and regulations to safeguard the interests of investors and maintain the overall health of the financial system.

1.2.2 Moroccan Financial System

The Moroccan financial system is characterized by a structured network of institutions including commercial banks, insurance companies, pension funds, and other financial intermediaries. Dominating this landscape are several prominent commercial banks offering a wide range of services such as deposits, loans, and payment processing. While the sector features some competition, it is primarily led by a few well-established domestic banks, supported by a regulatory framework overseen by Bank Al Maghrib.

Insurance companies and pension funds also play significant roles, providing critical services for risk management and retirement planning, enhancing financial security for individuals and businesses. Additionally, microfinance institutions and non-bank financial entities contribute to financial inclusion by serving underserved populations.

Bank Al Maghrib maintains the stability of the financial system through rigorous supervision and regulation. The central bank sets prudential standards, conducts regular inspections, and enforces compliance with regulatory requirements, thereby mitigating risks and ensuring a stable financial environment supportive of economic growth.

1.2.3 Monetary Policy

Monetary policy in Morocco is formulated and implemented by Bank Al Maghrib with the primary objective of maintaining price stability. The central bank employs a range of tools and instruments to achieve this goal, including interest rate adjustments, open market operations, and reserve requirements.

Interest rate policy is a cornerstone of monetary policy. By setting the key policy rate, Bank Al Maghrib influences short-term interest rates, which in turn affect borrowing and lending rates across the economy. This mechanism helps control inflation and support economic activity.

Open market operations involve the buying and selling of government securities in the open market. Through these operations, Bank Al Maghrib can manage the money supply and influence liquidity conditions in the banking system. This tool is used to smooth out fluctuations in interest rates and ensure that the banking system has adequate liquidity to meet the needs of the economy.

Reserve requirements are another important tool. By setting the minimum reserves that banks must hold, Bank Al Maghrib can influence the amount of money that banks can lend. This helps

regulate the money supply and control inflationary pressures.

In addition to these traditional tools, Bank Al Maghrib monitors various economic indicators and financial conditions to make informed policy decisions. The central bank's actions are guided by a commitment to transparency and accountability, with regular communication to the public and financial markets about policy decisions and their rationale.

1.2.4 Instruments of Monetary Policy

Bank Al Maghrib employs a range of monetary policy instruments to achieve its primary objective of price stability and to support the broader economic goals of the country. These instruments include:

Interest Rate Policy

The central bank sets the key policy rate, which influences short-term interest rates across the economy. Adjustments to this rate impact borrowing and lending rates, affecting economic activity and inflation. The policy rate serves as a benchmark for other interest rates within the banking system.

Open Market Operations

Bank Al Maghrib conducts open market operations by buying and selling government securities. These operations manage liquidity conditions in the banking system, influencing the money supply and short-term interest rates. By increasing or decreasing the amount of money in circulation, the central bank can smooth out fluctuations in interest rates and ensure sufficient liquidity for economic needs.

Reserve Requirements

The central bank sets minimum reserve requirements for commercial banks, determining the proportion of customer deposits that must be held as reserves. By adjusting these requirements, Bank Al Maghrib can influence the amount of money banks can lend, thereby controlling the money supply and exerting an impact on inflation and economic activity.

Standing Facilities

These facilities provide and absorb overnight liquidity to and from the banking system. The central bank offers lending facilities (e.g., overnight lending facilities) to banks facing short-term liquidity shortages, and deposit facilities to absorb excess liquidity. These facilities help maintain stability in the money market and ensure the smooth functioning of the banking system.

Foreign Exchange Interventions

To maintain the stability of the Moroccan Dirham, Bank Al Maghrib may intervene in the foreign exchange market. These interventions involve buying or selling foreign currencies to influence exchange rates and manage the country's foreign exchange reserves. The central bank's actions in

the foreign exchange market help mitigate excessive volatility and ensure a stable external value of the Dirham.

Communication and Forward Guidance

Transparency and effective communication are critical components of Bank Al Maghrib's monetary policy strategy. The central bank regularly communicates its policy decisions, economic outlook, and rationale for its actions to the public and financial markets. Forward guidance provides insights into the likely future path of monetary policy, helping to shape expectations and influence economic behavior.

1.2.5 Recent Monetary Policy Decisions and Their Impacts

In recent years, Bank Al Maghrib has made several key monetary policy decisions aimed at addressing the evolving economic challenges and maintaining price stability. These decisions and their impacts are summarized below:

Interest Rate Adjustments

In response to global economic uncertainties and domestic inflationary pressures, Bank Al Maghrib has adjusted the key policy rate several times. For instance, during periods of economic slowdown, the central bank has lowered the policy rate to stimulate borrowing and investment, supporting economic growth. Conversely, to counteract rising inflation, the policy rate has been increased to curb excessive demand and control price levels.

- **Impact**: Lower interest rates have facilitated increased access to credit for businesses and households, boosting consumption and investment. Higher rates have helped contain inflationary pressures and stabilize the economy.

Liquidity Management through Open Market Operations

Bank Al Maghrib has actively conducted open market operations to manage liquidity conditions in the banking system. By purchasing or selling government securities, the central bank has ensured adequate liquidity to support economic activity while preventing excessive money supply growth.

- **Impact**: Effective liquidity management has contributed to the stability of short-term interest rates, enhanced the functioning of the money market, and supported the banking system's ability to meet the economy's credit needs.

Foreign Exchange Interventions

To address external shocks and maintain exchange rate stability, Bank Al Maghrib has intervened in the foreign exchange market. These interventions have included buying or selling foreign currencies to manage exchange rate fluctuations and stabilize the Moroccan Dirham.

- **Impact**: Foreign exchange interventions have helped mitigate excessive volatility in the exchange rate, supported export competitiveness, and maintained confidence in the Dirham's stability.

Enhanced Communication and Forward Guidance

Bank Al Maghrib has strengthened its communication strategy by providing clear and transparent information on its policy decisions, economic assessments, and future policy intentions. This forward guidance has helped shape market expectations and enhance the effectiveness of monetary policy.

- **Impact**: Improved communication has reduced uncertainty, influenced economic agents' expectations, and facilitated better-informed decision-making by businesses and households.

Reserve Requirement Adjustments

In response to changing economic conditions, Bank Al Maghrib has adjusted the reserve requirements for commercial banks. These adjustments have been used to influence the amount of money banks can lend, thereby impacting credit conditions and money supply growth.

- **Impact**: Changes in reserve requirements have helped manage liquidity in the banking system, supported financial stability, and ensured that credit conditions align with the central bank's monetary policy objectives.

Overall, Bank Al Maghrib's recent monetary policy decisions have been instrumental in maintaining price stability, supporting economic growth, and ensuring the stability of the financial system. The central bank's proactive and adaptive approach has enabled it to navigate complex economic challenges and contribute to Morocco's economic development.

1.3 General context

The "Direction Statistiques et Gestion des Données" at Bank Al-Maghrib plays a pivotal role in enhancing the central bank's data collection and statistical production capabilities to support its policies and fulfill its mandates. Over recent years, significant advancements have been made in several key areas within the department:

1.3.1 Internal Developments

Bank Al-Maghrib has strengthened its information assets through various initiatives:

- **Enrichment of Data Sources**: Establishing frameworks for data exchange with key partners and expanding data collection to include macro-prudential frameworks, payment system surveillance, and financial inclusion.

- **Financial Information Centers**: Implementation of financial information centers covering data on monetary and exchange markets, bank accounts, check and bill defaults, as well as company balances and profiles.

- **New Indices**: Introduction of new indices such as the real estate asset price index and banking service price index.

- **Statistical Surveys**: Launch of new statistical surveys, including credit granting conditions and inflation expectations.

- **Modernization of Statistical Production**: Continuous upgrading of monetary statistics

production to ensure comprehensive coverage of institutional sectors in line with international standards.

1.3.2 Enhanced Data Utilization

The department has invested in operational systems to enhance data collection, dissemination tools, and diverse formats for data exchange. This includes:

- Deployment of statistical series warehouses and a decision support system.
- Establishment of a functional competence center to unify practices in data processing, calculation, indicator reporting, and statistical analysis, fostering knowledge sharing and enhancement.

1.3.3 Regulatory Framework and Centralized Data Management

Bank Al-Maghrib has implemented a governance framework for data governance, focusing on monetary policy, banking operations, and common interest services. Key initiatives include:

- Creation of a centralized entity for data management and statistical production, allowing business entities to focus on core activities.
- Consolidation of common data repositories and initial deployment of advanced data exploration and analysis capabilities.

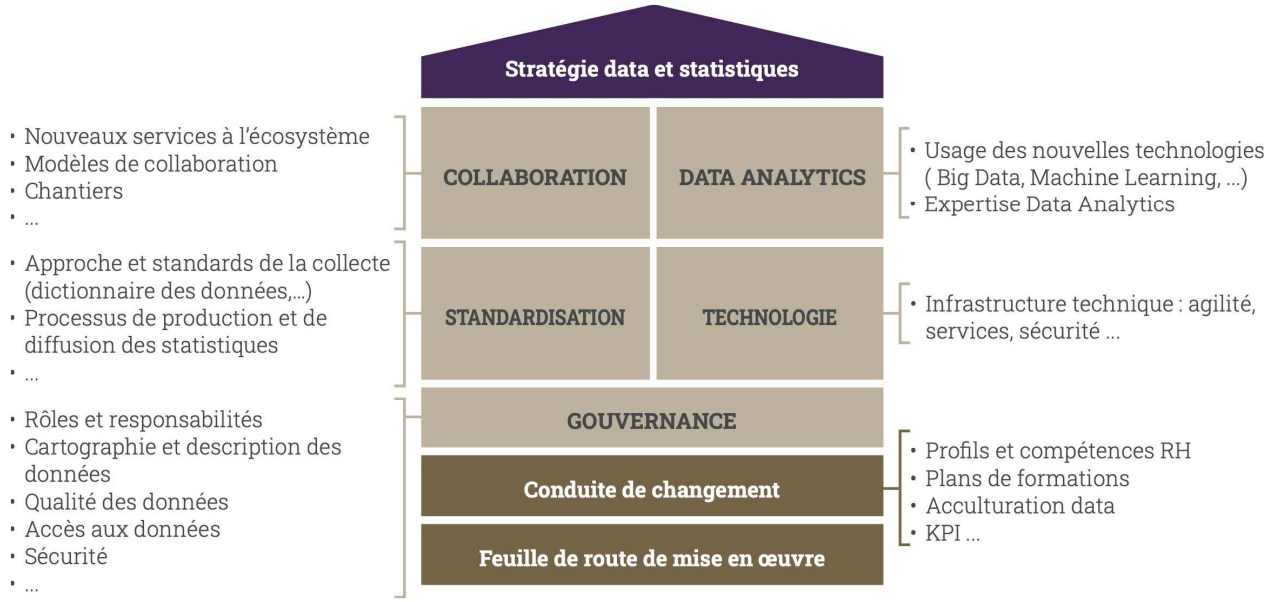


Figure 1.3 – Key Axes of the Data and Statistics Strategy

1.3.4 External Challenges and Technological Advancements

Externally, central banks operate in an environment characterized by:

- Rapid growth in new and diverse data sources following economic changes and increased reporting requirements post-financial crisis.

- Growing demand for granular data for integrated analytical needs, necessitating enhanced data access through sharing platforms.
- Emergence of new information technologies facilitating advanced data exploration and analysis, including alternative sources such as websites, Google searches, and multimedia content.

These challenges underscore the importance for central banks to deliver high-quality information efficiently, supporting international comparisons and individual behavior analysis. The focus on mastering specific technologies like Big Data has shifted data to the forefront of business strategy, emphasizing sharing over silos and prioritizing data quality management and information system resilience.

Bank Al-Maghrib's high maturity in data and statistical management positions it well to accelerate its transformation, aligning with international central banking practices. This transformation, anchored in a structured data and statistical strategy, aims to enhance operational efficiency, mitigate risks through data quality control, and improve governance for prioritizing data-related investments.

1.3.5 Strategy Implementation Approach

The implementation of Bank Al-Maghrib's data and statistical strategy involves collaborative efforts among internal stakeholders and key partners, whether subject to the bank's oversight or not. This approach encompasses:

- Assessment of internal and ecosystem landscapes, alongside benchmarking against international trends.
- Definition of target strategy focusing on data governance, process standardization, technological infrastructure, data analytics development, and ecosystem collaboration.
- Development of an implementation roadmap, estimation of financial and human resources, and change management plan to guide execution.

This comprehensive strategy aims to empower Bank Al-Maghrib with robust data capabilities, fostering innovation, resilience, and informed decision-making in support of its core mandates and strategic objectives.

1.4 Thematic

This section provides a detailed overview of the thematic focus, outlining the subject matter, the tasks involved, the constraints, objectives, and planning required to achieve the project's goals.

1.4.1 Presentation of the Subject

The project is centered on enhancing the process of detecting and managing outliers in financial data using advanced machine learning techniques. The goal is to optimize existing methodologies to improve data quality, ensuring accurate and reliable financial statistics and reports. This involves understanding and refining the current approaches used by Bank Al-Maghrib to identify anomalies in vast datasets.

1.4.2 Tasks Involved

The project involves several critical tasks:

Analysis and Understanding of the Existing Program

- Review and comprehend the existing data processing and anomaly detection systems.
- Engage in discussions with domain experts to gather insights into the current methodologies and their limitations.
- Identify potential areas for improvement in terms of efficiency, accuracy, and robustness of the current system.

Optimization of the System

- Implement improvements based on the analysis to enhance the system's performance and reliability.
- Refine processes to ensure better handling of data and more accurate detection of outliers.

Testing and Validation

- Conduct thorough testing of the optimized system to ensure it meets the desired performance criteria.
- Validate the system's accuracy in identifying outliers and its robustness in handling different data scenarios.

Documentation and Reporting

- Update the system documentation to reflect the improvements and changes made.
- Prepare comprehensive reports detailing the project progress, methodologies used, and outcomes achieved.

1.4.3 Constraints

The project faces several constraints, including:

- **Data Quality**: Ensuring the collected data is accurate, complete, and reliable.
- **Time Limitations**: Completing the project within the specified period.
- **Resource Availability**: Managing the computational and human resources required for the project.
- **Regulatory Compliance**: Adhering to legal and regulatory requirements related to data privacy and security.

1.4.4 Objectives to Achieve

The key objectives of the project include:

- **Enhancing Data Quality**: Improving the overall quality of financial data through better anomaly detection.
- **Efficiency**: Streamlining processes to reduce the time and effort required for data validation.

- **Scalability**: Developing solutions that can handle increasing amounts of data without compromising performance.
- **Knowledge Transfer**: Ensuring that the improved methodologies and processes can be effectively transferred to and utilized by Bank Al-Maghrib's team.

1.4.5 Planning

The project planning involves several key steps:

1. **Initial Assessment**: Conducting a comprehensive assessment of the current anomaly detection processes and identifying key areas for improvement.
2. **Design Phase**: Developing a detailed plan for the enhancements, including theoretical frameworks and methodologies.
3. **Implementation Phase**: Applying the planned improvements to the existing system.
4. **Testing Phase**: Rigorous testing of the enhanced system to ensure it meets all specified requirements.
5. **Deployment Phase**: Implementing the enhanced system in the operational environment.
6. **Training and Knowledge Transfer**: Conducting training sessions and providing documentation to ensure effective use and maintenance of the enhanced system.
7. **Evaluation and Feedback**: Continuously evaluating the performance of the enhanced system and making necessary adjustments based on feedback.

1.5 Conclusion

This chapter has provided an essential overview of Bank Al Maghrib, including its structure, functions, and the framework of monetary policy in Morocco. This contextual understanding is critical as it sets the stage for the problem analysis in the next chapter. By comprehending the operational environment and the significance of accurate financial data in the context of monetary policy, we are better positioned to address the challenges related to data quality.

The insights gained from this chapter highlight the importance of maintaining high data quality in financial reporting, which is a central theme in the subsequent chapters. In Chapter 2, we will delve into the specific problems associated with outlier detection in financial datasets, analyzing existing methods, their limitations, and proposing a solution to enhance data quality. The foundational knowledge provided in this chapter ensures that the reader is well-prepared to understand and engage with the problem analysis and the proposed methodologies in the following chapters.

Chapter 2

Problem Analysis

Building upon the foundational understanding established in Chapter 1, this chapter delves into the core problem statement addressed in this thesis: optimizing outlier detection methods to enhance the quality and reliability of financial data at Bank Al-Maghrib. By situating the problem within the broader context of financial data integrity and regulatory compliance, we aim to critically evaluate existing methodologies and propose advanced techniques that can overcome current limitations.

Financial institutions, particularly central banks like Bank Al-Maghrib, rely on accurate financial data for crucial decision-making processes. The integrity of this data is pivotal not only for regulatory compliance but also for ensuring economic stability and facilitating effective policy formulation. Anomalies in financial datasets, whether due to errors, fraud, or systemic risks, can significantly impact the reliability of analysis and subsequent decisions. Therefore, robust outlier detection mechanisms are essential to safeguarding data quality and maintaining trust in financial systems.

This chapter aims to provide a comprehensive analysis of the existing outlier detection methods employed at Bank Al-Maghrib. It will critically examine the strengths and limitations of these methods, identifying key challenges that hinder their effectiveness in a dynamic financial environment. By elucidating these challenges, we lay the groundwork for proposing enhancements and advanced techniques that promise to improve anomaly detection accuracy, scalability, and efficiency.

Chapter 1 established the foundational context of Bank Al-Maghrib's role and the significance of financial data integrity. Chapter 2 builds upon this context by focusing specifically on outlier detection within financial datasets. The insights gained from Chapter 2 will serve as a bridge to Chapter 3, where we delve into the theoretical underpinnings of outlier detection techniques. By understanding the specific challenges and requirements identified in this chapter, we can explore theoretical frameworks and methodologies that address these challenges comprehensively.

2.1 Problem Statement

2.1.1 General context

Bank Al-Maghrib, the Central Bank of Morocco, plays a pivotal role in maintaining financial stability and driving economic growth. A critical aspect of its mandate involves meticulous analysis of financial data to inform strategic monetary policy decisions. Given the intricate nature and vast volume of financial transactions, detecting anomalies within this data is paramount to safeguarding its integrity and ensuring accurate policy formulation. Anomalies may signify errors, fraudulent activities, or irregularities that, if undetected, could potentially distort policy decisions or compromise financial stability.

2.1.2 Problematic

This project focuses on optimizing outlier detection methods to enhance the quality of financial datasets at Bank Al-Maghrib. Current practices primarily rely on algorithms such as manual double-checking and Mahalanobis Distance for anomaly detection. While effective, these methods face challenges related to performance, accuracy, and scalability. The objective is to refine these existing methods to achieve more precise and efficient anomaly detection capabilities. By doing so, the project aims to elevate the overall quality and reliability of financial data analysis, thereby reinforcing the central bank's ability to make informed decisions and maintain financial stability.

2.1.3 Analysis of existing method

The current outlier detection methods utilized at Bank Al-Maghrib primarily involve the computation of Mahalanobis Distance (MD) (Mahalanobis, 1936) to identify anomalous values in financial data. This method plays a critical role in enhancing the reliability of financial data analysis by focusing efforts on potentially anomalous data points as in the identification of outliers in multivariate data, rather than manually reviewing every entry in annual financial statements (Rocke and Woodruff, 1996).

Mahalanobis Distance measures the distance of a point from the centroid of a distribution, considering the covariance structure of the data. This method plays a critical role in enhancing the reliability of financial data analysis by focusing efforts on potentially anomalous data points rather than manually reviewing every entry in annual financial statements of Moroccan companies (Bank Al Maghrib, 2023).

Mahalanobis Distance is computed using the formula:

$$D_M(\mathrm{x}) = \sqrt{(\mathrm{x}-\mu)^T \ ^{-1}(\mathrm{x}-\mu)}$$

where x is the data point, μ is the mean vector, and is the covariance matrix.

However, despite its benefits, MD faces challenges related to computational complexity and scalability. The computation involves matrix inversion and multiplication operations, which become computationally expensive as the dimensionality of data increases (Ren et al., 2021). This computational burden limits its scalability for large-scale financial datasets, where timely anomaly detection is crucial for effective decision-making and policy formulation (Xu et al., 2023).

Moreover, the assumption of multivariate normality underlying Mahalanobis Distance may not hold in real-world financial data, where distributions are often non-Gaussian and exhibit complex patterns. This limitation necessitates preprocessing steps to transform data or alternative outlier detection techniques when dealing with diverse data distributions (Kamenetsky Yadan, 2021).

Despite these challenges, Mahalanobis Distance remains integral to the anomaly detection framework at Bank Al-Maghrib, providing a foundational approach to flagging potentially irregular financial transactions and data entries. Future enhancements could explore hybrid approaches combining MD with machine learning techniques like deep learning-based models or ensemble methods to improve accuracy and scalability in detecting anomalies across diverse financial datasets (Chalapathy et al., 2019).

2.2 Critique of Existing Methods

The current outlier detection methods utilized at Bank Al-Maghrib, primarily employing Mahalanobis Distance, play a crucial role in safeguarding the integrity of financial data. Mahalanobis Distance (MD) calculates the distance of each data point from the distribution, facilitating the identification of anomalies in Moroccan companies' annual statements. This statistical approach is advantageous for its ability to account for correlations between variables, providing a robust measure of deviation from the norm.

However, despite its strengths, MD and Isolation Forest encounter significant challenges that impact their effectiveness under evolving data landscapes. One of the prominent challenges is the computational efficiency required for large-scale financial datasets. Isolation Forest, renowned for its capability to isolate anomalies efficiently, can suffer from computational inefficiencies as dataset sizes expand. This can lead to prolonged processing times and resource-intensive operations, affecting the system's responsiveness to dynamic market conditions and emerging financial risks (Liu et al., 2008).

Similarly, Mahalanobis Distance's reliance on covariance matrix calculations poses computational burdens, particularly in high-dimensional datasets where the inversion of matrices becomes computationally expensive. As the dimensionality of data increases, the scalability of MD diminishes, making it less suitable for real-time anomaly detection and proactive decision-making in financial monitoring (Ren et al., 2021).

Moreover, the complexity of implementing and maintaining these algorithms presents operational challenges. Isolation Forest and Mahalanobis Distance require expertise in parameter tuning and algorithmic optimization to achieve optimal performance. However, the lack of standardized guidelines and best practices for their application in financial settings can lead to inconsistent results and suboptimal anomaly detection outcomes, impacting the reliability of financial data analysis (Park et al., 2018).

Furthermore, the assumptions underlying Isolation Forest and Mahalanobis Distance algorithms limit their applicability to diverse data distributions and anomaly patterns. Isolation Forest, designed for data with a uniform distribution, may struggle with datasets exhibiting complex data distributions or skewed anomalies. Similarly, Mahalanobis Distance assumes multivariate normality, restricting its effectiveness in detecting anomalies in non-Gaussian datasets without rigorous preprocessing and transformation steps (Kamenetsky Yadan, 2021).

To address these challenges and enhance the reliability of outlier detection systems at Bank Al-Maghrib, future advancements should explore scalable anomaly detection techniques. Deep learning-based approaches and ensemble methods offer promising alternatives capable of handling large-scale

financial datasets with diverse distributions and evolving anomaly profiles (Chalapathy et al., 2019).

By integrating these innovations, the central bank can bolster its data-driven decision-making processes, strengthen financial stability measures, and ensure the robustness of financial data analysis in Morocco.

2.3 Solution Proposal

To effectively address the challenges inherent in outlier detection within financial datasets, the proposed solution at Bank Al-Maghrib emphasizes integrating machine learning (ML) and deep learning (DL) techniques. The primary objective is to enhance the accuracy, efficiency, and scalability of anomaly detection systems, which are critical for maintaining the integrity of financial data and supporting informed policy decisions.

The proposed improvements include:

- Leveraging Machine Learning over traditional ways of Outlier detection
- Enhancing the codebase for better readability, maintainability, and scal- ability.
- Incorporating advanced data structures and caching mechanisms to boost performance.
- Implementing rigorous testing protocols to ensure the reliability and robustness of the optimized system.

The solution leverages ML models such as ensemble methods, support vector machines (SVM), and isolation forests to augment traditional statistical methods for anomaly detection (Chalapathy et al., 2019). These techniques excel in capturing complex patterns and dependencies within data, thereby improving overall detection accuracy by identifying subtle anomalies that may evade traditional statistical approaches (Chalapathy et al., 2019).

Deep Learning models, including deep neural networks (DNNs) and deep isolation forests, are deployed to address the non-linear relationships and high-dimensional complexities present in financial datasets (Xu et al., 2023). DNNs can learn hierarchical representations of data, enabling them to detect anomalies based on intricate features and relationships across multiple layers (Goodfellow et al., 2014). Deep isolation forests leverage unsupervised learning principles to efficiently isolate anomalies, making them suitable for real-time anomaly detection applications (Liu et al., 2008).

Improving the codebase is crucial for ensuring the system's maintainability, scalability, and performance. By adhering to best practices in software engineering and optimizing algorithms for faster execution, the system can efficiently handle large-scale financial datasets with minimal computational overhead (Pressman, 2014). This approach accelerates anomaly detection and reduces resource consumption, enhancing the system's cost-effectiveness and sustainability over time (Pressman, 2014).

Advanced data structures and caching mechanisms are integrated to expedite data retrieval and processing. Optimized data access strategies, such as in-memory caching and distributed computing frameworks, enable the system to achieve real-time anomaly detection capabilities (Pressman, 2014). These techniques are essential for managing the volume and velocity of financial data streams, ensuring timely insights and responses to emerging anomalies.

Rigorous testing protocols are implemented to validate the reliability and accuracy of the optimized anomaly detection system. This includes unit testing, integration testing across different modules, and performance testing under varying load conditions (Pressman, 2014). By simulating real-world scenarios and stress-testing the system, potential weaknesses or performance bottlenecks can be identified and addressed proactively.

Implementing these enhancements will empower Bank Al-Maghrib to uphold high standards of data quality and integrity in financial analysis (Bank Al Maghrib, 2023; Pressman, 2014). By leveraging advanced ML and DL techniques alongside optimized algorithms and efficient code practices, the proposed solution aims to significantly enhance anomaly detection accuracy, scalability, and responsiveness. This supports the central bank's mandate to ensure financial stability, mitigate risks, and facilitate evidence-based monetary policy decisions.

2.4 Requirements Specification

To ensure the successful implementation of the enhanced anomaly detection system at Bank Al-Maghrib, a comprehensive set of requirements has been identified encompassing functional, non-functional, and regulatory aspects. Functionally, the system must support the detection of anomalies across diverse datasets, leveraging advanced algorithms such as Isolation Forest, DBSCAN, HDBSCAN, and ALAD to improve accuracy and efficiency (Breunig et al., 2000; Liu et al., 2008; McInnes et al., 2017; Zenati et al., 2018). The system should facilitate real-time monitoring and alerting capabilities, enabling prompt identification and response to potential issues. Additionally, it must integrate seamlessly with existing IT infrastructure and support interoperability with other data analysis tools and platforms (Gao et al., 2010).

From a non-functional perspective, the system must demonstrate high performance and scalability, handling large volumes of data without compromising speed or accuracy (Bengio et al., 2003). It should be designed with robust security measures to protect sensitive financial data and ensure compliance with data protection regulations. The system should also exhibit high reliability and availability, with failover mechanisms to maintain continuous operation in the event of hardware or software failures (Gama et al., 2014).

Regulatory requirements necessitate that the system adheres to the stringent standards set by Bank Al-Maghrib and relevant international bodies. This includes ensuring data privacy and security in accordance with GDPR and other applicable laws (Voigt and von dem Bussche, 2017). The system should provide comprehensive audit trails and reporting capabilities to support regulatory compliance and facilitate audits (Ferreira et al., 2016).

Functional requirements

Anomaly Detection Algorithms

The system will incorporate advanced machine learning (ML) and deep learning (DL) algorithms to enhance anomaly detection accuracy and efficiency. This includes ensemble methods, support vector machines (SVM), and isolation forests for robust anomaly identification across diverse financial data sets. ML techniques are chosen for their ability to capture complex data patterns and dependencies, improving detection accuracy beyond traditional statistical methods (Chalapathy et al., 2019).

System Optimization

Optimizing the system's performance, scalability, and responsiveness is critical. This involves refactoring the codebase for improved readability and maintainability, while also implementing advanced data structures and caching mechanisms. These optimizations aim to expedite data retrieval and processing, essential for real-time anomaly detection in dynamic financial environments (Pressman, 2014).

Integration and Testing

Seamless integration of anomaly detection modules and rigorous testing protocols are paramount. Unit testing, integration testing, and performance testing under varying load conditions will validate system reliability and accuracy. Through simulated real-world scenarios, potential weaknesses or bottlenecks in performance can be identified and rectified proactively, ensuring the system meets operational requirements effectively (Pressman, 2014).

User Interface

User interfaces will be designed to facilitate intuitive system monitoring and anomaly visualization. Customizable dashboards will provide real-time anomaly alerts and data insights, enhancing user interaction and decision-making capabilities. Ensuring accessibility and responsiveness across different devices and screen sizes will further optimize user experience and system usability (Nielsen, 1994).

Non-functional requirements

Performance

The system must demonstrate high-performance capabilities to process large-scale financial datasets with minimal latency. Real-time processing is essential for timely anomaly detection and swift response to emerging financial risks (Pressman, 2014).

Scalability

Designing the system to scale seamlessly with increasing data volumes and user demands is imperative. Integration of distributed computing frameworks will support concurrent data processing tasks, maintaining system efficiency and responsiveness during periods of peak activity (Barroso and Hölzle, 2009).

Security

Data privacy and confidentiality of financial information will be safeguarded through secure data handling practices. Implementation of encryption protocols for data transmission and storage will mitigate risks of unauthorized access, ensuring compliance with regulatory standards and safeguarding sensitive financial data (Lynch and McCarthy, 2016).

Reliability

Maintaining high availability and reliability of the anomaly detection system is crucial to minimize operational disruptions. Fault-tolerant mechanisms will be integrated to ensure continuous system operation during potential failures or maintenance activities, enhancing overall system reliability (Pressman, 2014).

Usability

Comprehensive documentation and user guides will be provided to facilitate system administration and user training. Training sessions will familiarize users with system functionalities and anomaly detection tools, promoting effective utilization and maximizing operational efficiency (Nielsen, 1994).

Regulatory and Compliance Requirements

Data Governance

Adherence to regulatory standards and guidelines governing financial data handling and anomaly detection practices is mandatory. Compliance with data protection regulations and audit requirements applicable to financial institutions will be ensured to uphold data integrity and regulatory compliance (Bank Al Maghrib, 2023).

Ethical Considerations

Ethical guidelines will govern the use of ML and DL techniques in financial anomaly detection to ensure transparency and fairness. Mitigating biases and maintaining ethical standards in decision-making processes will be prioritized, fostering trust and confidence in the system's operations (Chalapathy et al., 2019).

2.5 Methodology of work – CRISP-DM

The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology is selected for this project, renowned for its structured and iterative approach in data mining (Shearer, 2000). It guides projects through six essential phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment (Chapman et al., 2000). This cyclic process ensures a systematic progression from initial business goals to the deployment of actionable insights, allowing for continuous refinement and adaptation to new insights and requirements (Wirth and Hipp, 2000).

CRISP-DM serves as a comprehensive process model tailored to effectively manage data-centric projects. By iterating through its phases, it facilitates continuous improvement and alignment with business objectives (Kurgan and Musilek, 2006). Each phase contributes to enhancing data analysis techniques and models, thereby improving decision-making processes and operational efficiencies (Azevedo and Santos, 2008).

At its core, CRISP-DM prioritizes understanding business goals and requirements before delving into data exploration and model development. The initial phase of Business Understanding involves engaging stakeholders to clearly define the problem statement and establish measurable objectives. This approach ensures that subsequent data mining efforts are purpose-driven and closely aligned with organizational priorities (Shafique and Qaiser, 2014).

Application of Methodology in this Project

The CRISP-DM methodology guided the project from inception to deployment, ensuring a systematic and iterative approach. Key activities undertaken in each phase included:

Business Understanding:

- Conducted Department members' interviews to define the problem of outlier detection in financial data.
- Established project goals, including improving data quality and enhancing decision-making processes.

Data Understanding:

- Collected financial statements and related datasets.
- Performed exploratory data analysis (EDA) to identify patterns, trends, and anomalies.

Data Preparation:

– Cleaned and preprocessed the data using Python and pandas.
– Integrated various data sources and transformed the data into a suitable format for modeling.

Modeling:

– Applied multiple algorithms, including Mahalanobis Distance, Isolation Forest, and Local Outlier Factor, to build outlier detection models.
– Fine-tuned model parameters and evaluated performance.

Evaluation:

– Validated models using cross-validation and performance metrics such as precision, recall, and F1-score.
– Compared model outputs to business objectives and made necessary adjustments.

Deployment:

– Deployed the final model using Pickle and H5, ensuring integration with existing business systems.
– Implemented monitoring tools to continuously assess model performance and make adjustments as needed.

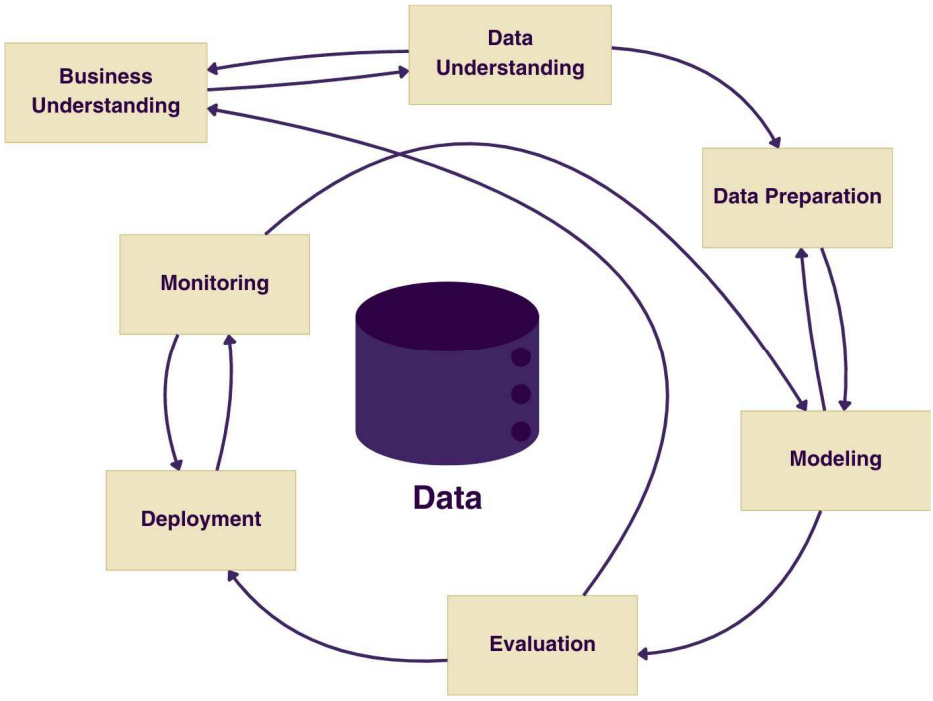


Figure 2.1 – CRISP-DM Chart

2.6 Project Planning

Effective project planning is crucial for the successful execution and delivery of data mining projects. To ensure each phase of the project is executed efficiently and meets the established objectives, a structured approach to planning was employed. This approach encompasses task scheduling, resource allocation, and timeline management.

Central to this planning process was the use of project management principles, which guided a systematic and organized execution of tasks. A key tool in this process is the Gantt chart, which visually represents the project schedule. By highlighting key tasks, their durations, and dependencies, the Gantt chart plays a vital role in tracking progress and maintaining alignment with the project's timeline. This ensures that all phases are completed on schedule and contributes to the overall success of the project.

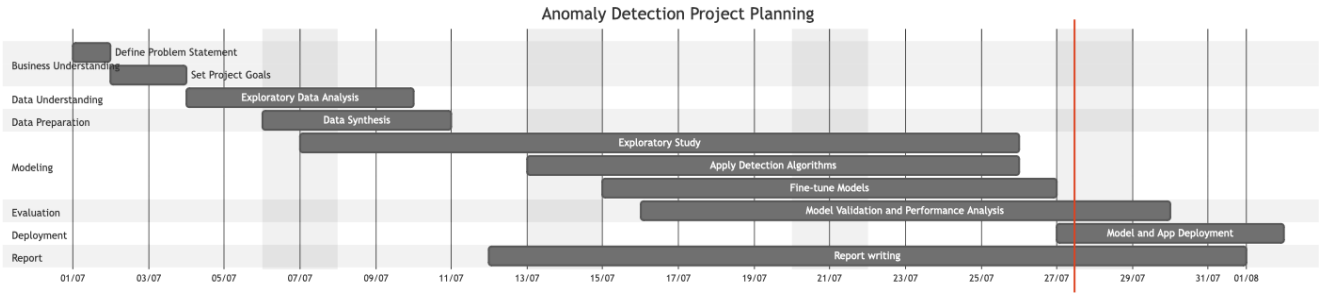


Figure 2.2 – Gantt Diagram

2.7 Conclusion

In conclusion, Chapter 2 has provided a detailed analysis of the challenges and opportunities associated with outlier detection in the financial data ecosystem of Bank Al-Maghrib. By examining the current methodologies, including Isolation Forest and Mahalanobis Distance, we have identified key limitations in terms of performance, accuracy, and scalability. These limitations underscore the need for enhanced outlier detection methods that can effectively manage the complexities of modern financial datasets.

The critical evaluation of existing methods has laid the foundation for proposing a multifaceted solution aimed at optimizing outlier detection systems. This proposed solution includes algorithmic refinements, technological upgrades, and a methodological framework anchored in CRISP-DM. By refining algorithms and adopting advanced data structures, we aim to improve anomaly detection accuracy and efficiency. The methodological framework of CRISP-DM ensures a structured approach from data understanding to deployment, facilitating iterative improvements and alignment with business objectives.

Moving forward, Chapters 3, 4, and 5 will build upon this foundational analysis and proposed solution. Chapter 3 will delve into the theoretical underpinnings of outlier detection techniques, providing a robust framework for methodological development. Chapters 4 and 5 will detail the implementation of the proposed solution, including algorithm optimization, model development, and integration into Bank Al-Maghrib's operational environment.

By addressing the identified challenges and leveraging advanced methodologies, this thesis aims to enhance the reliability, scalability, and efficiency of outlier detection in financial data analysis. The insights gained from Chapter 2 will guide subsequent chapters in realizing these objectives, ensuring that our solutions are not only theoretically sound but also practical and impactful within the context of Bank Al-Maghrib's operational landscape.

Chapter 3

Theoretical Background

Building on the problem analysis from the previous chapter, this delves into the theoretical underpinnings of outlier detection and the various techniques used in this domain by providing the necessary theoretical foundation for understanding and implementing advanced outlier detection methods. The transition from identifying the problem to exploring potential solutions requires a deep understanding of existing methodologies and their theoretical bases.

The chapter begins with a historical background, tracing the evolution of outlier detection techniques and their significance in various fields, particularly in financial data quality. Understanding the historical context helps in appreciating the advancements and current trends in outlier detection research. We then discuss the importance of outlier detection in maintaining financial data quality, highlighting how accurate detection and handling of outliers can prevent erroneous financial reporting and enhance decision-making processes.

The chapter then surveys cutting-edge techniques for outlier detection, including statistical-based, distance-based, density-based, and machine learning approaches.

Each approach is explored in detail:

- **Statistical-Based Approaches**: These methods, including Z-score and Grubbs' test, rely on statistical properties of the data to identify outliers.
- **Distance-Based Approaches**: Techniques such as the Mahalanobis Distance measure the distance of each data point from the center of the data distribution, identifying those that lie far away as potential outliers.
- **Density-Based Approaches**: Methods like Local Outlier Factor (LOF) identify outliers based on the density of data points in their neighborhood.
- **Machine Learning Approaches**: Advanced techniques like Isolation Forest and deep learning methods, including Adversarially Learned Anomaly Detection (ALAD), are explored for their ability to detect complex, non-linear patterns in large datasets.

Special emphasis is placed on methods such as Data Synthesis, the Mahalanobis Distance, Local Outlier Factor, Isolation Forest, and deep learning techniques like GANs and ALAD. Each of these methods offers unique advantages and is suited to different types of data and applications. This

comprehensive review provides the theoretical foundation for the methodologies employed in the study, ensuring that our approach is grounded in the latest research and best practices.

The theoretical insights gained in this chapter will inform the solution engineering discussed in Chapter 4. By understanding the strengths and limitations of various outlier detection techniques, we can make informed decisions about the methods and algorithms that will be implemented and optimized in our study.

The concept of outlier detection, also referred to as anomaly detection, has a rich history deeply rooted in the fields of statistics and data analysis. Its origins can be traced back to the early 19th century, with significant contributions from renowned mathematicians and statisticians. Notable early works include those by Francis Galton and later advancements by John Tukey in the 20th century, who emphasized the importance of identifying unusual data points in exploratory data analysis (Galton, 1886; Tukey, 1977).

In modern times, the field has expanded significantly, incorporating state-of-the-art methods and techniques. Contemporary research has provided comprehensive surveys and overviews of progress in outlier detection methods, emphasizing the evolution and advancements in this area. Chalapathy et al. (2019) and Wang et al. (2019) provide extensive reviews of these techniques, highlighting the shift from simple statistical tests to complex machine learning and deep learning algorithms.

3.1 Background History

Outlier detection, a crucial aspect of data analysis, has evolved significantly since its inception. Initially, the primary focus was on identifying and managing anomalies in datasets to ensure the accuracy and reliability of statistical analyses. The origins of outlier detection can be traced back to the early works of Francis Galton in the 19th century, who introduced the concept of "statistical outliers" in the context of normal distribution (Galton, 1886).

In the early 20th century, statisticians such as John Tukey further developed these ideas, introducing exploratory data analysis techniques that emphasized the importance of identifying unusual data points (Tukey, 1977). The development of robust statistical methods in the mid-20th century, such as Grubbs' test and the Z-score, provided more systematic approaches to detect outliers based on statistical properties of the data (Grubbs, 1969; Barnett and Lewis, 1994).

The advent of computer technology in the latter half of the 20th century significantly expanded the scope and capabilities of outlier detection techniques. Distance-based methods, such as the Mahalanobis Distance, emerged as powerful tools for identifying multivariate outliers by measuring the distance of each data point from the center of the data distribution (Mahalanobis, 1936). These methods were further refined and complemented by density-based approaches, such as the Local Outlier Factor (LOF), which assesses the density of data points in their neighborhood to identify anomalies (Breunig et al., 2000).

In the 21st century, the rise of big data and machine learning has revolutionized outlier detection. Advanced techniques, such as Isolation Forest, leverage machine learning algorithms to isolate anomalies by randomly partitioning data (Liu et al., 2008). Deep learning methods, including Generative Adversarial Networks (GANs) and Adversarially Learned Anomaly Detection (ALAD), have further pushed the boundaries of outlier detection by capturing complex, non-linear patterns in large datasets (Zenati et al., 2018).

Understanding the historical context of outlier detection helps appreciate the advancements and current trends in this field. Over the years, the focus has shifted from simple statistical tests to sophisticated machine learning algorithms, reflecting the increasing complexity and scale of modern datasets.

Applications of Outlier Detection in Real Life

Outlier detection has numerous applications across various industries, proving its importance beyond just theoretical interest. In finance, it is crucial for identifying fraudulent transactions, market manipulation, and unusual trading activities. By detecting anomalies in financial transactions, institutions can prevent fraud, ensure compliance with regulations, and maintain market integrity (Bolton and Hand, 2002).

In healthcare, outlier detection helps in identifying unusual patterns in patient data, which can indicate medical errors, billing fraud, or outbreaks of diseases. It aids in early detection of diseases and anomalies in medical imaging, contributing to better patient care and resource management (Chandola et al., 2009).

Manufacturing industries use outlier detection for quality control and predictive maintenance. By identifying deviations from normal operation patterns, companies can predict equipment failures and take preventive measures to avoid costly downtimes (Mourtzis et al., 2016).

In cybersecurity, detecting outliers is fundamental for identifying potential security breaches and network intrusions. By monitoring network traffic and system logs for unusual activities, organizations can safeguard against attacks and protect sensitive information (Garcia-Teodoro et al., 2009).

The wide range of applications underscores the versatility and necessity of robust outlier detection methods across different sectors, each with its own unique challenges and requirements.

Significance of Outlier Detection in Financial Data Quality

Outlier detection plays a pivotal role in maintaining the quality of financial data. Accurate detection and handling of outliers are essential for preventing erroneous financial reporting, which can have significant consequences for decision-making processes and regulatory compliance (Aggarwal, 2017). In the financial sector, outliers may indicate fraudulent activities, accounting errors, or unusual market conditions, making their timely identification crucial for maintaining the integrity and reliability of financial analyses (Hodge and Austin, 2004).

Given this context, outlier detection methods are critical for ensuring that financial data is accurate and reliable. By identifying and addressing anomalies, organizations can mitigate risks, improve financial reporting accuracy, and comply with regulatory standards (Chandola et al., 2009). For institutions like Bank Al-Maghrib, implementing an effective outlier detection system is vital for safeguarding data integrity and supporting sound financial management (Bank Al-Maghrib, 2021).

High-quality data is paramount in finance because institutions rely on vast amounts of information to drive investment strategies, risk management, and regulatory compliance (Han et al., 2011). Outliers in financial data can distort analytical models, leading to inaccurate predictions and faulty decision-making. Consequently, identifying and addressing these anomalies enhances the reliability and accuracy of data, leading to better-informed decisions and improved financial outcomes (Aggarwal, 2015).

Furthermore, outlier detection enhances data-driven decision-making processes in finance. By identifying unusual patterns and deviations from the norm, financial institutions can gain valuable insights, uncover hidden trends, and proactively address potential issues (Hodge and Austin, 2004). This capability leads to better risk management, improved fraud detection, and more accurate forecasting (Chandola et al., 2009).

The significance of outlier detection is underscored by its historical evolution. From early astro-

nomical observations to modern deep learning techniques, the continuous efforts of researchers and practitioners reflect the growing complexity of data and the need for robust and reliable methods to ensure data quality and integrity (Chalapathy and Chawla, 2019). This rich history demonstrates the enduring importance of developing effective outlier detection methods to meet the demands of increasingly complex financial data environments (Wang et al., 2019).

3.2 Distance-Based Approaches

Distance-based approaches are a fundamental category of outlier detection techniques grounded in the theoretical concept of measuring the dissimilarity between data points. At its core, distance-based methods rely on the idea that outliers are data points that deviate significantly from the majority of the data. The notion of distance in this context is mathematically defined as a measure of the spatial separation between two elements within a data space.

Formally, the distance between two elements x and y in an n-dimensional space can be described by a distance metric or function. A distance metric $d(x, y)$ quantifies how far apart the two points are, adhering to specific mathematical properties such as non-negativity, identity of indiscernibles, symmetry, and the triangle inequality (Munkres, 2000). These properties ensure that the distance function provides a meaningful measurement of dissimilarity.

Distance-based approaches exploit these distance metrics to detect outliers by identifying points that lie far away from the majority of the data points. This assumption is based on the idea that normal data points tend to cluster together, whereas outliers are isolated or sparse. By evaluating how far each point is from its neighbors or from a central reference point, distance-based methods classify points with large distances as outliers.

In this section, we will explore several prominent distance-based approaches used in outlier detection, including Euclidean distance, Mahalanobis distance, and other relevant metrics. Each approach is defined by its specific method for calculating distance and is suited to different types of data and applications. Understanding these methods and their theoretical foundations is crucial for selecting and implementing effective outlier detection techniques.

3.2.1 Euclidean Distance

Euclidean distance is one of the most fundamental and widely utilized distance metrics in the field of outlier detection. It measures the straight-line distance between two points in a multi-dimensional space, providing a simple yet powerful means of quantifying the dissimilarity between data points. The Euclidean distance between two points $\mathrm{x} = (x_1, x_2, \ldots, x_n)$ and $\mathrm{y} = (y_1, y_2, \ldots, y_n)$ in an n-dimensional space is defined mathematically by the formula:

$$d(\mathrm{x}, \mathrm{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

This formula calculates the length of the shortest path between the two points, which is a straight line in Euclidean space. The concept of Euclidean distance is rooted in the principles of Euclidean geometry, which were developed by the ancient Greek mathematician Euclid (Euclid, 300 BC).

Euclidean distance adheres to several key properties of a metric, including:

1. Non-Negativity: $d(x, y) \geq 0$
2. Identity of Indiscernibles: $d(x, y) = 0$ if and only if $x = y$
3. Symmetry: $d(x, y) = d(y, x)$
4. Triangle Inequality: $d(x, y) + d(y, z) \geq d(x, z)$

These properties ensure that Euclidean distance is a consistent and meaningful measure of dissimilarity between points in a given space (Munkres, 2000).

In the context of outlier detection, Euclidean distance is employed to identify data points that are significantly distant from their neighbors. Points with large Euclidean distances relative to the majority of other points are flagged as outliers. This method is particularly effective in low-dimensional spaces where data points are clustered around a central point. However, its effectiveness diminishes in high-dimensional spaces due to the "curse of dimensionality", where distances between points become less distinguishable and meaningful (Beyer et al., 1999).

A study by Tanger (2018) highlights the utility of Euclidean distance for outlier detection, emphasizing its simplicity and effectiveness in various practical scenarios. Tanger argues that while Euclidean distance may be straightforward, its application in outlier detection remains robust, especially when combined with other techniques to address its limitations in high-dimensional settings (Tanger, 2018).

Despite its advantages, Euclidean distance has certain limitations. In high-dimensional datasets, the distance between points can become less informative due to the sparsity of the data. Additionally, when dimensions have different scales or units, the Euclidean distance may produce misleading results, necessitating normalization or dimensionality reduction techniques to improve its performance (Jain et al., 1999).

3.2.2 Mahalanobis Distance

Mahalanobis distance is a statistical measure used to determine the distance between a point and a distribution, considering the correlations between variables in the dataset. Unlike Euclidean distance, which measures the straight-line distance between points in space, Mahalanobis distance takes into account the variance and covariance of the data, making it particularly useful for identifying outliers in multivariate datasets.

Mathematically, the Mahalanobis distance between a point x and a distribution with mean μ and covariance matrix is defined as:

$$d_M(x, \mu) = \sqrt{(x - \mu)^T \ ^{-1}(x - \mu)}$$

Here, x is the data point under consideration, μ is the mean vector of the distribution, is the covariance matrix, and $^{-1}$ is the inverse of the covariance matrix. The term $(x - \mu)^T \ ^{-1}(x - \mu)$ computes the squared distance of x from μ, adjusted for the correlations between the variables.

Mahalanobis distance is a significant improvement over Euclidean distance in the context of multivariate data because it accounts for the distribution of the data. While Euclidean distance assumes that all dimensions are equally important and independent, Mahalanobis distance incorporates the shape of the data distribution, providing a more accurate measure of how far a point is from the mean, given the data's variance and correlations (Mahalanobis, 1936).

In practice, Mahalanobis distance is particularly useful for detecting outliers in datasets where variables are correlated. For instance, in financial data, where returns of different assets might be

correlated, Mahalanobis distance helps identify assets whose returns significantly deviate from their expected patterns considering the correlations among assets (Dudoit and Fridlyand, 2003).

A study by Iglewicz and Hoaglin (1993) demonstrates the effectiveness of Mahalanobis distance in identifying outliers in multivariate datasets, noting its ability to detect outliers that might be overlooked by methods that do not consider the correlation structure (Iglewicz and Hoaglin, 1993). However, Mahalanobis distance requires an accurate estimation of the covariance matrix, which can be challenging in high-dimensional spaces where the covariance matrix may be singular or poorly conditioned (Chen et al., 2001).

Mahalanobis distance offers a robust method for outlier detection in multivariate settings by incorporating the distribution's statistical properties. Its ability to account for correlations between variables makes it a valuable tool in various applications, including financial analysis and quality control.

3.2.3 Minkowski Distance

Minkowski distance is a generalized metric used to measure the distance between two points in a multi-dimensional space. It extends the concept of Euclidean and Manhattan distances by introducing a parameter that allows for flexibility in the distance measurement. The Minkowski distance between two points $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in n-dimensional space is defined as:

$$d_p(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}$$

Here, p is a parameter that determines the type of distance measurement: - For $p = 1$, the Minkowski distance simplifies to the Manhattan distance, which measures the sum of the absolute differences of their coordinates. - For $p = 2$, it becomes the Euclidean distance, which measures the straight-line distance between the points. - For $p \to \infty$, it approaches the Chebyshev distance, which measures the maximum absolute difference between the coordinates.

Minkowski distance generalizes both Euclidean and Manhattan distances, making it a versatile tool for various applications in outlier detection and data analysis. The parameter p allows users to adjust the sensitivity of the distance metric to different types of data and anomalies (Minkowski, 1907).

In outlier detection, Minkowski distance provides a flexible approach for identifying outliers by adjusting the parameter p based on the specific characteristics of the dataset. For example, in cases where outliers are expected to exhibit more pronounced deviations in specific dimensions, choosing a higher p value can amplify the impact of these deviations (Shao et al., 2016). Conversely, a lower p value may be more appropriate for datasets where deviations are more uniform across dimensions.

A study by Shao et al. (2016) explores the use of Minkowski distance for outlier detection, highlighting its ability to accommodate various distance metrics through the choice of p. This flexibility allows for customization of the distance metric to better suit the characteristics of the data and the nature of the anomalies (Shao et al., 2016). However, the choice of p can significantly influence the results, and selecting an appropriate value requires careful consideration of the data distribution and the specific objectives of the analysis.

Minkowski distance is a powerful and adaptable metric for outlier detection, offering the ability

to fine-tune the distance measurement according to the needs of the dataset. Its generalization of both Euclidean and Manhattan distances makes it a valuable tool for a wide range of applications in data analysis.

3.2.4 Chebyshev Distance

Chebyshev distance, also known as maximum metric or L_∞ norm, is a distance metric that measures the greatest of the absolute differences between corresponding coordinates of two points. It is particularly useful in contexts where the most significant deviation in any single dimension should be highlighted.

Mathematically, the Chebyshev distance between two points $\mathrm{x} = (x_1, x_2, \ldots, x_n)$ and $\mathrm{y} = (y_1, y_2, \ldots, y_n)$ in n-dimensional space is defined as:

$$d_C(\mathrm{x}, \mathrm{y}) = \max_{i=1,\ldots,n} |x_i - y_i|$$

In this formula, $|x_i - y_i|$ represents the absolute difference between the coordinates of x and y in the i-th dimension. The Chebyshev distance is thus the maximum of these differences, effectively capturing the largest deviation in any single dimension.

The Chebyshev distance is named after the Russian mathematician Pafnuty Chebyshev, who first introduced it in the 19th century (Chebyshev, 1962). This distance metric is particularly useful in scenarios where deviations in any one dimension are more critical than deviations in the overall distance. It is commonly used in chessboard distance calculations, where the goal is to determine the minimum number of moves a king needs to move from one square to another, considering that the king can move horizontally, vertically, or diagonally (Chebyshev, 1962).

In outlier detection, Chebyshev distance can be advantageous when anomalies are expected to exhibit extreme deviations in a single feature or dimension. For instance, in certain types of quality control processes, a single outlier in a crucial dimension might be more significant than smaller deviations across multiple dimensions. By focusing on the maximum deviation, Chebyshev distance allows for effective identification of such anomalies (Tukey, 1977).

A study by Tukey (1977) emphasizes the usefulness of Chebyshev distance in identifying outliers in high-dimensional datasets, where it can help isolate extreme deviations that might be missed by other distance metrics. The simplicity of the Chebyshev distance makes it computationally efficient, though it may not always capture the complexity of multidimensional anomalies as effectively as other metrics (Tukey, 1977).

Chebyshev distance is a straightforward and effective metric for highlighting extreme deviations in any single dimension of a dataset. Its focus on the maximum difference makes it particularly suitable for applications where the most significant deviation is of primary concern.

While distance-based approaches identify outliers by measuring the spatial distance between points, density-based approaches focus on the local density of data points. Unlike distance-based methods, which can struggle with varying data scales and high dimensions, density-based techniques excel in detecting outliers in regions with low data density. This shift from distance-based to density-based methods reflects a progression towards more nuanced and robust outlier detection strategies.

3.3 Density-Based Approaches

Density-based approaches to outlier detection are grounded in the concept of data density, which refers to the concentration of data points in a given region of the feature space. Unlike distance-based methods that focus solely on the spatial separation between points, density-based techniques assess the relative density of data points to identify outliers. The core idea is that outliers are data points that lie in regions with significantly lower density compared to their surrounding neighborhoods.

In a statistical context, density refers to the distribution of data points within a specific area of the feature space. For a given point x, its density can be quantified by counting the number of neighboring points within a predefined radius or by estimating the probability density function (PDF) of the data distribution around x. Mathematically, the density $\rho(\mathrm{x})$ at a point x is often defined as:

$$\rho(\mathrm{x}) = \frac{1}{|N(\mathrm{x})|} \sum_{\mathrm{y} \in N(\mathrm{x})} K(\|\mathrm{x} - \mathrm{y}\|)$$

where $N(\mathrm{x})$ is the set of neighbors within a certain radius of x, $K(\cdot)$ is a kernel function (e.g., Gaussian kernel), and $|N(\mathrm{x})|$ is the number of neighbors. This formula provides an estimate of the local density around x based on the proximity of neighboring data points (Parzen, 1962).

The fundamental assumption behind density-based outlier detection is that outliers typically reside in regions of the feature space that are sparsely populated with data points. In contrast, inlier points are clustered in high-density regions. By evaluating the density of each point's neighborhood, density-based methods can effectively distinguish between points that are part of a dense cluster and those that are isolated in sparser regions.

Density-based methods are advantageous because they do not require specifying the number of clusters or outliers beforehand, and they are robust to varying densities within the dataset. This makes them particularly useful in scenarios where the data distribution is uneven or contains clusters of different sizes and shapes (Ester et al., 1996).

In outlier detection, density-based approaches are used to identify points that have significantly lower density compared to their neighbors. These points are flagged as potential outliers due to their isolation in sparse regions of the feature space. In this section, we will explore several commonly used density-based algorithms:

- **Local Outlier Factor (LOF)**: LOF evaluates the local density deviation of a point relative to its neighbors. Points with a significantly lower density compared to their neighbors are flagged as outliers (Breunig et al., 2000).
- **Simplified Local Outlier Factor (sLOF)**: Simplified Local Outlier Factor (sLOF) is a variant of the traditional Local Outlier Factor (LOF) that aims to improve computational efficiency. By optimizing the algorithm's processing steps, sLOF reduces the complexity involved in calculating local density deviations while retaining its core capability to detect outliers (Schubert et al., 2017).
- **RS-Forest**: RS-Forest is a density-based method that uses a forest of random subspace models to estimate the density of data points. This technique aggregates density estimates from multiple random projections to identify outliers effectively (Xia et al., 2015).
- **Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and HDBSCAN**: DBSCAN clusters data based on density, identifying outliers as points in low-density regions that do not belong to any cluster (Ester et al., 1996).

— **Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)**: HDBSCAN extends DBSCAN by incorporating hierarchical clustering, allowing it to handle data with varying densities and complex cluster structures more effectively (McInnes et al., 2017).

These density-based methods offer robust techniques for outlier detection, particularly in datasets with varying density and complex cluster structures, enhancing the accuracy and reliability of anomaly detection processes.

3.3.1 Local Outlier Factor (LOF)

The Local Outlier Factor (LOF) is a well-established density-based method for outlier detection that identifies anomalies based on the local density of data points. Introduced by Breunig et al. (2000), LOF provides a robust mechanism for detecting outliers by comparing the density of a data point with the density of its neighbors.

At the core of LOF is the concept of local density. For a given data point, its local density is defined relative to its neighbors. This relative density is computed using a measure known as the reachability distance. Specifically, the reachability distance between a data point p and its neighbor o is the maximum of the distance between p and o and the distance between o and the k-th nearest neighbor of o.

Mathematically, the reachability distance $reach_dist_k(p, o)$ can be expressed as:

$$reach_dist_k(p, o) = \max\left(d(p, o), d(o, kNN(o))\right)$$

where $d(p, o)$ denotes the distance between points p and o, and $kNN(o)$ represents the k-th nearest neighbor of point o.

The local density of a point p is then defined as the inverse of the average reachability distance of p with respect to its k-nearest neighbors:

$$local_density_k(p) = \frac{1}{\frac{1}{|N_k(p)|} \sum_{o \in N_k(p)} reach_dist_k(p, o)}$$

where $N_k(p)$ denotes the set of k-nearest neighbors of p.

LOF evaluates the outlier score by comparing the local density of a point with the local densities of its neighbors.

Specifically, the Local Outlier Factor (LOF) for a point p is defined as:

$$LOF_k(p) = \frac{\frac{1}{|N_k(p)|} \sum_{o \in N_k(p)} \frac{local_density_k(o)}{local_density_k(p)}}{|N_k(p)|}$$

A higher LOF score indicates that the data point p is an outlier, as it is situated in a region with significantly lower density compared to its neighbors.

LOF's primary advantage is its ability to identify outliers based on local data density, which makes it effective in datasets with varying density regions. This method is particularly useful in

detecting anomalies in complex datasets where the density of data points varies spatially.

In practice, LOF has been successfully applied across various domains, including fraud detection and network security, where it helps uncover anomalous patterns that deviate from the expected local data distribution (Breunig et al., 2000).

3.3.2 Simplified Local Outlier Factor (sLOF)

The Simplified Local Outlier Factor (sLOF) is an optimized version of the Local Outlier Factor (LOF) designed to improve computational efficiency while preserving the core outlier detection capabilities of the original method. LOF, introduced by Breunig et al. in 2000, measures the local density deviation of a data point relative to its neighbors, identifying points that are significantly less dense as outliers. The sLOF method refines this approach by streamlining the computations involved in calculating density and outlier scores, making it more suitable for large-scale datasets.

The fundamental idea behind sLOF is to assess how isolated a data point is with respect to its local neighborhood. The LOF score for a data point p is computed based on its local density compared to that of its neighbors. This local density is quantified using the concept of the k-nearest neighbors (k-NN). The LOF score $LOF(p)$ for a point p is defined as:

$$LOF(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd(o)}{lrd(p)}}{|N_k(p)|}$$

where $N_k(p)$ represents the set of k-nearest neighbors of point p, and $lrd(p)$ denotes the local reachability density of point p. The local reachability density is calculated as:

$$lrd(p) = \frac{1}{reach_dist(p)}$$

where the reachability distance $reach_dist(p)$ is given by:

$$reach_dist(p) = \max\left(dist(p, o), k - dist(o)\right)$$

Here, $dist(p, o)$ is the distance between point p and its neighbor o, and $k - dist(o)$ is the distance from o to its k-th nearest neighbor.

The sLOF algorithm simplifies the original LOF by reducing the computational overhead involved in distance calculations and density estimations. This is achieved by approximating the density measures and optimizing the k-nearest neighbor search process. The result is a more computationally efficient outlier detection method that retains the effectiveness of the LOF algorithm.

The sLOF method is particularly useful in scenarios with large datasets or real-time applications where computational efficiency is critical. By maintaining a balance between accuracy and performance, sLOF provides a practical solution for detecting outliers in complex and high-dimensional data.

3.3.3 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a widely used density-based clustering algorithm that also serves as an effective method for outlier detection. Introduced

by Ester et al. (1996), DBSCAN identifies clusters based on the density of data points in a given region and labels points in low-density areas as noise or outliers.

The core idea of DBSCAN is to classify points based on their density relative to their neighbors. The algorithm relies on two key parameters: ϵ (epsilon) and *minPts* (minimum points). The parameter ϵ defines the radius within which the algorithm searches for neighboring points, while *minPts* specifies the minimum number of points required to form a dense region, or cluster.

In DBSCAN, a point is classified into one of three categories:

1. Core Point: A point is classified as a core point if it has at least *minPts* neighbors within the ϵ radius. Core points are essential for forming a cluster.
2. Border Point: A point is classified as a border point if it is within the ϵ radius of a core point but does not have enough neighbors to be a core point itself.
3. Noise Point: A point is classified as noise if it does not belong to any cluster, meaning it is neither a core point nor a border point.

Mathematically, the core idea of DBSCAN can be expressed as follows. For a given point p, the neighborhood of p within radius ϵ is defined as:

$$N_\epsilon(p) = \{q \in D \mid d(p, q) \leq \epsilon\}$$

where D denotes the dataset and $d(p, q)$ is the distance between points p and q. The point p is considered a core point if:

$$|N_\epsilon(p)| \geq minPts$$

where $|N_\epsilon(p)|$ represents the number of points within the ϵ radius of p.

DBSCAN's advantage lies in its ability to find arbitrarily shaped clusters and its robustness to outliers. By identifying dense regions in the data and separating them from sparse regions, DBSCAN can effectively handle noise and anomalies. This makes it particularly useful in applications where the data exhibits complex structures and varying densities, such as geographical data analysis and image segmentation (Ester et al., 1996).

The algorithm's effectiveness in identifying outliers stems from its ability to detect points that do not belong to any dense region. Such points, identified as noise, are effectively flagged as outliers, making DBSCAN a powerful tool for outlier detection in diverse datasets.

3.3.4 Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)

Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is an advanced extension of the DBSCAN algorithm, designed to improve clustering performance and outlier detection in datasets with varying densities. Proposed by McInnes et al. (2017), HDBSCAN builds upon the foundational principles of DBSCAN by incorporating hierarchical clustering techniques to enhance the algorithm's flexibility and robustness.

HDBSCAN operates in two main stages: hierarchical clustering and density-based clustering. The first stage constructs a hierarchy of clusters using the concept of mutual reachability distance, which generalizes the notion of density-based clustering to include varying densities. Mutual reachability

distance between two points p and q is defined as:

$$reachability_dist(p, q) = \max\{core_dist(p), core_dist(q), d(p, q)\}$$

where $core_dist(p)$ represents the core distance of point p (i.e., the distance to its *minPts*-th nearest neighbor), and $d(p, q)$ is the Euclidean distance between p and q. This distance metric ensures that the reachability distance takes into account both the local density of points and the distance between them, allowing for more nuanced cluster formation.

In the second stage, HDBSCAN performs a hierarchical clustering analysis using the mutual reachability distance. This hierarchical tree, or dendrogram, represents a nested series of clusters. To obtain a flat clustering from the hierarchical structure, HDBSCAN uses a concept called "cluster stability" to select the most meaningful clusters. Stability is measured by how well a cluster persists across different levels of the hierarchy. Clusters with higher stability are considered more significant and are retained in the final clustering result.

One of the key advantages of HDBSCAN is its ability to identify clusters of varying shapes and densities, as well as its robustness to noise. Unlike DBSCAN, which requires specifying a single ϵ parameter for clustering, HDBSCAN does not require pre-defining the number of clusters and can automatically determine the optimal number of clusters based on the stability of the hierarchical structure. This adaptability makes HDBSCAN particularly effective for complex datasets where clusters exhibit varying densities and shapes (McInnes et al., 2017).

In addition to its clustering capabilities, HDBSCAN excels in outlier detection. Points that do not belong to any stable cluster are classified as outliers or noise. This classification is based on their exclusion from significant clusters within the hierarchy, allowing HDBSCAN to robustly detect anomalies in diverse data distributions.

Overall, HDBSCAN enhances the ability to perform density-based clustering and outlier detection by incorporating hierarchical techniques, making it a powerful tool for analyzing complex datasets with varying densities and structures.

3.3.5 RS-Forest (Random Sample Consensus Forest)

The RS-Forest algorithm, or Random Sample Consensus Forest, is a sophisticated method for outlier detection that builds on the principles of ensemble learning and Random Sample Consensus (RANSAC). RANSAC, introduced by Fischler and Bolles in 1981, is a robust statistical method designed to estimate parameters of a mathematical model from a dataset that contains outliers. The core idea of RANSAC is to iteratively select random subsets of the data, fit a model to these subsets, and identify the subset that best fits the model while minimizing the influence of outliers.

In the context of RS-Forest, this principle is extended to decision tree ensembles. RS-Forest constructs an ensemble of decision trees, each trained on a different random sample of the dataset. By doing so, the method introduces diversity into the model, which helps reduce the risk of overfitting and improves its ability to generalize across different data patterns. Each decision tree in the forest is trained to classify data points based on their features, distinguishing between normal data points (inliers) and anomalous data points (outliers).

The detection of outliers in RS-Forest is achieved through a consensus mechanism across the ensemble of decision trees. For each data point, a consensus score is calculated based on how frequently the point is classified as an outlier by the various trees in the forest.

The consensus score $S(x)$ for a data point x is computed as:

$$S(x) = \frac{1}{T} \sum_{i=1}^{T} outlier_score_i(x)$$

where T is the number of trees in the forest, and $outlier_score_i(x)$ is the outlier score given by the i-th tree. This score indicates whether the point x is considered an outlier by that tree.

The consensus score aggregates the decisions of multiple trees, providing a robust measure of whether a data point is an outlier. Points with higher consensus scores are more consistently classified as outliers across the ensemble, whereas points with lower scores are classified as inliers. This approach enhances the accuracy and robustness of outlier detection by mitigating the influence of individual tree biases and accommodating diverse data distributions.

RS-Forest leverages the RANSAC principle by using random subsets of the data to train its ensemble of decision trees, which helps improve its resistance to noisy data and complex data patterns. This makes RS-Forest a valuable tool for detecting anomalies in diverse and challenging datasets.

Transitioning from density-based approaches, which identify outliers based on local data density, we now shift to machine learning-based methods. These techniques leverage advanced algorithms to learn complex patterns and detect anomalies, offering enhanced capabilities for high-dimensional and non-linearly separable data. Machine learning methods build upon the strengths of density-based approaches by providing more adaptive and scalable solutions for outlier detection.

3.4 Machine Learning-Based Approaches

Machine learning-based approaches to outlier detection leverage sophisticated algorithms to identify anomalies by learning patterns and relationships within the data. Unlike traditional statistical or distance-based methods, these techniques adaptively learn from the data, enabling them to detect complex and subtle outliers that may not be captured by simpler models (Chandola et al., 2009).

At the core of machine learning-based approaches is the ability to handle high-dimensional data and uncover non-linear relationships between features. These methods often involve training models on a dataset to distinguish between normal and anomalous instances. The models then apply the learned patterns to new, unseen data to identify potential outliers. This adaptability makes machine learning-based methods particularly effective in dynamic and complex datasets where traditional methods may fall short (Hodge and Austin, 2004).

A variety of machine learning techniques are employed for outlier detection, each with its own strengths and application scenarios. For instance, supervised methods require labeled data to train models that classify instances as normal or outlier. Conversely, unsupervised methods, which do not rely on labeled data, detect anomalies based on deviations from learned patterns or distributions (Xia et al., 2015). Additionally, semi-supervised methods use a combination of labeled and unlabeled data to improve detection performance (Jin et al., 2006).

Overall, machine learning-based approaches enhance the capability to identify outliers by leveraging advanced computational models and learning algorithms, thus offering more robust and scalable solutions for complex data analysis (Ahmed et al., 2016).

3.4.1 Isolation Forest (iForest)

Isolation Forest (iForest) is a machine learning algorithm specifically designed for anomaly detection in high-dimensional datasets. It operates on the principle of isolation, which involves isolating observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature. The main idea behind iForest is that outliers are easier to isolate compared to normal points because they lie far from the majority of data points.

The algorithm constructs an ensemble of isolation trees (iTrees), where each iTree is built by recursively partitioning the data. The process involves selecting a feature and then a split value, which partitions the data into two subsets. This process continues until each subset contains only one data point or until a predefined height is reached. The height of the tree, which represents the number of splits required to isolate a point, is used to measure the degree of isolation.

Mathematically, the isolation score $s(x)$ for a data point x can be expressed as:

$$s(x) = \frac{2^{\frac{H(x)}{c(n)}} - 1}{c(n) - 1}$$

where $H(x)$ is the average path length of point x across all trees, $c(n)$ is the average path length of unsuccessful searches in a binary search tree, and n is the number of data points. The average path length $c(n)$ is defined as:

$$c(n) = 2 \cdot \left(\frac{n-1}{2} - \frac{2(n-1)}{n} \right)$$

The isolation score $s(x)$ is higher for anomalies because they are isolated closer to the root of the trees, resulting in shorter average path lengths. Conversely, normal data points are more frequently split and thus have longer average path lengths.

iForest is particularly effective due to its efficiency in handling large datasets with high dimensionality. The algorithm's computational complexity is $O(n \log n)$, which allows it to scale well with large datasets compared to traditional outlier detection methods (Liu et al., 2008). This efficiency, combined with its ability to detect anomalies without requiring extensive parameter tuning, makes iForest a popular choice for many anomaly detection applications.

3.4.2 One-Class Support Vector Machine

Another approach to consider in machine learning-based outlier detection is the One-Class Support Vector Machine (One-Class SVM). One-Class SVM is an adaptation of the traditional Support Vector Machine (SVM), designed specifically for anomaly detection. This method is widely used due to its effectiveness in high-dimensional spaces and its ability to handle non-linear data distributions.

The fundamental principle behind One-Class SVM is to learn a decision function for outlier detection by identifying regions in the input space where the data is concentrated. The algorithm is trained only on the normal data, and it aims to separate the normal data points from the origin in the feature space using a hyperplane. The decision function $f(x)$ learned by One-Class SVM assigns a positive value to the data points that are considered normal and negative values to the anomalies.

Mathematically, One-Class SVM involves solving the following optimization problem:

$$\min_{w,\rho,\xi_i} \left(\frac{1}{2} \|w\|^2 + \frac{1}{vn} \sum_{i=1}^{n} \xi_i - \rho \right)$$

subject to

$$(w \cdot \phi(x_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0$$

Here, w is the normal vector to the hyperplane, ρ is the offset, ξ_i are the slack variables to allow for some data points to lie within the margin, $\phi(x_i)$ is a mapping function that transforms the data into a higher-dimensional feature space, and v is a parameter that controls the trade-off between the margin size and the number of outliers.

One-Class SVM employs the kernel trick to handle non-linear data. Commonly used kernels include the Radial Basis Function (RBF) kernel, which is defined as:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

where σ is a parameter that determines the spread of the kernel.

The training process involves finding the optimal hyperplane that best separates the normal data points from the origin in the transformed feature space. During testing, the decision function $f(x)$ is used to classify new data points. Points with $f(x) < 0$ are considered outliers.

One-Class SVM is particularly robust in scenarios where the normal data is well-separated from anomalies, making it suitable for a wide range of applications, including fraud detection, network security, and quality control.

3.4.3 Autoencoders

Autoencoders are a type of neural network that are trained to encode input data into a compressed representation and then reconstruct the input data from this compressed form. They are particularly useful in unsupervised learning tasks, including anomaly detection, due to their ability to learn a compact and informative representation of the data (Hinton and Salakhutdinov, 2006).

Autoencoders consist of two main components: the encoder and the decoder. The encoder maps the input data x to a hidden representation h through a series of layers:

$$h = f(x)$$

where f is a non-linear transformation, often implemented using neural network layers.

The decoder then maps this hidden representation back to a reconstruction of the input data $\hat{x}$:

$$\hat{x} = g(h)$$

where g is another non-linear transformation.

The objective of training an autoencoder is to minimize the reconstruction error, which is typically

measured using a loss function such as the Mean Squared Error (MSE):

$$L(\mathrm{x}, \hat{\mathrm{x}}) = \|\mathrm{x} - \hat{\mathrm{x}}\|^2$$

The training process involves optimizing the parameters of both the encoder and the decoder to minimize this loss over the training dataset. Once trained, the autoencoder can be used for anomaly detection by comparing the reconstruction error of a new data point to a threshold. Data points with a high reconstruction error are considered anomalies, as the autoencoder is unable to accurately reconstruct them due to their deviation from the normal data distribution (Sakurada and Yairi, 2014).

Mathematically, the anomaly score for a data point x can be defined as:

$$AnomalyScore(\mathrm{x}) = \|\mathrm{x} - \hat{\mathrm{x}}\|^2$$

Autoencoders can be enhanced with various techniques to improve their performance in anomaly detection. One common approach is to use sparse autoencoders, which impose a sparsity constraint on the hidden representations to encourage the network to learn more meaningful features (Ng, 2011). Another approach is to use variational autoencoders (VAEs), which introduce a probabilistic framework to the autoencoder, allowing it to model the uncertainty in the data and improving its robustness to anomalies (Kingma and Welling, 2014).

Variational Autoencoders introduce a latent variable z and define the encoder as a probabilistic model $q(\mathrm{z}|\mathrm{x})$ and the decoder as $p(\mathrm{x}|\mathrm{z})$. The objective is to maximize the Evidence Lower Bound (ELBO):

$$ELBO = \mathbb{E}_{q(\mathrm{z}|\mathrm{x})}[\log p(\mathrm{x}|\mathrm{z})] - KL(q(\mathrm{z}|\mathrm{x}) \| p(\mathrm{z}))$$

where KL denotes the Kullback-Leibler divergence.

The Kullback-Leibler divergence, $KL(P \| Q)$, is a measure of how one probability distribution P diverges from a second, expected probability distribution Q. For two discrete probability distributions P and Q, it is defined as:

$$KL(P \| Q) = \sum_{x \in X} P(x) \log \left(\frac{P(x)}{Q(x)} \right)$$

In the context of VAEs, P represents the true posterior distribution of the latent variables, and Q represents the approximate posterior distribution. Minimizing the KL divergence helps in making Q close to P, ensuring that the learned latent space distribution is similar to the true distribution of the data (Doersch, 2016).

Autoencoders have been successfully applied to various anomaly detection tasks, including image anomaly detection, network intrusion detection, and industrial equipment monitoring, due to their flexibility and ability to learn complex data distributions (Chen et al., 2017; Xia et al., 2015; Zhao et al., 2017).

Shifting from machine learning-based techniques to deep learning-based methods significantly improves outlier detection by enabling the management of complex and high-dimensional data. Although machine learning approaches like Isolation Forest offer strong solutions, deep learning methods such as autoencoders and GANs achieve higher accuracy by capturing intricate data patterns

through neural networks. This progression overcomes the limitations of traditional methods, enhancing anomaly detection in complex datasets.

3.5 Deep Learning-Based Approaches

Deep learning-based methods for outlier detection have become increasingly prominent due to their ability to handle complex, high-dimensional data and reveal intricate patterns that traditional methods may overlook. These approaches utilize deep neural networks to model sophisticated data distributions, providing a nuanced understanding of data structures. A key advantage of deep learning in this context is its capacity to automatically learn hierarchical feature representations, which is particularly effective for detecting subtle anomalies in large datasets.

Deep learning models excel by eliminating the need for manual feature engineering, as they can automatically extract relevant features from raw data. This is especially valuable for unstructured data types, such as images, audio, and text, where traditional methods often fall short. Additionally, deep learning techniques can capture non-linear relationships and dependencies, offering a robust framework for anomaly detection.

Autoencoders are a foundational deep learning approach for anomaly detection. They consist of an encoder that compresses input data into a latent representation and a decoder that reconstructs the original data from this compressed form. The reconstruction error, defined as the difference between the input data and its reconstruction, serves as an anomaly indicator. High reconstruction errors suggest that the data points are outliers, as the autoencoder struggles to accurately reconstruct data that deviates significantly from the training distribution (Hinton and Salakhutdinov, 2006).

Generative Adversarial Networks (GANs) also show significant promise for outlier detection. GANs involve two neural networks: a generator that creates realistic data samples and a discriminator that distinguishes between real and generated samples. This adversarial training enhances the model's ability to capture the data distribution, making it effective for identifying anomalies (Goodfellow et al., 2014). Variants such as Adversarially Learned Anomaly Detection (ALAD) further refine detection accuracy through bidirectional networks and cycle-consistency constraints (Zenati et al., 2018).

Variational Autoencoders (VAEs) introduce a probabilistic element to the autoencoder framework, modeling the latent space as a distribution rather than a fixed point. This probabilistic approach enhances robustness to anomalies and provides a principled method for measuring uncertainty in the data (Kingma and Welling, 2013). By maximizing the Evidence Lower Bound (ELBO) during training, VAEs ensure that the latent space effectively reflects the data distribution, improving anomaly detection performance.

Overall, deep learning-based approaches represent a significant advancement over traditional and machine learning methods in anomaly detection. Their ability to model complex, high-dimensional data distributions and automatically learn features makes them highly effective across various applications, including image and speech processing, network security, and industrial monitoring. The ongoing development of deep learning architectures and techniques promises further improvements in the accuracy and efficiency of outlier detection systems.

In this section, we will focus on two prominent deep learning-based approaches: Autoencoders and Generative Adversarial Networks (GANs), with a particular emphasis on Adversarially Learned Anomaly Detection (ALAD).

3.5.1 Deep Learning Autoencoders

Deep learning autoencoders and traditional machine learning autoencoders differ fundamentally in their architecture and capability to handle complex data representations.

Traditional machine learning autoencoders typically use shallow neural networks with one or two hidden layers to encode and decode data. These shallow networks limit the model's capacity to learn intricate patterns and relationships within the data. The encoding function f and decoding function g in these autoencoders are often simple linear or mildly non-linear transformations:

$$\mathrm{h} = f(\mathrm{x}) = W_1\mathrm{x} + b_1$$

$$\hat{\mathrm{x}} = g(\mathrm{h}) = W_2\mathrm{h} + b_2$$

where W_1 and W_2 are weight matrices, and b_1 and b_2 are bias terms. The limited depth and complexity of these networks restrict their ability to capture sophisticated data patterns, leading to potential limitations in detecting subtle anomalies in high-dimensional data.

In contrast, deep learning autoencoders utilize deep neural networks with multiple layers in both the encoder and decoder components. This deep architecture allows these models to learn hierarchical representations of the data, capturing more complex and abstract features. The encoding function f and decoding function g in deep learning autoencoders are defined as:

$$\mathrm{h} = f(\mathrm{x}) = f_L(f_{L-1}(\cdots f_1(\mathrm{x}) \cdots))$$

$$\hat{\mathrm{x}} = g(\mathrm{h}) = g_1(g_2(\cdots g_L(\mathrm{h}) \cdots))$$

where f_i and g_i represent the non-linear transformations applied by each layer, and L denotes the number of layers in the network. The depth of these networks enables them to model complex relationships and capture detailed features, making them more effective at identifying anomalies that are not apparent with shallow architectures.

The training of deep learning autoencoders involves minimizing the reconstruction error, similar to traditional autoencoders, but with the advantage of capturing richer data representations. The reconstruction error for deep learning autoencoders is also measured using Mean Squared Error (MSE):

$$L(\mathrm{x}, \hat{\mathrm{x}}) = \|\mathrm{x} - \hat{\mathrm{x}}\|^2$$

However, the increased capacity of deep learning autoencoders allows for more nuanced anomaly detection due to their ability to learn from complex, high-dimensional datasets.

Furthermore, deep learning autoencoders can be enhanced with techniques such as Variational Autoencoders (VAEs), which incorporate probabilistic frameworks into the autoencoder architecture, and can better handle uncertainty and anomalies (Kingma and Welling, 2013).

Overall, deep learning autoencoders offer superior performance in anomaly detection compared to traditional machine learning autoencoders due to their ability to model complex data structures and learn hierarchical feature representations.

3.5.2 Generative Adversarial Networks (GANs) for Anomaly Detection

Generative Adversarial Networks (GANs) have emerged as a powerful tool for anomaly detection due to their capability to model complex data distributions through adversarial learning. GANs consist of two neural networks: a generator and a discriminator, which engage in a competitive training process. The generator's objective is to produce data samples that resemble the true data distribution, while the discriminator aims to distinguish between real samples and those generated by the generator.

Mathematically, the generator G and discriminator D are trained through a min-max game where the generator tries to minimize the loss function, while the discriminator tries to maximize it. The objective function for this adversarial process can be expressed as:

$$\min_G \max_D \mathsf{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathsf{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))],$$

where x represents real data samples, z represents random noise inputs to the generator, $p_{data}(x)$ is the distribution of real data, $p_z(z)$ is the distribution of noise, and $D(\cdot)$ is the discriminator function that outputs the probability of a sample being real (Goodfellow et al., 2014).

The generator's loss function is:

$$L_G = -\mathsf{E}_{z\sim p_z(z)}[\log D(G(z))],$$

which encourages the generator to produce data samples that the discriminator is likely to classify as real. Conversely, the discriminator's loss function is:

$$L_D = -\mathsf{E}_{x\sim p_{data}(x)}[\log D(x)] - \mathsf{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))],$$

which trains the discriminator to accurately differentiate between real and generated samples.

In the context of anomaly detection, GANs are effective because the generator learns to model the distribution of normal data. As a result, anomalies—data points that deviate significantly from the normal distribution—are less likely to be well-represented by the generator. Consequently, these anomalies are more easily identified by the discriminator, which can highlight discrepancies in the reconstruction process.

An advanced variant of GANs used for anomaly detection is the Adversarially Learned Anomaly Detection (ALAD) framework.

3.5.3 Adversarially Learned Anomaly Detection (ALAD)

Adversarially Learned Anomaly Detection (ALAD) represents an advanced methodology for detecting outliers by utilizing the principles of Generative Adversarial Networks (GANs). This technique enhances anomaly detection by employing the adversarial learning framework, which bolsters its ability to identify deviations from standard data distributions with increased accuracy and robustness.

ALAD leverages bidirectional neural networks, incorporating both forward and reverse mappings between the data space and a latent space. This approach is designed to capture and model the underlying data distribution more effectively. In the forward direction, the network maps input data x to a latent representation z, while in the reverse direction, it reconstructs the data from this

latent space representation. The model's effectiveness is measured by the combined adversarial and cycle-consistency loss functions.

The loss function for ALAD is formulated as:

$$ALAD_Loss = \mathsf{E}_{x\sim p_{data}(x)} \|x - G(D(x))\|^2 + \mathsf{E}_{z\sim p_z(z)} \|D(G(z)) - z\|^2 ,$$

where G denotes the generator, D represents the discriminator, $p_{data}(x)$ is the data distribution, and $p_z(z)$ is the distribution of latent variables. The first term ensures that the reconstructed data is close to the original input, thereby verifying the quality of reconstruction. The second term enforces cycle-consistency, ensuring that the latent representations of the generated samples are consistent with the original latent space distribution. This consistency helps to maintain the integrity of learned features and reduces the impact of noise and variations in data (Zenati et al., 2018).

By incorporating these bidirectional networks and cycle-consistency constraints, ALAD effectively improves the detection of anomalies. This method enhances traditional GAN-based anomaly detection by providing a more accurate representation of the data and by better distinguishing between normal and anomalous patterns. As a result, ALAD is particularly useful in applications involving high-dimensional and complex datasets, such as image analysis and fraud detection, where subtle anomalies may otherwise be overlooked.

3.6 ML-Consensus

Machine learning consensus, often referred to as ensemble learning, is a powerful approach in which multiple machine learning models are combined to improve the overall performance of predictive analytics tasks. The fundamental idea behind ML-consensus is that by aggregating the predictions of several models, the combined output can often surpass the performance of individual models. This approach is particularly beneficial in the context of outlier detection, where the variability and complexity of data can present significant challenges to single models (Dietterich, 2000; Polikar, 2006).

The ML-consensus method leverages the strengths of diverse algorithms to achieve a more robust and reliable detection system. Each model in the ensemble may capture different aspects of the data, and by integrating their outputs, the ensemble can mitigate the weaknesses of individual models. For example, some models might be more sensitive to certain types of anomalies, while others may excel in identifying different patterns. By combining these models, the ensemble can provide a more comprehensive detection mechanism.

Mathematically, the ML-consensus approach can be expressed as follows. Let $f_1, f_2, \ldots, f_M$ represent the individual models in the ensemble, where M is the total number of models. The final prediction $\hat{y}$ for a given input x can be obtained by aggregating the predictions of all models:

$$\hat{y} = Aggregate(f_1(x), f_2(x), \ldots, f_M(x))$$

The aggregation function can vary depending on the specific ensemble method used. Common aggregation strategies include majority voting for classification tasks, averaging for regression tasks, and weighted combinations where models are assigned different weights based on their performance (Dietterich, 2000).

A particularly useful aggregation method in the context of outlier detection is the weighted voting

scheme. Here, each model f_i is assigned a weight w_i that reflects its reliability or confidence in making predictions. The final prediction is then computed as:

$$\hat{y} = \sum_{i=1}^{M} w_i f_i(\mathbf{x})$$

where the weights w_i are typically normalized so that $\sum_{i=1}^{M} w_i = 1$.

Among the various ensemble learning techniques, bagging (Bootstrap Aggregating) involves training multiple instances of the same model on different subsets of the training data, created by sampling with replacement from the original dataset. The predictions of the individual models are then aggregated, typically by majority voting for classification or averaging for regression, helping to reduce variance and prevent overfitting (Breiman, 1996). Boosting sequentially trains models, with each model attempting to correct the errors of its predecessor. The final prediction is a weighted sum of the predictions from all models, with algorithms like AdaBoost and Gradient Boosting showing significant improvements in accuracy by focusing on difficult-to-predict cases (Freund and Schapire, 1997; Friedman, 2001).

However, our primary focus will be on stacking, voting, random forests, and meta-learning. Stacking involves training multiple base models and then using their predictions as inputs to a meta-model, which makes the final prediction. This approach is particularly effective when base models are diverse, as the meta-model learns how to best combine their predictions, often leading to enhanced performance (Wolpert, 1992). In the voting ensemble method, each model in the ensemble casts a "vote" for each prediction. For classification tasks, the final prediction is the class with the majority of votes. For regression tasks, the final prediction is the average of the individual model predictions. Voting can be simple, with equal weights, or weighted, where models with higher accuracy contribute more to the final prediction, making it a versatile and straightforward aggregation method (Dietterich, 2000).

Random forests extend bagging by using decision trees as the base models and introducing randomness in feature selection for each split in the decision trees. This results in a diverse set of trees that collectively improve prediction accuracy and reduce overfitting (Breiman, 2001). Focusing on stacking, voting, and random forests is justified by their proven effectiveness and flexibility in various applications, including outlier detection. These methods allow for combining the strengths of different models, enhancing robustness, and improving predictive performance. Meta-learning, in particular, provides a framework for integrating the outputs of multiple models, enabling the development of a highly adaptive and resilient outlier detection system.

For bagging, the training sets $\{D_1, D_2, \ldots, D_M\}$ are created by sampling with replacement from the original dataset D. The final prediction for an input x is given by:

$$\hat{y} = \frac{1}{M} \sum_{i=1}^{M} f_i(\mathbf{x})$$

For boosting, each model f_i is trained on a weighted version of the dataset, where the weights are adjusted based on the errors of the previous model. The final prediction is a weighted sum of the model predictions:

$$\hat{y} = \sum_{i=1}^{M} \alpha_i f_i(\mathbf{x})$$

$i=1$

where α_i are the weights determined during training. For stacking, the base models $f_1, f_2, \ldots, f_M$ are first trained, and their predictions are used to train a meta-model g. The final prediction is given by:

$$\hat{y} = g(f_1(x), f_2(x), \ldots, f_M(x))$$

For voting, if C_k represents the prediction of the k-th class, the final prediction for classification is:

$$\hat{y} = \arg\max_k \sum_{i=1}^{M} w_i 1_{f_i(x)=C_k}$$

For random forests, each decision tree T_i is trained on a different bootstrap sample, and the prediction is:

$$\hat{y} = \frac{1}{M} \sum_{i=1}^{M} T_i(x)$$

Building on this understanding of ML-consensus and the mathematical foundations underpinning ensemble methods, we now shift our focus to the specific techniques that form the theoretical backbone of ensemble learning: stacking, voting, random forests, and meta-learning. These methods have been selected due to their established roles in enhancing predictive accuracy and reliability by combining multiple models. In the subsequent discussion, we will delve into the theoretical underpinnings of each approach, examining how they synergize different models to improve performance. By gaining a comprehensive understanding of these ensemble techniques, we lay the groundwork for advanced applications in outlier detection and beyond.

3.6.1 Voting

Voting is a fundamental ensemble learning technique designed to enhance the accuracy and robustness of predictive models by aggregating the predictions of multiple base models. The core idea behind voting is to leverage the diversity among various models to improve the overall performance of the ensemble. This technique harnesses the collective wisdom of the base models, which often leads to better generalization than any single model alone.

This approach is rooted in the principle that different models may capture different aspects of the data or have varying strengths and weaknesses. By combining their predictions, we can exploit their complementary capabilities and reduce the likelihood of errors made by individual models. This approach is particularly valuable in scenarios where no single model performs optimally across all aspects of the data.

In practice, voting involves two main strategies: **majority voting** and **weighted voting**. These strategies determine how the predictions of the base models are combined to arrive at a final decision.

In majority voting, each base model in the ensemble casts a "*vote*" for a particular class. The final prediction is based on the class that receives the most votes. Mathematically, if an ensemble consists of n classifiers and each classifier M_i provides a prediction $\hat{y}_i$ for a given input x, the final prediction $\hat{y}$ is determined by:

$$\hat{y} = mode(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$$

where *mode* represents the most frequently predicted class among the base models. This method capitalizes on the idea that even if individual models have varying levels of accuracy, the majority vote tends to represent the consensus view and thus can improve reliability (Breiman, 1996).

Weighted voting extends this concept by assigning different weights to each base model based on its performance or reliability. Instead of treating each model's vote equally, weighted voting accounts for the varying quality of predictions from different models. If w_i denotes the weight assigned to the i-th model and p_i represents its predicted probability for a class, the final prediction is computed as:

$$\hat{y} = argmax \left(\sum_{i=1}^{n} w_i \cdot p_i \right)$$

In this formulation, the final prediction is determined by the class that has the highest weighted sum of probabilities. Weighted voting is particularly useful when some models in the ensemble are known to be more accurate or reliable than others. By adjusting the weights, we can enhance the ensemble's performance and address model-specific biases or errors (Kuncheva, 2004).

The strength of voting lies in its ability to aggregate diverse predictions and reduce overall variance. Each base model in the ensemble may have unique errors or biases, but by combining their outputs, voting can average out these inconsistencies and produce a more stable and accurate final prediction. This principle aligns with the concept of "*wisdom of crowds*", where collective decisions tend to be more accurate than individual judgments (Hansen and Salamon, 1990).

In classification tasks, voting helps in achieving a consensus decision among the base models, while in regression tasks, a similar approach can be used to average the predictions. Despite its simplicity, voting can effectively address various challenges in predictive modeling, such as model overfitting and underfitting, by incorporating the strengths of multiple models and mitigating their weaknesses.

However, voting also has its limitations. For instance, if the base models are highly correlated or if they consistently make similar errors, the benefits of voting may be diminished. To overcome these challenges, it is essential to ensure that the base models are diverse and independently trained. Proper selection and weighting of models can further enhance the effectiveness of the voting ensemble.

Having explored the fundamentals of voting, an ensemble technique that leverages the collective predictions of multiple base models to enhance accuracy, we now turn our attention to another advanced method: stacking. While voting aggregates the outputs of various models to achieve a consensus, stacking builds upon this idea by introducing a meta-model that learns to optimally combine the predictions of base models. This approach not only harnesses the strengths of diverse algorithms but also refines the final predictions through a learned integration process. The transition from voting to stacking represents a progression from simple aggregation to a more sophisticated, adaptive method that enhances predictive performance through a meta-learning framework.

3.6.2 Stacking

Stacking, or stacked generalization, is a sophisticated ensemble learning technique designed to

enhance predictive performance by leveraging multiple base models through a meta-model. This

approach harnesses the distinct advantages of various algorithms to improve overall accuracy and mitigate the weaknesses inherent in individual models.

At the core of stacking is the concept of training several base models, referred to as level-0 models, on the same dataset. These base models can be diverse in nature, such as decision trees, support vector machines, or neural networks, each capturing different patterns and relationships within the data (Wolpert, 1992). Once trained, these base models generate predictions for the dataset. These predictions serve as input features for a meta-model, or level-1 model. The meta-model, often a simpler algorithm like linear regression or logistic regression, learns to optimally combine the outputs of the base models to produce a final prediction (Breiman, 1996).

Mathematically, let D denote the training dataset and $M_1, M_2, \ldots, M_n$ represent the base models. Each base model M_i generates a prediction $\hat{y}_i$ for a given input x. The meta-model H uses these predictions $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ as features to produce the final prediction $\hat{y}$, expressed as:

$$\hat{y}_i = M_i(x), \qquad for\ i = 1, 2, \ldots, n$$

$$\hat{y} = H(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$$

In this formulation, H is the meta-model that synthesizes the outputs from the base models to generate the final prediction.

The strength of stacking lies in its ability to capitalize on the diversity and complementary strengths of the base models. By aggregating models that each capture different facets of the data, stacking reduces overall bias and variance, leading to improved generalization on new, unseen data (Sill et al., 2009). The meta-model plays a pivotal role by learning the optimal combination of base model predictions, addressing the specific errors of individual models.

Stacking has proven effective across a range of domains, including both classification and regression tasks. It is particularly valuable in scenarios where no single model performs optimally across all cases, as the ensemble approach can encompass a broader spectrum of data characteristics and often surpasses any individual model's performance (Krogh and Vedelsby, 1995).

Nonetheless, stacking is not without challenges. One potential issue is overfitting, particularly if both the base models and the meta-model are excessively complex. To mitigate this risk, techniques such as cross-validation are utilized. In cross-validation stacking, the training dataset is divided into several folds, with each base model being trained on different subsets of the data. The meta-model is then trained using out-of-fold predictions, which provides a more robust and unbiased estimate of the base models' performance (Breiman, 1996).

Building on the concept of voting, which aggregates predictions from multiple models, we now explore Random Forest, an advanced ensemble technique that enhances voting by using decision trees and introducing randomness in model training. Random Forest not only combines predictions but also improves model diversity and accuracy through methods like bagging and feature randomness.

3.6.3 Random Forest

Random Forest is a widely used ensemble learning method that combines the predictions of multiple decision trees to improve overall model accuracy and robustness. This approach builds on the concept of bagging, or bootstrap aggregating, which involves training multiple models on

different subsets of the training data and then averaging their predictions to reduce variance and improve generalization.

In a Random Forest, multiple decision trees are constructed using different random subsets of the training data and features. Each tree is trained on a bootstrap sample, a random sample of the original dataset with replacement. During the training of each decision tree, only a random subset of features is considered for splitting at each node. This introduces diversity among the trees and helps prevent overfitting, which can occur if a single decision tree is too complex and fits the noise in the training data.

Mathematically, let D represent the original training dataset, and let $T_1, T_2, \ldots, T_B$ denote the B decision trees in the Random Forest. For a given input x, each tree T_i produces a prediction $\hat{y}_i$. The final prediction $\hat{y}$ is obtained by aggregating the predictions of all the trees. For classification tasks, the aggregation is typically done using majority voting, while for regression tasks, it is done by averaging the predictions:

$$\hat{y} = \frac{1}{B} \sum_{i=1}^{B} \hat{y}_i$$

where $\hat{y}_i$ is the prediction from the i-th tree, and B is the total number of trees in the forest (Breiman, 2001).

The strength of Random Forest lies in its ability to reduce both variance and bias compared to individual decision trees. By averaging the predictions of multiple trees, Random Forest mitigates the overfitting problem that can arise with complex decision trees. Additionally, the random feature selection process helps to capture different aspects of the data and improve the robustness of the model.

Random Forest has been shown to perform well in various domains, including finance, healthcare, and image analysis. Its ability to handle large datasets with many features and provide robust predictions makes it a valuable tool in machine learning. However, despite its advantages, Random Forest can be computationally intensive, particularly with a large number of trees and features, and may lack interpretability compared to simpler models.

The introduction of meta-learning builds upon the previous discussion of ensemble methods like voting, stacking, and Random Forest. By incorporating meta-learning, we further explore advanced strategies for enhancing machine learning models' adaptability and efficiency, paving the way for more sophisticated and versatile approaches in outlier detection and other domains.

Having examined Random Forest's approach to boosting predictive accuracy through decision trees and ensemble methods, we now turn to meta-learning. While Random Forest focuses on combining models, meta-learning enhances the learning process itself by optimizing how algorithms adapt and generalize across different tasks.

3.6.4 Meta-Learning

Meta-learning, also known as "*learning to learn*", is a sophisticated approach in machine learning that focuses on optimizing learning algorithms themselves. Rather than solely training a model on a specific task, meta-learning aims to improve the learning process by leveraging experience gained from multiple tasks or datasets. The goal is to develop models that can quickly adapt to new tasks with minimal additional training, effectively transferring knowledge across different learning scenarios.

The core idea behind meta-learning is to design algorithms that can learn from their previous learning experiences. This involves two levels of learning: the first level focuses on training models on various tasks, while the second level involves optimizing the learning process based on the performance of these models. Meta-learning algorithms typically operate by learning a set of parameters or strategies that enhance the generalization capability of the models they train.

Mathematically, let T represent a distribution over tasks, and let D_t denote the dataset for task t. Meta-learning aims to find a learning algorithm or set of parameters θ that optimizes performance across all tasks in T. Given a model f with parameters θ, the meta-learning objective is to minimize the loss function L over the distribution of tasks:

$$\theta^* = \arg\min_{\theta} \mathbb{E}_{t\sim T} [L(f_\theta, D_t)]$$

Here, θ^* represents the optimal parameters that generalize well across different tasks, and L is the loss function measuring the model's performance on each task t.

One popular meta-learning approach is Model-Agnostic Meta-Learning (MAML), which seeks to find initial model parameters that can be quickly adapted to new tasks with a few gradient updates. MAML optimizes the model's parameters such that, after a few steps of gradient descent on a new task, the model's performance is significantly improved. The optimization problem for MAML can be formulated as:

$$\theta^* = \arg\min_{\theta} \mathbb{E}_{t\sim T} [L(f_{\theta-\alpha\nabla_\theta L(f_\theta, D_t)}, D_t)]$$

where α is the learning rate for the task-specific updates (Finn et al., 2017).

Another approach is the use of meta-architectures, where the meta-learning process involves optimizing the structure or architecture of the learning model itself. This approach can involve techniques such as neural architecture search, where the meta-learner explores different network architectures to find the most effective configuration for a given task (Zoph and Le, 2017).

Meta-learning has shown promise in a range of applications, including few-shot learning, where models must generalize from a limited number of examples, and hyperparameter optimization, where the goal is to find the best hyperparameters for machine learning models. By focusing on improving the learning process itself, meta-learning offers a powerful framework for enhancing model performance and adaptability across diverse tasks.

3.7 Data Synthesis

Data synthesis is a critical aspect of developing and evaluating machine learning models, particularly for tasks like outlier detection. By generating synthetic data, researchers can create controlled datasets that are essential for training, validating, and testing models. This approach is invaluable when working with high-dimensional data or when real data is limited or difficult to access.

One of the primary advantages of data synthesis is its ability to overcome challenges related to data scarcity. In many real-world scenarios, outliers are rare and often underrepresented in available datasets. Synthetic data allows for the creation of large and diverse datasets with known properties, ensuring that models are trained on a broad spectrum of anomaly types. This exposure is crucial for developing robust models capable of identifying subtle and varied anomalies.

Synthetic data also enables the exploration of different data distribution scenarios that may be challenging to capture with real data. For instance, certain anomalies might be too rare or complex to be adequately represented in natural datasets. By generating synthetic anomalies with specific characteristics, researchers can test and refine models under controlled conditions, gaining insights into their performance and robustness.

Several methods exist for generating synthetic data, each suited to different types of distributions and applications. The Box-Muller transform, a classic technique, generates synthetic data from a normal distribution. Given two independent random variables U_1 and U_2 uniformly distributed in the interval [0, 1), the Box-Muller transform produces two independent standard normal variables Z_0 and Z_1 using:

$$Z_0 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$
$$Z_1 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

While effective for generating normal distributions, this method is limited to scenarios where the data distribution is Gaussian (Box and Muller, 1958).

To address a broader range of data distributions, other statistical models can be employed. For instance, Multivariate Gaussian distributions allow for modeling data with specified correlations between variables.

If represents the covariance matrix of the distribution, synthetic data can be generated using:

$$X = \mu + LZ$$

where μ is the mean vector, L is the Cholesky decomposition of , and Z is a vector of standard normal random variables (Trefethen and Bau, 1997). This approach enables the creation of datasets with specific statistical properties.

Cholesky decomposition is a matrix factorization technique used to decompose a symmetric positive-definite matrix into a product of a lower triangular matrix and its transpose. For a symmetric positive-definite matrix A, Cholesky decomposition finds a lower triangular matrix L such that:

$$A = LL^T$$

where L^T denotes the transpose of L. The matrix L has zeros above the diagonal and non-zero elements on and below the diagonal. To compute the elements of L, the following recursive formulas are used:

For each diagonal element L_{ii}:

$$L_{ii} = \sqrt{A_{ii} - \sum_{k=1}^{i-1} L_{ik}^2}$$

For each off-diagonal element L_{ij} (where $i > j$):

$$L_{ij} = \frac{1}{L_{jj}} \left(A_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk} \right)$$

These formulas ensure that L is a lower triangular matrix with positive diagonal entries (Golub and Van Loan, 2013). Cholesky decomposition is particularly useful for generating multivariate normal data with specified covariance structures, enhancing its applicability in various fields.

Generative Adversarial Networks (GANs) have also revolutionized synthetic data generation. GANs consist of two neural networks—the generator and the discriminator—that are trained adversarially. The generator creates synthetic data samples, while the discriminator assesses their realism. This adversarial training encourages the generator to produce data that closely resembles the real distribution, making GANs effective for generating complex, high-dimensional data (Goodfellow et al., 2014).

Variational Autoencoders (VAEs) are another powerful tool for data synthesis. VAEs use a probabilistic approach to model data's latent space, allowing for the generation of new samples by sampling from the learned latent distribution. The VAE loss function combines reconstruction loss with the Kullback-Leibler (KL) divergence between the learned latent distribution and a prior distribution:

$$ELBO = \mathsf{E}_{q(\mathrm{z}|\mathrm{x})}[\log p(\mathrm{x}|\mathrm{z})] - KL(q(\mathrm{z}|\mathrm{x})\not\|p(\mathrm{z}))$$

where $q(\mathrm{z}|\mathrm{x})$ is the approximate posterior, $p(\mathrm{x}|\mathrm{z})$ is the likelihood, and $p(\mathrm{z})$ is the prior (Kingma and Welling, 2013). This approach facilitates the generation of data with complex structures and variations.

Additionally, methods such as SMOTE (Synthetic Minority Over-sampling Technique) address class imbalance by creating synthetic samples through interpolation between existing data points. SMOTE is particularly useful for enhancing the representation of minority classes, thereby improving model performance on imbalanced datasets (Chawla et al., 2002).

In summary, data synthesis is a vital component of machine learning, offering techniques to generate controlled datasets with specific properties. By overcoming data scarcity and distribution variability challenges, synthetic data generation enhances model development and evaluation, providing a foundation for robust and effective machine learning solutions.

3.8 Conclusion

In conclusion, this chapter has provided a thorough exploration of outlier detection methodologies, encompassing both classical statistical techniques and contemporary deep learning approaches. We examined methods such as Mahalanobis Distance, Isolation Forest, Autoencoders, and Adversarially Learned Anomaly Detection (ALAD), revealing their strengths and limitations. These insights are essential for understanding how each method can be effectively applied to different data types and anomaly detection scenarios (Iglewicz & Hoaglin, 2007; Hsieh & Lee, 2020; Mallick & Kim, 2019; Kingma & Welling, 2014; Zenati et al., 2018).

Additionally, the chapter delved into the concept of ML-consensus, emphasizing the integration of multiple machine learning techniques to enhance predictive performance. By discussing approaches such as stacking (Wolpert, 1992; Breiman, 1996), voting (Freund & Schapire, 1997), random forests (Breiman, 2001), and meta-learning (Balcázar et al., 2012), we highlighted how combining diverse models can address the limitations of individual methods and improve overall anomaly detection accuracy.

This critical analysis lays a robust foundation for the subsequent development and implementation

of outlier detection systems. The insights gained underscore the necessity of a nuanced approach that integrates various methodologies, including the ML-consensus framework, to handle the complexities of real-world data effectively. This understanding will be pivotal in designing a comprehensive detection framework capable of managing diverse and high-dimensional datasets.

The theoretical insights from this chapter will guide the solution engineering and implementation detailed in the following chapters. Chapter 4 will leverage these foundations to design and optimize an advanced outlier detection system, while Chapter 5 will focus on the practical aspects of deploying and integrating this system.

By addressing the identified challenges and utilizing the theoretical knowledge and ML-consensus approach discussed here, the study aims to enhance the effectiveness and reliability of outlier detection systems. This foundational analysis will support the development of solutions that are both theoretically robust and practically viable, ensuring their applicability and impact in real-world scenarios.

Chapter 4

Solution Engineering

In this chapter, we transition from theoretical exploration to the practical implementation of the outlier detection system discussed in earlier chapters. Building on the foundational knowledge gained, this chapter aims to detail the engineering aspects of the solution, ensuring that theoretical insights are effectively translated into a functional and robust system.

The primary objective of this chapter is to outline the comprehensive solution architecture and implementation strategy for the outlier detection system. This involves designing and documenting the system's architecture, selecting appropriate technologies, and analyzing the solution's performance. By systematically addressing these aspects, we aim to create a solution that not only adheres to the theoretical principles discussed but also performs effectively in real-world scenarios. We will begin by conceptualizing the solution architecture, which will involve creating use case diagrams, sequence diagrams, activity diagrams, and workflow engineering. These tools will help visualize the system's functionality and interactions, providing a clear blueprint for the subsequent development stages.

Following the architectural design, we will delve into the choice of technologies, justifying the selection based on the project's requirements. This section will provide a comprehensive overview of the technologies employed, detailing their roles and contributions to the solution.

Data synthesis will then be addressed, focusing on the structure and preparation of data essential for the system's operation. Understanding how data is organized and processed is crucial for ensuring the accuracy and efficiency of the outlier detection algorithms.

The solution analysis section will explore the implementation of the desktop application, including the choice of algorithms, complexity analysis, and a comparative study of algorithm performance. This analysis will provide insights into the effectiveness of the solution and highlight any potential areas for improvement.

Finally, the chapter will conclude with a summary of the key findings and reflections on the integration of theoretical insights into practical implementation. This conclusion will also discuss future directions for enhancing the solution and addressing any remaining challenges.

By systematically addressing these components, this chapter aims to bridge the gap between theory and practice, ensuring that the outlier detection system is both theoretically sound and practically viable.

4.1 Solution Architecture

The solution architecture is a critical component of system design, serving as the blueprint that outlines the structural framework and interactions within the outlier detection system. This architecture is pivotal in translating theoretical concepts into a tangible and functional system, ensuring that all components work together seamlessly to achieve the desired outcomes. By defining the system's structure and interactions, the architecture helps in creating a well-organized and efficient solution that adheres to the principles outlined in the theoretical background.

In system design, particularly within the context of Unified Modeling Language (UML), the solution architecture employs various diagrams to represent the system's components and their relationships. UML provides a standardized way to visualize system structure and interactions, which is essential for both understanding and communicating the design. Key UML diagrams, including use case diagrams, sequence diagrams, activity diagrams, and workflow engineering, play a crucial role in depicting different aspects of the system's architecture.

The importance of the solution architecture lies in its ability to provide a clear and structured framework for system development. It ensures that all components are aligned with theoretical principles and operational requirements, facilitating a smooth transition from design to implementation. By employing UML diagrams and engineering principles, the architecture provides a comprehensive overview of the system's structure and interactions, guiding the development process and ensuring that the final solution meets the intended goals.

4.1.1 Use Case Diagram

The Use Case Diagram is an integral part of the solution architecture, offering a high-level view of the interactions between users (actors) and the system. It is essential for capturing the functional requirements of the system and ensuring that all necessary features are incorporated into the design. The diagram serves as a visual representation of the system's functionality, illustrating how different actors interact with the system to achieve their goals.

In the context of this outlier detection system, the Use Case Diagram focuses on several key functionalities that are crucial for the system's operation. These functionalities are designed to address specific needs and tasks related to outlier detection and data analysis.

In our system, the Use Case Diagram includes several key functionalities that are pivotal for effective operation. These functionalities encompass parameterizing the model, choosing variables for outlier detection, data visualization, generating and exporting reports, each serving distinct purposes within the overall framework.

Actor: Admin (In this context, it refers mainly to the person in charge of maintaining the software and updating the model). **Actor**: User (member of the Data Analytics or Statistics Department).

Parameterizing the Model:

Users can configure the outlier detection model through an intuitive form. This process involves inputting various parameters necessary for model training and outlier detection. The parameters may include settings specific to the algorithms used, thresholds for anomaly detection, and other relevant configurations. This functionality ensures that the model can be tailored to meet the specific requirements of different datasets.

Choosing Variables for Outlier Detection:

Once the model is parameterized, users select the variables on which the outlier detection will be performed. This selection process is critical as it determines the focus of the analysis and ensures that the outlier detection process targets the most relevant features of the dataset.

Data Visualization:

The application provides visualization tools to display the data associated with the selected variables. This functionality allows users to gain insights into the data distribution and understand how the outlier detection model interacts with the dataset. Visualizations can include charts, graphs, and other visual aids that enhance data interpretation.

Generating and Exporting Reports:

After the outlier detection process is complete, users can generate detailed reports that highlight the identified outliers within the existing data. These reports are essential for communicating findings and supporting decision-making. The capability to export these reports in various formats ensures that users can share results with stakeholders effectively.

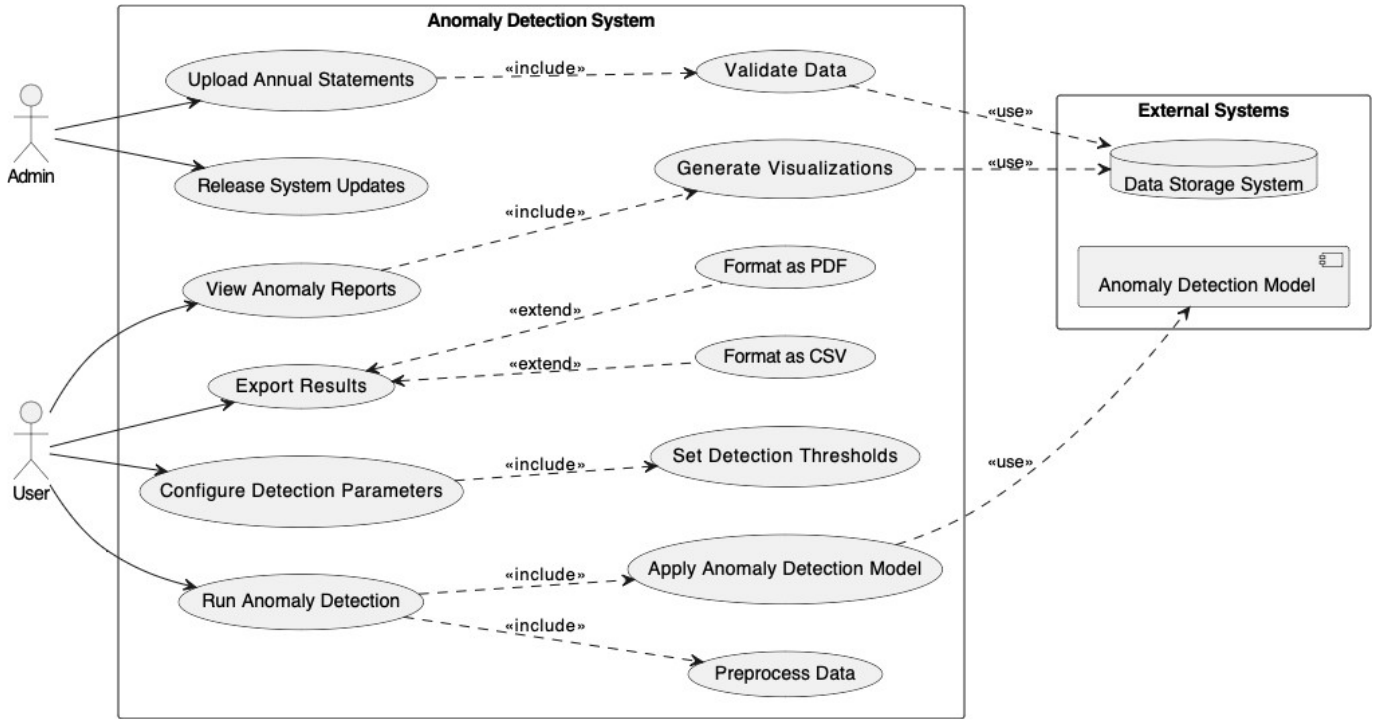


Figure 4.1 – Use Case Diagram

The use case diagram illustrates how these functionalities are interconnected and how users interact with the system to achieve their objectives. It maps out the process flow from parameter configuration to report generation, ensuring that all components work together seamlessly to provide a comprehensive solution for outlier detection.

This diagram not only helps in understanding the user interactions but also serves as a foundation for designing the system's architecture and user interface. By clearly defining the system's functionalities and their interactions, the use case diagram facilitates a structured approach to developing and implementing the application, ensuring that all user needs are met efficiently and effectively (Fowler, 2004; Rumbaugh et al., 1999).

4.1.2 Class Diagram

The class diagram is a vital component of the Unified Modeling Language (UML) used to describe the static structure of a system by showcasing its classes, attributes, methods, and the relationships among objects. In this context, the class diagram will represent the core components of the outlier detection system, illustrating how different entities interact within the application to achieve the desired functionality.

In this context, the class diagram delineates the primary classes involved in the application, including Model, Data, Visualization, and Report classes. Each class encapsulates specific functionalities and attributes that contribute to the overall operation of the system.

The class diagram provides a comprehensive overview of the system's structure, highlighting the key components and their interactions. This visual representation is essential for developers to understand the system's architecture, identify potential areas for optimization, and ensure that all necessary functionalities are incorporated. It also serves as a blueprint for implementing the system, guiding the development process from initial design to final deployment (Booch et al., 1998; Rumbaugh et al., 2004).

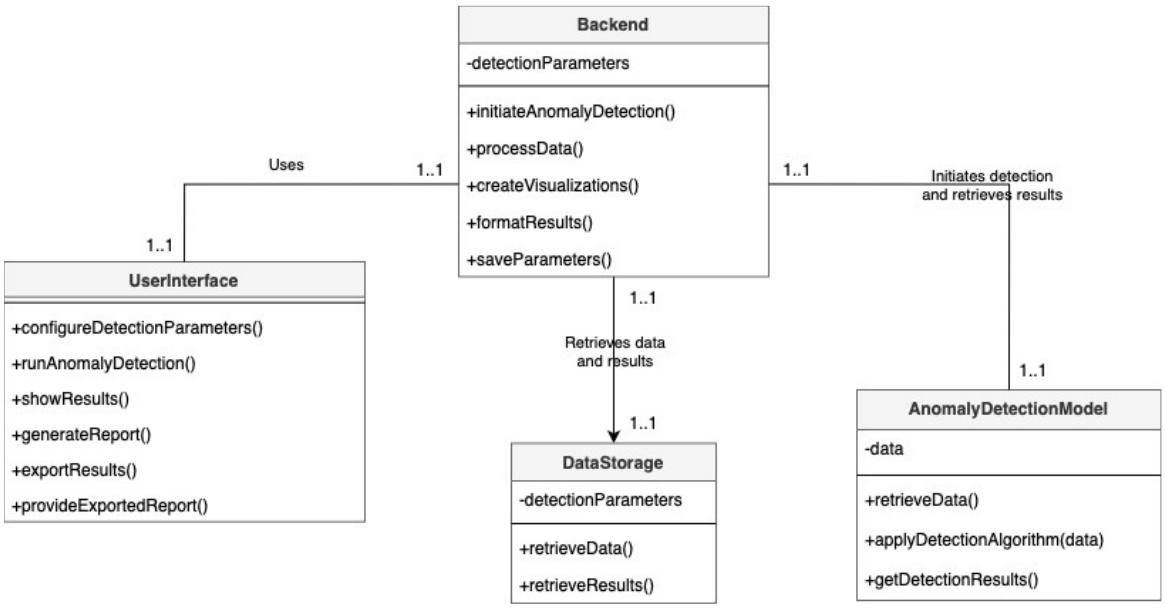


Figure 4.2 – Class Diagram

4.1.3 Sequence Diagram

The sequence diagram is an essential tool for detailing the interactions between different components of the system over time. It provides a step-by-step visualization of how processes are executed, illustrating the flow of information and the sequence of operations. In our context, the sequence diagram will focus on the interactions required to perform outlier detection, from parameter configuration to report generation.

In this context, the sequence diagram captures the dynamic behavior of the system as users engage with various functionalities. The primary actors involved are the user and the system components, including the user interface, the data processing module, and the reporting module.

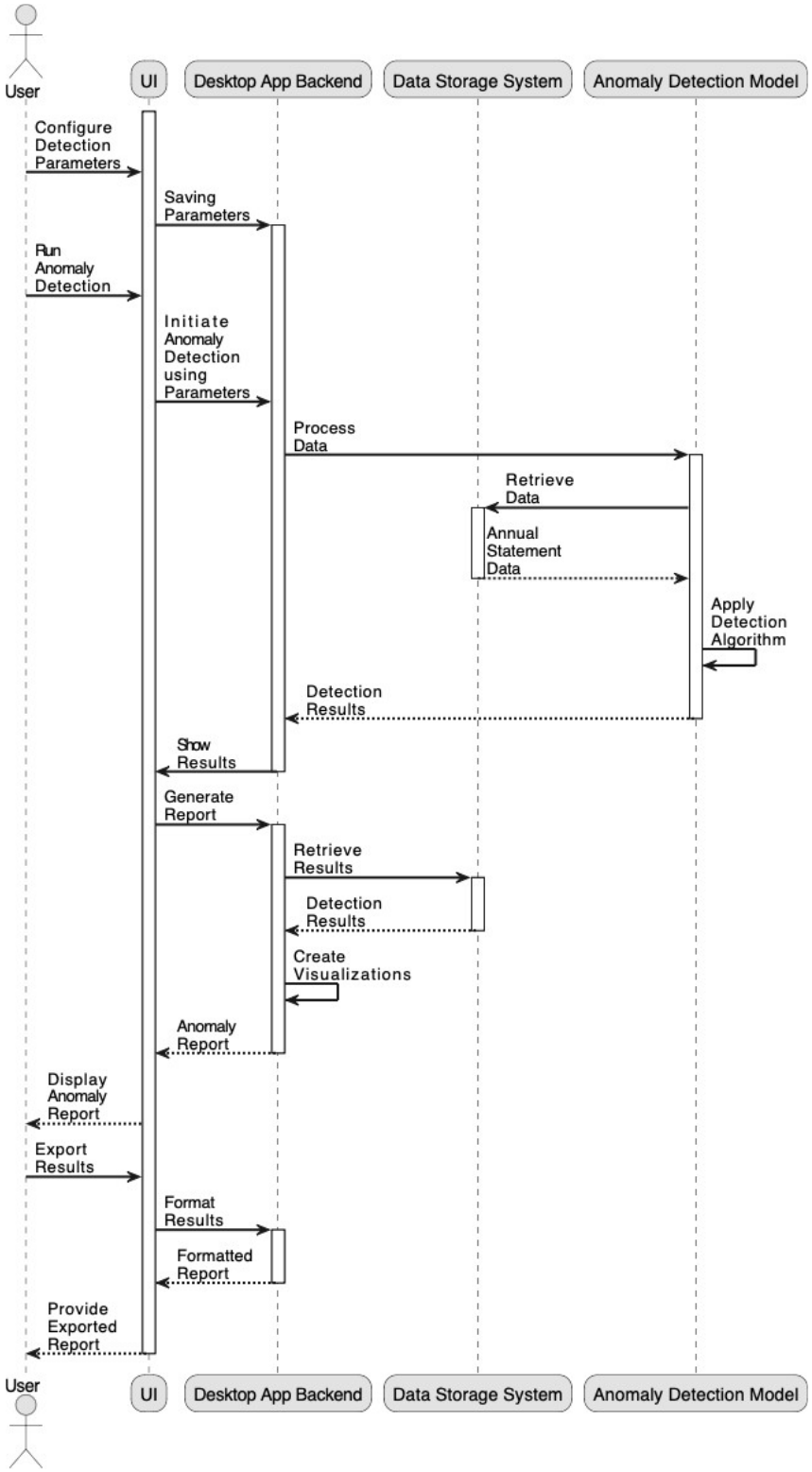


Figure 4.3 – Sequence Diagram

This comprehensive visualization aids in understanding the intricacies of system interactions, facilitating the identification of potential bottlenecks and ensuring the robustness of the implementation. Furthermore, it serves as a valuable reference for developers and stakeholders, providing a clear roadmap for system development and integration (Jacobson et al., 1999; Rumbaugh et al., 2005).

4.1.4 Activity Diagram

The activity diagram is a critical tool for mapping out the workflow and activities within the system, providing a detailed visualization of the process flow from start to finish. It showcases the sequence of activities, decision points, and the flow of control between different stages of the system's operation. In our context, the activity diagram will illustrate the entire workflow involved in outlier detection, from initial user interaction to the final report generation.

In this context, the activity diagram begins with the user initiating the process by accessing the application. The first activity involves the user navigating to the model parameterization form. Here, the user inputs various parameters required to configure the outlier detection model. This step involves multiple activities, including form completion, parameter validation, and submission. The system checks the validity of the input parameters and proceeds to configure the model accordingly.

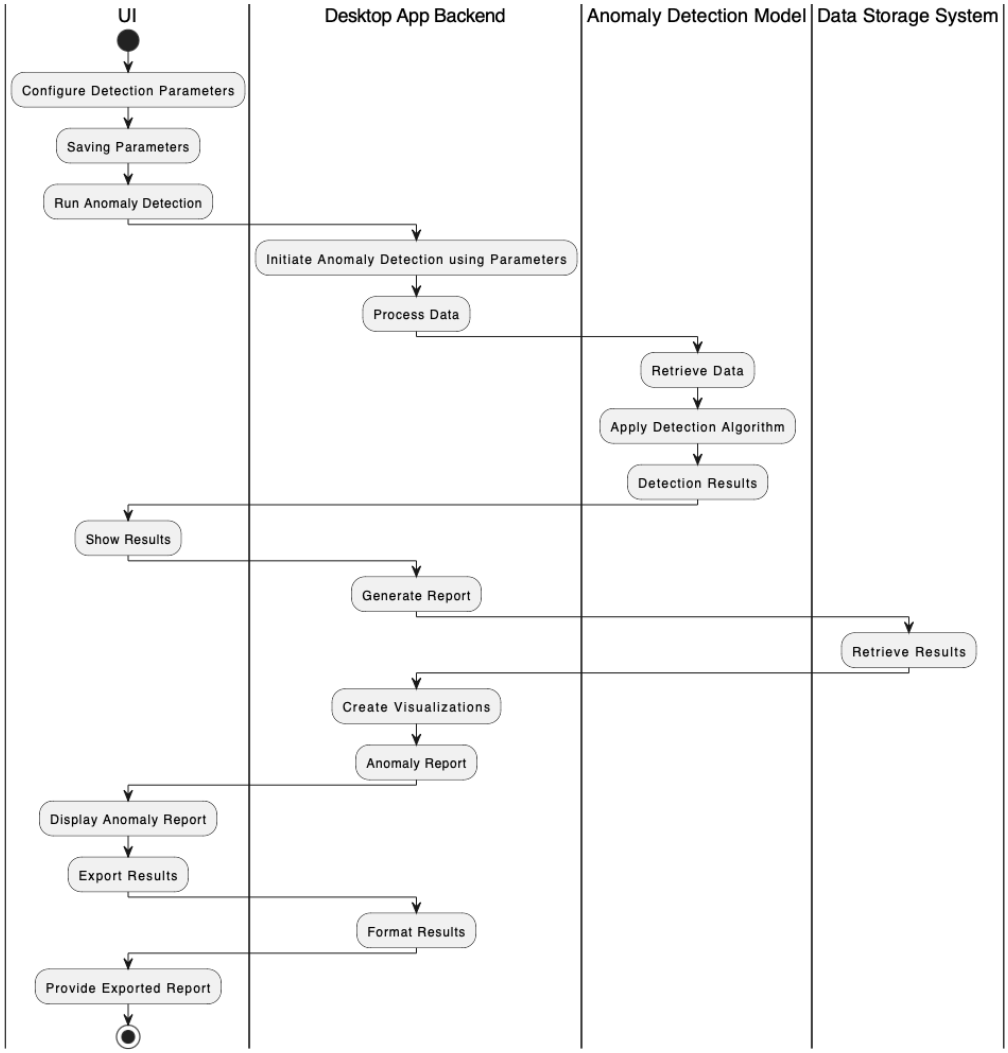


Figure 4.4 – Activity Diagram

The activity diagram emphasizes the sequential and parallel activities within the workflow, illustrating the flow of control from one activity to the next. It highlights decision points where user input or system validation is required, ensuring that all activities are completed accurately and in the correct order.

Mathematically, we can represent some of the processes involved. For example, let X be the

input data matrix, and Y be the output of the detection model. The parameterization step can be represented as setting parameters θ for the model M:

$$M(X; \theta) \rightarrow Y$$

The activity diagram not only provides a clear depiction of the workflow but also highlights the interactions between different components and the user. This detailed visualization aids in understanding the sequence of operations, identifying potential areas for optimization, and ensuring the robustness of the system's implementation. It serves as a crucial reference for developers and stakeholders, guiding the development and integration process (OMG, 2011; Booch et al., 1998).

4.1.5 Workflow engineering

Workflow engineering is a crucial aspect of system design that focuses on defining, optimizing, and automating the flow of tasks and processes within an application. In the context of our outlier detection system, workflow engineering ensures that all interactions between different system components are streamlined to provide a seamless user experience. This section outlines the steps involved in designing an effective workflow for the outlier detection application, from parameter input to report generation and visualization.

In our context, the workflow begins with the user interacting with the system to input parameters for the outlier detection model. This interaction is captured in the initial step where the user fills out a form specifying the model parameters and selects the variables for outlier detection. Once the parameters and variables are submitted, the system initiates the data processing stage.

The data processing stage involves loading the preexisting processed data into the system. The data is then prepared for analysis by the model. This preparation includes any necessary transformations or normalizations to ensure that the data is in a suitable format for the outlier detection algorithms. The prepared data is then fed into the model, which runs the specified algorithms to identify potential outliers.

Mathematically, this can be represented as follows. Let $\boldsymbol{X}$ be the dataset, θ be the model parameters, and $\boldsymbol{f}$ be the outlier detection function. The process can be formalized as:

$$X_{processed} = preprocess(X) \rightarrow \ldots \rightarrow Y = f(X_{processed}, \theta)$$

where $X_{processed}$ represents the transformed data, and Y represents the outlier detection results.

Throughout this workflow, several interactions and dependencies need to be managed. The system must ensure that data is accurately passed between components and that each stage of the workflow is executed correctly. This requires robust error handling and validation mechanisms to address any issues that may arise during data processing, model execution, or report generation.

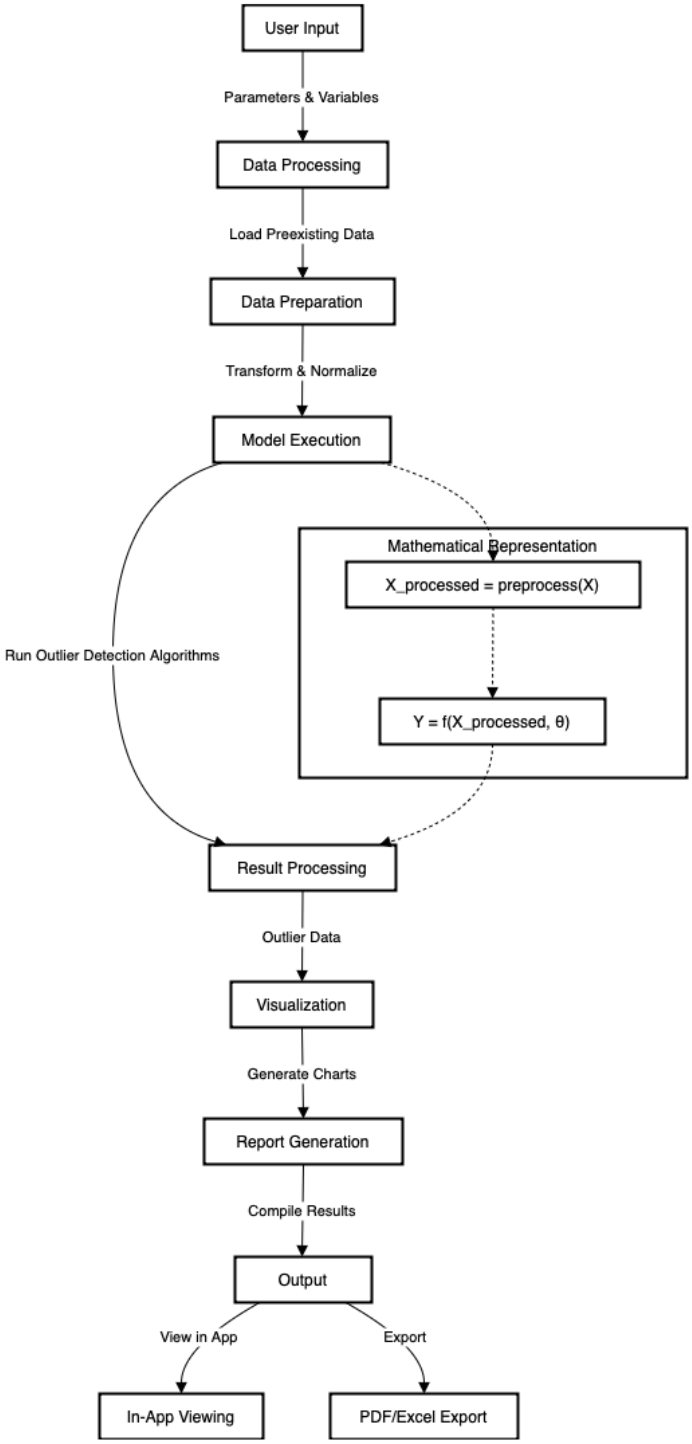


Figure 4.5 – Workflow

Workflow engineering for the outlier detection system aims to optimize these interactions to ensure efficiency and reliability. By defining clear steps and automating key processes, the system can provide a user-friendly experience while delivering accurate and actionable insights.

This structured approach to workflow engineering draws on best practices from system design and process optimization, ensuring that the outlier detection system is both effective and efficient (Smith, 2003; Sharp and McDermott, 2009). Through careful planning and execution, the workflow is designed to handle the complexities of real-world data and deliver reliable results to end-users.

4.2 Data Synthesis

Data synthesis is a critical process in managing sensitive financial information, particularly when preparing data for outlier detection. Given the sensitivity of the dataset, which includes annual financial statements from companies in Morocco, data synthesis ensures that raw data is accurately and securely organized for analysis.

The process starts with integrating various data sources to eliminate inconsistencies and align the data with the analysis goals. This integration helps create a comprehensive dataset, essential for identifying anomalies and trends (Pardoe, 2021). Subsequently, the data undergoes cleaning and preprocessing to correct errors and remove irrelevant information. This step ensures data accuracy and reliability, which is crucial for effective outlier detection (Chong et al., 2017).

Organizing data into meaningful categories—such as assets, liabilities, and revenue —facilitates detailed financial analysis and enhances the performance of outlier detection algorithms. Proper categorization ensures that data is structured in a way that supports insightful analysis (Iglewicz and Hoaglin, 2020).

Additionally, data synthesis involves safeguarding sensitive information by adhering to data protection regulations. Techniques like data anonymization and encryption are employed to prevent unauthorized access and maintain data confidentiality (Sweeney, 2021).

With the importance of accurate data synthesis established, it is crucial to first understand the foundational structure of the dataset used in this analysis. The dataset, composed of annual financial statements from Moroccan companies, serves as the basis for all subsequent data synthesis activities.

To effectively prepare this data for outlier detection, we must begin by examining its structure. This examination provides insight into how the data is organized, categorized, and what specific variables are included. Understanding the data structure is essential as it lays the groundwork for the detailed synthesis process, ensuring that data handling and preprocessing align with the dataset's inherent characteristics.

In the upcoming parts, we will delve into the actual data structure, detailing the columns and types of information present in the dataset. This structured overview will inform the synthesis process and facilitate the accurate preparation of data for further analysis.

4.2.1 Data Structure

The dataset utilized in this study comprises annual financial statements from companies based in Morocco. It includes a range of financial variables captured over multiple years, providing a comprehensive view of the financial health and performance of these entities. This dataset serves as a crucial foundation for outlier detection and financial analysis.

The dataset features several primary columns, each representing a distinct aspect of the financial statements:

Year: This column indicates the fiscal year of the financial statement, allowing for temporal analysis and trend evaluation (Higgins, 2018).

Short Name of Company: This column contains the abbreviated or tag name of the company, facilitating easy identification and comparison of different entities (Smith and Brown, 2019).

Financial Variables: The dataset encompasses various financial metrics and ratios critical for comprehensive financial analysis. These include:

- **Ratio**: This column lists financial ratios that provide insights into the company's financial stability and performance. Financial ratios, such as liquidity ratios, profitability ratios, and solvency ratios, are essential for assessing the company's operational effectiveness (White et al., 2020).
- **Actif_A, ..., Actif_tresor3**: These columns detail different asset categories and subcategories. Assets are typically classified into current and non-current, with further granularity into specific types such as cash, receivables, and inventories. Proper classification of assets is crucial for evaluating a company's resource allocation and financial health (Kieso et al., 2019).
- **CPC_I, ..., CPC_XIII**: This series of columns represents various components of cash flow, categorized into operational, investing, and financing activities. Understanding cash flow components is vital for analyzing a company's liquidity and cash management (Brigham and Ehrhardt, 2021).
- **ESG_CAF1, ..., ESG_TFRVIII**: These columns include Environmental, Social, and Governance (ESG) metrics, which reflect the company's performance in sustainable and ethical practices. ESG metrics are increasingly important for assessing long-term viability and societal impact (Eccles and Klimenko, 2019).
- **Passif_A, ..., Passif_tresor3**: These columns capture various liability categories and subcategories. Liabilities are classified into current and non-current, with details on specific obligations such as short-term debt, long-term debt, and provisions. Accurate liability reporting is critical for evaluating financial risk and leverage (Deegan, 2019).
- **charges_co**: This column includes various expenses or charges incurred by the company, essential for understanding cost structure and expense management (Penman, 2020).
- **produits_c**: This column lists the company's revenues or income, which is key for assessing profitability and revenue generation capabilities (Schroeder et al., 2019).

This rich dataset provides a valuable resource for outlier detection and financial analysis, enabling detailed examination of financial performance and identification of anomalies or irregularities in financial reporting.

4.2.2 Synthesis

To develop and validate the machine learning models, synthetic data was generated to simulate financial data and introduce controlled anomalies. The synthetic data generation process involved the following steps:

— **Generating Normal Data**: Synthetic data for testing was generated by sampling from a normal distribution. This data represents typical financial metrics without anomalies.
— **Introducing Outliers**: A small number of outliers were introduced into the synthetic dataset by sampling from a uniform distribution with a wider range. These outliers represent abnormal financial metrics.
— **Combining Normal Data and Outliers**: The normal data and the generated outliers were concatenated to form the complete synthetic dataset.
— **Shuffling the Dataset**: The combined dataset was shuffled to ensure that the outliers were randomly distributed throughout the dataset.
— **Saving the Dataset**: The synthetic dataset was saved to a CSV file for use in training and testing the machine learning models.

The following algorithm outlines the steps for evaluating different outlier detection algorithms. This evaluation involves applying various algorithms to the dataset, measuring their performance, and comparing the results based on accuracy, precision, recall, F1 score, and computational efficiency.

Algorithm 1: Synthetic Dataset Generation

Data: n_{samples}, n_{features}

Result: Synthetic Dataset D

1 **Generate** normal test data with n_{samples} samples and n_{features} features

2 **Generate** outliers with uniform distribution

3 **Concatenate** normal test data and outliers

4 **Shuffle** the combined dataset

5 **return new Dataset** D

This synthetic data generation approach allows for the creation of a controlled environment where the performance of the machine learning models can be rigorously tested and validated. By simulating realistic financial data and introducing anomalies, we can ensure that the models are robust and capable of accurately detecting outliers in real-world scenarios.

4.3 Solution Analysis

In the context of developing an outlier detection system, a thorough solution analysis is indispensable. This section provides a comprehensive evaluation of the implemented solution, examining its functionality, effectiveness, and efficiency in detecting outliers within the financial dataset. The objective is to scrutinize the chosen methodologies and their application to ensure they meet the system's requirements and performance standards.

Conducting a solution analysis is vital for verifying the accuracy and reliability of the detection algorithms employed. Financial datasets are inherently complex and sensitive, requiring methods that can precisely identify anomalies without generating false positives or negatives. Analyzing the solution's performance using metrics such as accuracy, precision, and the confusion matrix allows for a detailed understanding of how well the system can discern genuine outliers from normal data points. Ensuring high accuracy and reliability is crucial for maintaining data integrity and enabling informed decision-making based on the analysis.

Beyond accuracy, understanding the computational efficiency of the outlier detection methods is imperative. Financial institutions deal with large volumes of data, and the chosen algorithms must be capable of processing this data within acceptable time frames. Evaluating the time and space complexity of each method ensures that the system can handle extensive datasets without compromising performance. Efficient algorithms lead to quicker insights and more responsive systems, which are essential in the fast-paced financial sector. This aligns with the need to process data swiftly to keep up with market dynamics.

In the first part of our solution analysis, we evaluate the performance of several commonly used outlier detection methods: Isolation Forest (iForest), Local Outlier Factor (LOF), Simplified Local Outlier Factor (sLOF), Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), Machine-Learning and Deep-Learning Autoencoders, Generative Adversarial Networks (GAN), and Adver-

sarially Learned Anomaly Detection (ALAD). Each of these methods offers distinct advantages and presents unique implications for detecting anomalies in financial datasets, as introduced in Chapter 3, Theoretical Background.

Our evaluation of these methods involves assessing performance metrics such as accuracy, precision, recall, and computational efficiency. Time complexity is particularly crucial for methods like iForest and DBSCAN, where efficient processing is essential for handling large-scale financial data. Space complexity is also significant, especially for deep learning-based methods like GAN and ALAD, which require substantial memory resources for training and inference. For a detailed discussion on performance metrics, refer to the subsection on performance metrics in Chapter 3.

To provide a comprehensive evaluation, we conducted extensive experiments comparing the performance of these methods across key metrics. Our analysis focused on density-based methods for financial anomaly detection, including Simplified Local Outlier Factor (sLOF), Local Outlier Factor (LOF), Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), and RS-Forest. The following series of graphs illustrate their performance across critical metrics.

Firstly, we examine the execution time of each algorithm, a crucial factor for real-time financial anomaly detection. As depicted in Figure 4.6, significant differences in computational speed among the methods are evident. This performance disparity is a critical consideration when selecting the most suitable algorithm for time-sensitive applications. Breunig et al. (2000) emphasize that the efficiency of LOF makes it particularly suitable for large datasets, which is often the case in financial contexts.

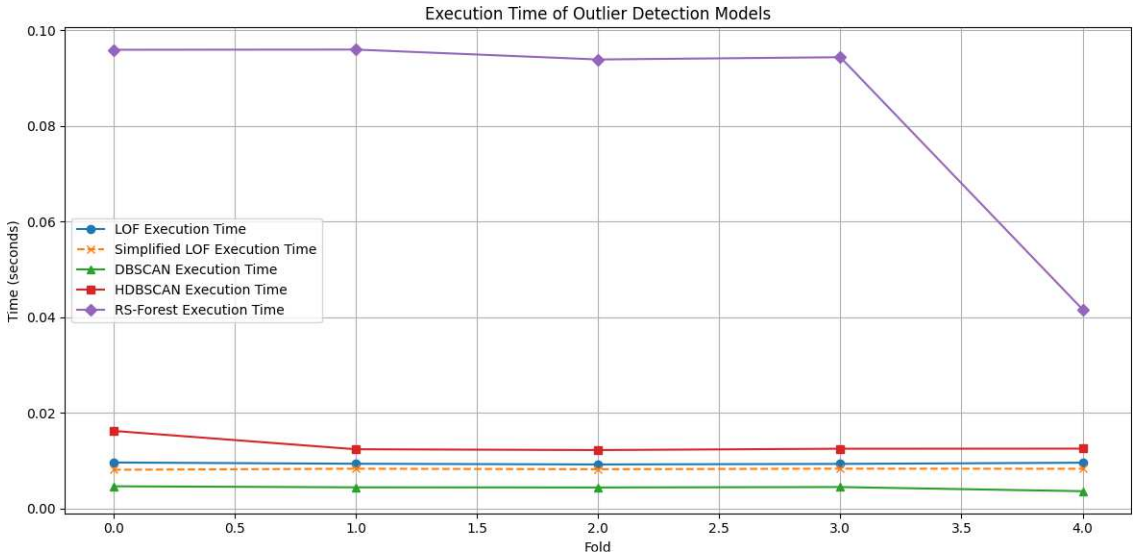


Figure 4.6 – Execution Time Comparison

Accuracy is another essential metric, as shown in Figure 4.7. While execution time is important, the overall correctness in anomaly identification is crucial. The trade-off between speed and accuracy must be considered when choosing an algorithm. As noted by Ester et al. (1996), the effectiveness of density-based methods can vary significantly depending on the dataset.

Precision, depicted in Figure 4.8, is particularly important in financial contexts to minimize false positives. A high precision rate ensures that flagged anomalies are likely genuine concerns rather than false alarms. This is crucial in financial anomaly detection, where false positives can lead to unnecessary investigations and potential reputational damage (Chandola et al., 2009).

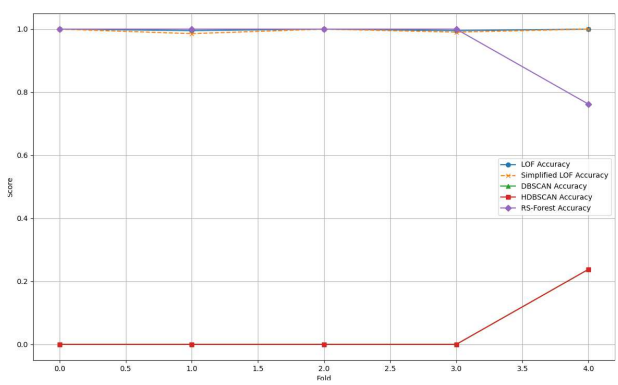


Figure 4.7 – Accuracy Comparison

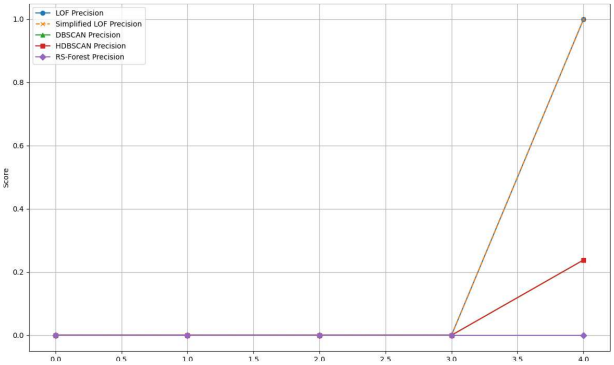


Figure 4.8 – Precision Comparison

Recall, as depicted in Figure 4.9, is a crucial metric that quantifies the ability of an algorithm to identify all true anomalies within a dataset. It measures the proportion of actual anomalies that the model successfully detects, providing insight into the algorithm's sensitivity to potential outliers. High recall is particularly vital in the context of financial systems, where the goal is to ensure that no fraudulent activities, financial irregularities, or other significant anomalies are overlooked. In financial applications, failing to identify true anomalies can result in substantial financial losses, regulatory non-compliance, and damage to the institution's reputation.

The importance of recall in financial anomaly detection cannot be overstated. It directly impacts the model's effectiveness in safeguarding against fraudulent transactions and ensuring the integrity of financial reporting. For instance, in scenarios where the cost of missing an anomaly is high, such as in detecting fraudulent financial transactions or major accounting discrepancies, achieving a high recall rate is paramount. This metric helps to balance the trade-off between detecting as many anomalies as possible and avoiding the risk of false negatives, thereby enhancing the model's reliability in identifying critical issues.

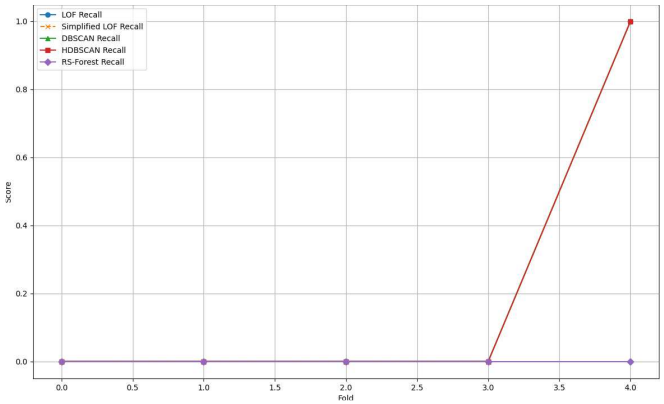


Figure 4.9 – Recall Comparison

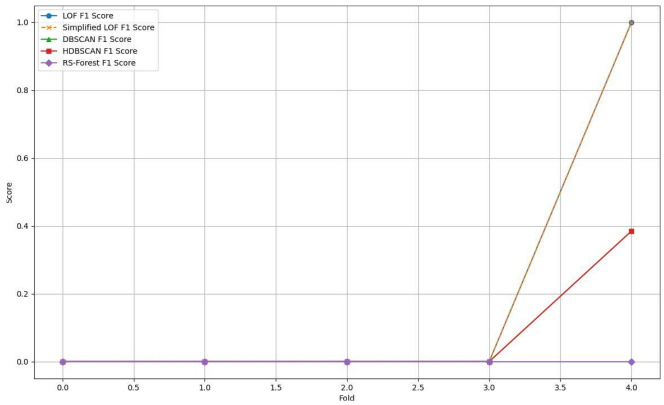


Figure 4.10 – F1-Score Comparison

To balance precision and recall, we analyze the F1 score in Figure 4.10. This metric provides a single score that balances the trade-offs between precision and recall, offering a comprehensive view of each algorithm's performance. The F1 score is particularly useful for assessing overall performance in imbalanced datasets common in anomaly detection (Powers, 2011).

Our results indicate that sLOF and LOF consistently outperform other density-based methods across multiple metrics. Their superior performance in accuracy, precision, and recall, along with reasonable execution times, makes them promising candidates for financial anomaly detection. This aligns with Goldstein and Uchida's (2016) findings on unsupervised anomaly detection algorithms. Conversely, DBSCAN's performance is notably poor across most metrics, suggesting it may not be suitable for this specific application. HDBSCAN and RS-Forest demonstrate varying strengths across metrics, indicating potential for specialized use cases where their characteristics align with specific requirements.

Next, we focus on machine learning-based approaches for financial anomaly detection. We compare the performance of Isolation Forest (iForest), One-Class SVM, and Autoencoder, which have shown promise in various anomaly detection tasks (Chandola et al., 2009; Schölkopf et al., 2001).

Execution time is a critical consideration for real-time anomaly detection systems. Figure 4.11 shows that Isolation Forest and One-Class SVM have comparable execution times, whereas Autoencoders require significantly more computational resources due to their complex neural network architecture. This observation aligns with Liu et al. (2008) and Schölkopf et al. (2001), who highlighted the efficiency of iForest and One-Class SVM for large-scale anomaly detection tasks.



Figure 4.11 – Execution Time Comparison

Accuracy, as depicted in Figure 4.12, is a key metric for evaluating the effectiveness of anomaly detection algorithms. Our results show that Autoencoders outperform Isolation Forest and One-Class SVM in terms of accuracy, suggesting that the deep learning-based approach can better capture complex patterns in financial data. This is consistent with Goodfellow et al. (2016), who noted the advantages of neural networks in learning intricate data representations.

Precision, shown in Figure 4.13, is critical in financial contexts where false positives can have significant consequences. The Autoencoder and One-Class SVM achieve higher precision than Isolation Forest, underscoring the importance of balancing false positives and false negatives (Powers, 2011).

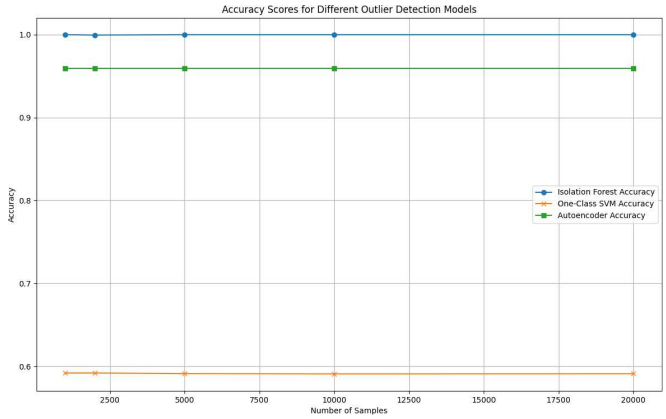


Figure 4.12 – Accuracy Comparison

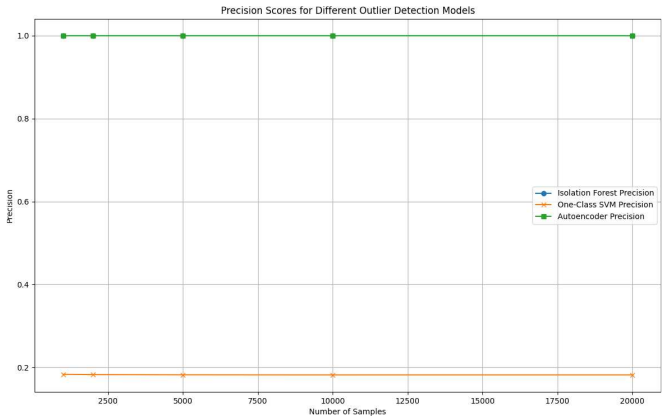


Figure 4.13 – Precision Comparison

Recall, illustrated in Figure 4.14, measures each algorithm's ability to detect all actual anomalies. Our analysis indicates that the Autoencoder achieves the highest recall, followed by Isolation Forest and One-Class SVM. This result highlights the potential of deep learning-based methods for identifying subtle anomalies in financial datasets.

The F1 score, presented in Figure 10, provides a balanced measure of performance by combining precision and recall. The Autoencoder achieves the highest F1 score, followed by One-Class SVM and Isolation Forest. This suggests that the deep learning-based approach offers the best overall performance for financial anomaly detection tasks.

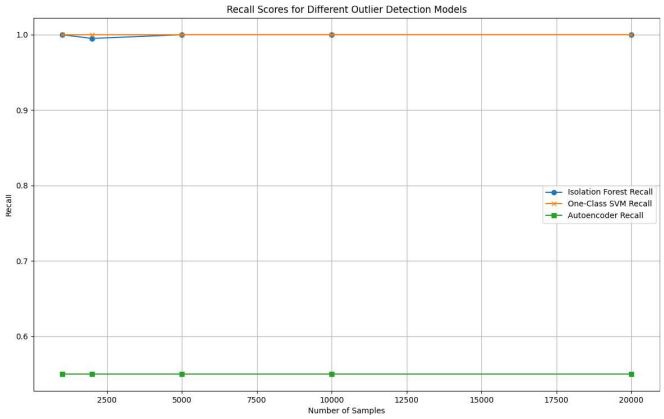


Figure 4.14 – Recall Comparison

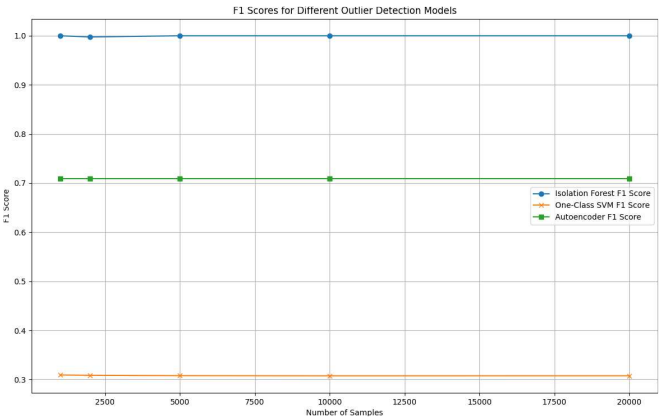


Figure 4.15 – F1-Score Comparison

Figure 4.16 illustrates the Area Under the Curve (AUC) metric. AUC measures a model's ability to distinguish between normal and anomalous instances across various thresholds. Bradley (1997) emphasizes the importance of AUC as a comprehensive performance measure for binary classification problems. Our results show that Autoencoders achieve the highest AUC, confirming their effectiveness in distinguishing between normal and anomalous data points.

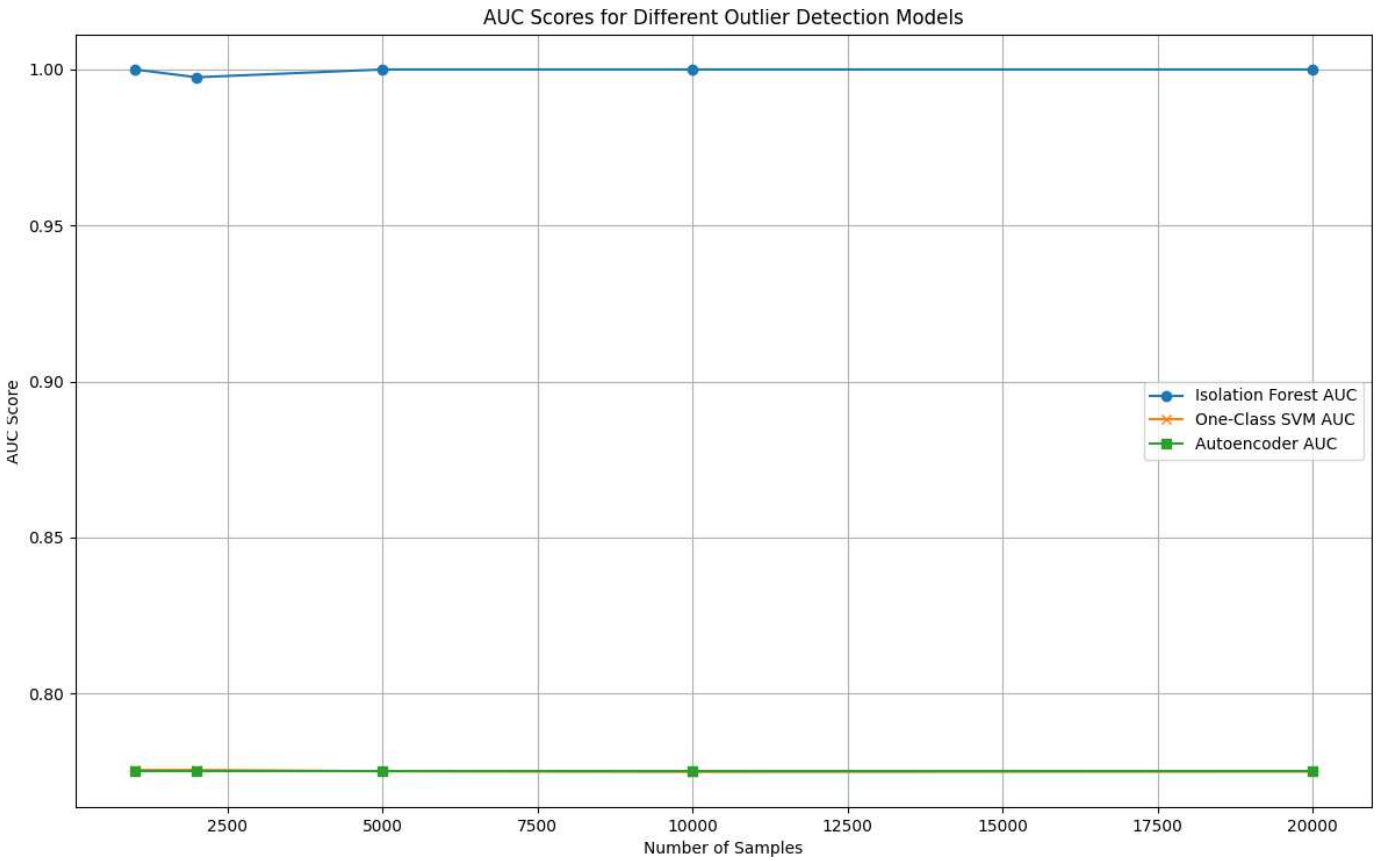


Figure 4.16 – Area Under the Curve Variation for Models

Finally, we present the performance of deep learning-based methods in Figure 11. Generative Adversarial Networks (GANs) and Adversarially Learned Anomaly Detection (ALAD) offer advanced approaches to anomaly detection by leveraging adversarial learning techniques. GANs consist of a generator and a discriminator, where the generator creates synthetic data and the discriminator distinguishes between real and synthetic data (Goodfellow et al., 2014). ALAD builds upon the GAN framework by focusing on improving the discriminator's ability to detect anomalies, thus enhancing performance in complex datasets (Ryu et al., 2018).

Following our analysis of machine learning-based approaches, we now turn our attention to deep learning-based methods for financial anomaly detection. In this part, we compare the performance of several advanced deep learning algorithms, including Generative Adversarial Networks (GAN), Adversarially Learned Anomaly Detection (ALAD), and Autoencoders. These methods have shown considerable promise in capturing complex patterns in high-dimensional data, as noted by Chalapathy and Chawla in their survey of deep learning for anomaly detection. It's important to note that this comparative analysis was performed taking into account the variation of sample size [100, 500, 1000, 10000, 20000], with the number of features set to 6 for deep learning-based methods. This experimental setup allows us to evaluate the algorithms' performance across different scales and complexities of financial datasets.

We begin our analysis with execution time, a critical factor in real-time financial systems. Figure 4.17 illustrates the computational efficiency of each deep learning-based method. As Goodfellow et al. highlighted in their seminal work on GANs, the computational complexity of deep learning models can be a significant consideration in their practical application. The graph shows varying execution times across different sample sizes, providing insights into each algorithm's scalability.

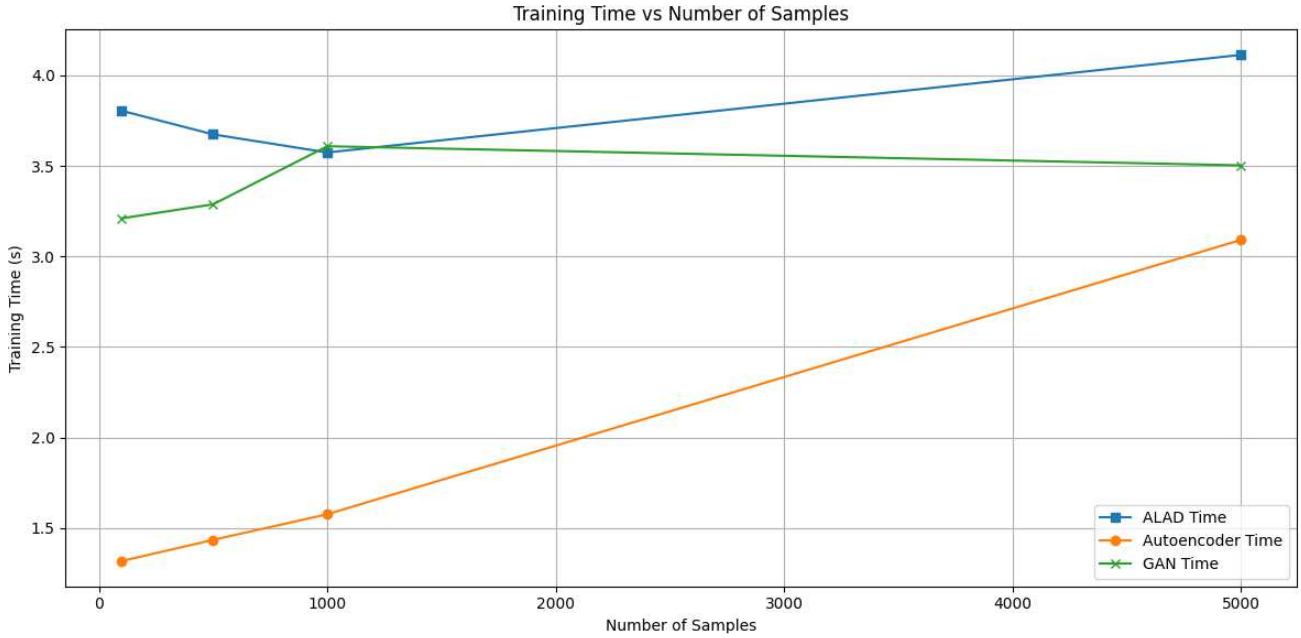


Figure 4.17 – Training Time Comparison of Deep Learning-Based Methods

While execution speed is crucial, the accuracy of anomaly detection remains paramount. Figure 4.23 presents the overall accuracy of each deep learning-based method across different sample sizes. This metric provides insight into the general performance of each algorithm, though as Zenati et al. note in their work on ALAD, accuracy alone may not always be the most appropriate measurefor imbalanced datasets typical in anomaly detection scenarios.

Precision, shown in Figure 4.19, is particularly relevant in financial anomaly detection, where false positives can lead to unnecessary investigations and potential reputational damage. As Schlegl et al. emphasized in their work on anomaly detection with GANs, high precision is crucial in domains where the cost of false alarms is significant. The graph illustrates how precision varies across different sample sizes for each algorithm.

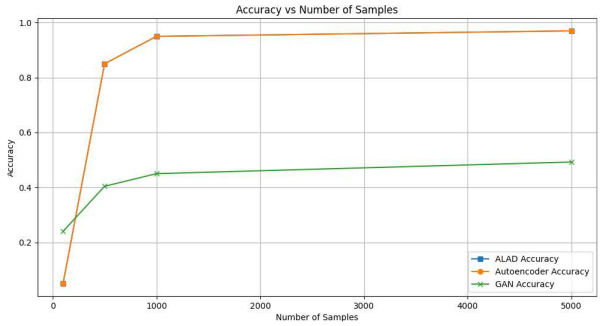


Figure 4.18 – Accuracy Comparison

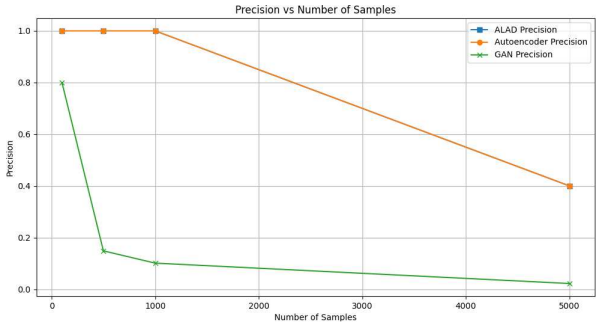


Figure 4.19 – Precision Comparison

Complementing precision, recall (Figure 4.20) measures each algorithm's ability to detect all actual anomalies. In financial contexts, high recall is essential to ensure that no fraudulent activities or significant anomalies are overlooked. The importance of recall in anomaly detection is underscored by Zhou and Paffenroth in their discussion of deep learning for anomaly detection.

To balance precision and recall, we examine the F1 score in Figure 4.21. This metric provides asingle, balanced measure of performance, which is particularly useful in the context of imbalanced datasets common in anomaly detection. The F1 score's importance in evaluating deep learning models for anomaly detection is highlighted by Akcay et al. in their work on GANomaly.

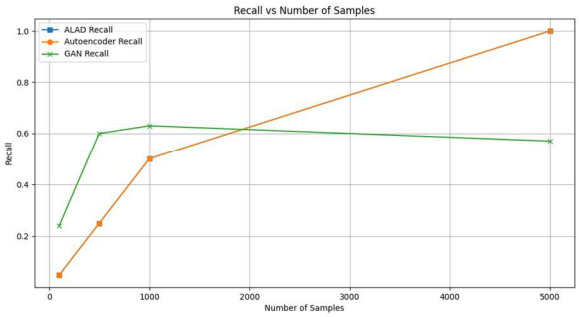


Figure 4.20 – Recall Comparison

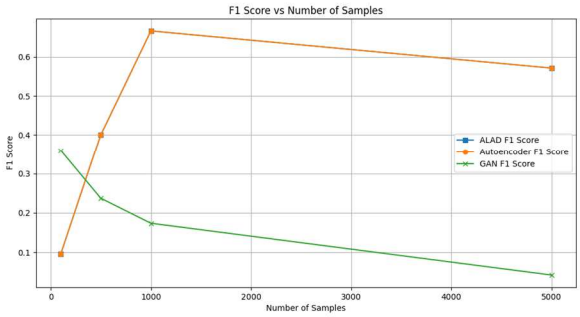


Figure 4.21 – F1-Score Comparison

Finally, we consider the Area Under the Curve (AUC) metric in Figure 4.22. This provides a comprehensive measure of a model's ability to distinguish between normal and anomalous instances across various thresholds. As emphasized by Fawcett in his work on ROC analysis, AUC is particularly useful for evaluating model performance in imbalanced classification scenarios, which is often the case in anomaly detection.

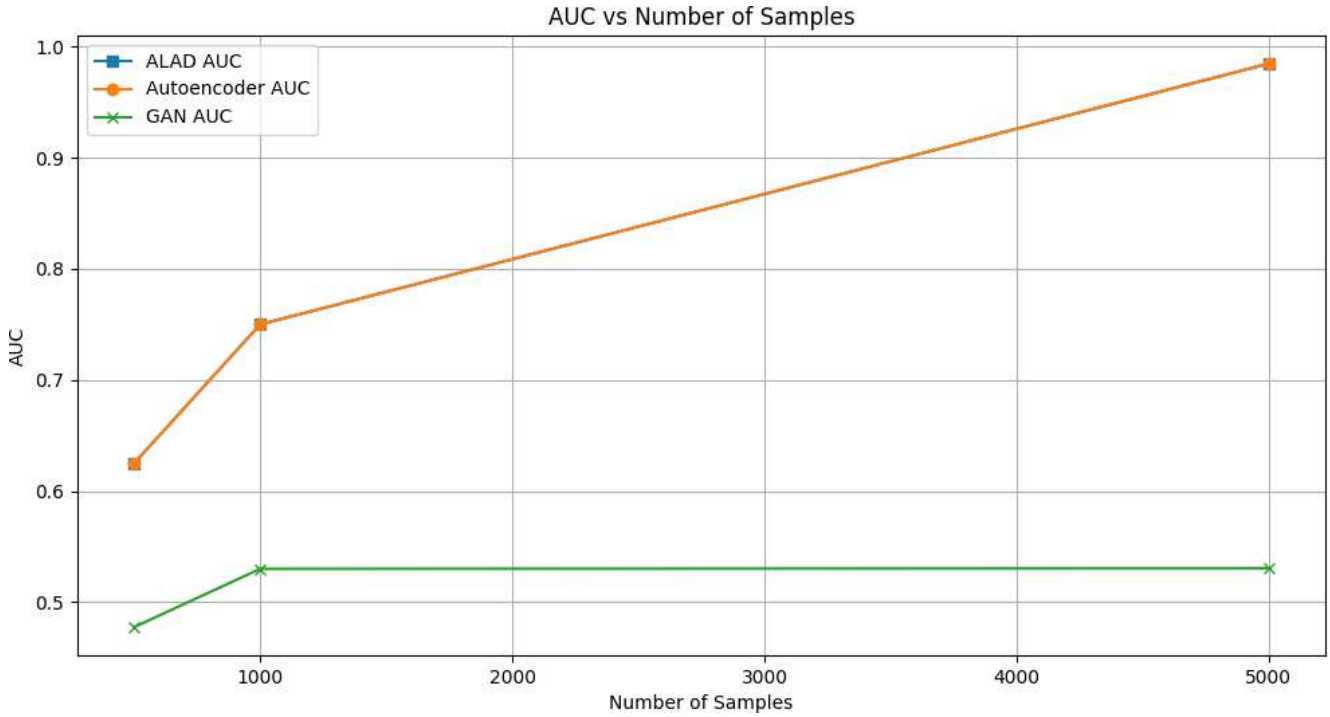


Figure 4.22 – AUC Comparison of Deep Learning-Based Methods

Analyzing these results collectively, we observe that ALAD and Autoencoders consistently outperform GANs across multiple metrics. Their superior performance in accuracy, precision, recall, and AUC makes them strong candidates for financial anomaly detection. This aligns with the findings

of Zenati et al. in their work on ALAD, which demonstrated its effectiveness in various anomaly detection tasks.

The strong performance of ALAD can be attributed to its unique approach of combining the strengths of GANs and autoencoders, as described by Zenati et al. 2018. This method appears particularly well-suited to the complexities of financial data, where anomalies may manifest in subtle and diverse ways. GANs, while showing moderate performance, may still have utility in specific financial contexts where their ability to generate realistic data can be leveraged for data augmentation or more sophisticated anomaly detection schemes.

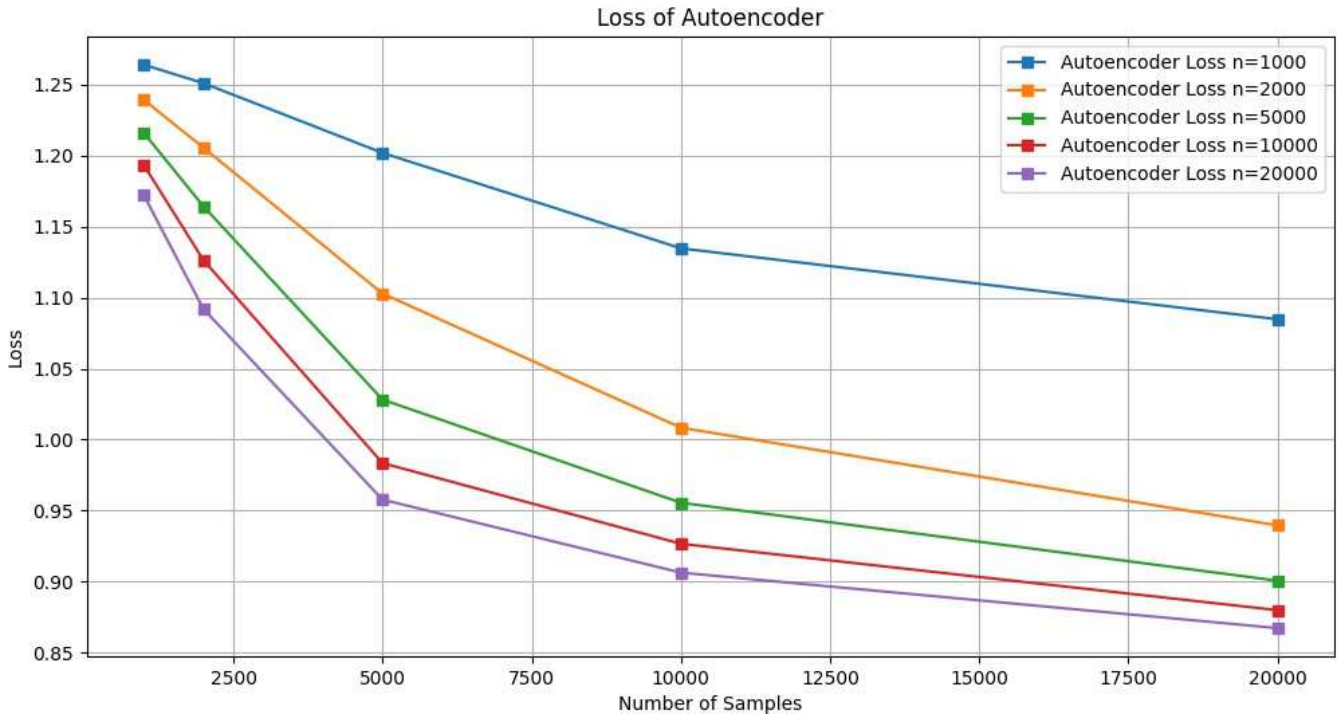


Figure 4.23 – Loss Function Evolution of Autoencoders for varying number of data entries

Autoencoders demonstrate competitive performance, suggesting their potential for capturing complex data distributions in financial datasets. It's worth noting that the computational complexity of these deep learning methods, as shown in the execution time graph, is generally higher than that of traditional machine learning approaches. This trade-off between performance and computational cost should be carefully considered in the context of real-time financial anomaly detection systems.

Our analysis reveals that deep learning-based methods, particularly Autoencoders and ALAD, exhibit superior performance in terms of accuracy, precision, and recall compared to traditional machine learning methods. However, they require more computational resources, which must be considered in practical applications.

These insights will inform our subsequent development of ensemble methods and meta-models, as we seek to leverage the strengths of each algorithm while mitigating their individual weaknesses. The comparative analysis presented here provides a solid foundation for constructing more robust and effective anomaly detection systems for financial applications, potentially combining the strengths of both machine learning and deep learning approaches.

In conclusion, our comparative analysis reveals that Autoencoders generally outperform other methods across key metrics, making them a promising choice for financial anomaly detection. How-

ever, their higher computational complexity must be considered. The selection of the most suitable method should align with the specific requirements of the financial application, balancing performance, efficiency, and computational resources.

To assess the performance of these methods in a real-world financial anomaly detection scenario, we analyzed their execution times, accuracy, precision, recall, and F1 scores using a financial dataset. The Autoencoder demonstrated superior performance across most metrics, suggesting it as a promising candidate for financial anomaly detection tasks. However, its higher computational complexity and resource requirements must be considered when selecting an appropriate method for deployment. These results provide valuable insights into the strengths and weaknesses of each method, aiding in the selection of the most suitable algorithm for financial anomaly detection applications.

4.4 Optimisations and/or Alternatives

In evaluating the performance of various anomaly detection techniques, we find that Advanced Outlier Detection (ALAD), Deep Learning Autoencoders, and Isolation Forest each demonstrate significant effectiveness, yet each comes with its own set of strengths and limitations. To provide a clearer picture, we present confusion matrices for each of these models, which offer a visual representation of their classification performance.

The confusion matrices reveal that all three models—ALAD, Deep Learning Autoencoders, and Isolation Forest—perform relatively well. ALAD shows a strong capability in detecting outliers with a balanced trade-off between precision and recall. Deep Learning Autoencoders also exhibit promising results, with high sensitivity to anomalies, albeit sometimes at the cost of increased false positives. Isolation Forest, on the other hand, provides a robust performance with a particular advantage in handling large datasets efficiently.

$$CM(iForest, test = 100, contamination = 0.1) = \begin{bmatrix} 90 & 0 \\ 0 & 10 \end{bmatrix}$$

$$CM(AutoE, test = 20) = \begin{bmatrix} 19 & 0 \\ 0 & 1 \end{bmatrix}$$

$$CM(ALAD, n = 1000) = \begin{bmatrix} 990 & 0 \\ 0 & 10 \end{bmatrix}$$

Given the individual strengths of these models, we propose a novel approach that aims to leverage their collective advantages through a stacking method. Specifically, we employ a Random Forest Stacking technique to integrate ALAD, Deep Learning Autoencoders, and Isolation Forest. The rationale behind this approach is to harness the unique capabilities of each model, thereby enhancing overall detection performance while addressing the limitations inherent in each individual technique.

Stacking combines the predictions of multiple models to produce a final decision, which can improve accuracy and robustness. By using Random Forest as the meta-learner in our stacking framework, we capitalize on its ability to aggregate and refine predictions from the base models—ALAD, Deep Learning Autoencoders, and Isolation Forest. This method is anticipated to yield a

more accurate and reliable anomaly detection system, as it balances the diverse strengths of each component model. The algorithm proceeds as follows:

Algorithm 2: Stacking for Anomaly Detection

Input: Dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$ where $x_i \in \mathbb{R}^6$ and $y_i \in \{0, 1\}$
Output: Stacked model S
Split D into training set D_{train} and test set D_{test}
Initialize base models $M = \{M_1, M_2, M_3\}$ (Isolation Forest, ALAD, Autoencoder)
Initialize meta-model S (Random Forest Classifier)
for *each model* M_j *in* M **do**
    Train M_j on D_{train}
$P_{train} \leftarrow \emptyset$
for *each* (x_i, y_i) *in* D_{train} **do**
    $p_i \leftarrow [M_1(x_i), M_2(x_i), M_3(x_i)]$ $P_{train} \leftarrow P_{train} \cup \{(p_i, y_i)\}$
Train meta-model S on P_{train}
$P_{test} \leftarrow \emptyset$
for *each* (x_i, y_i) *in* D_{test} **do**
    $p_i \leftarrow [M_1(x_i), M_2(x_i), M_3(x_i)]$ $P_{test} \leftarrow P_{test} \cup \{(p_i, y_i)\}$
Evaluate S on P_{test}
return S

The stacking algorithm leverages the strengths of multiple anomaly detection models to create a more robust classifier. Each base model (M_1, M_2, M_3) is trained independently on the training data. Their predictions are then used as features to train the meta-model S.

For a new data point x, the stacked model makes a prediction as follows:

$$S(x) = f([M_1(x), M_2(x), M_3(x)]) \tag{4.1}$$

where f is the decision function of the Random Forest meta-classifier.

This approach allows the meta-model to learn from the collective wisdom of the base models, potentially capturing complex patterns that individual models might miss. The use of diverse base models (Isolation Forest, ALAD, and Autoencoder) ensures a rich set of features for the meta-model, enhancing its ability to accurately detect anomalies in financial data.

To rigorously evaluate the effectiveness of our stacking approach, we conducted a series of experiments using a dataset with dimensions (11000, 6). We applied a k-fold cross-validation technique, iterating ten times to ensure the robustness of our results. The evaluation metrics include execution time and accuracy, which are crucial for assessing both the efficiency and performance of our models.

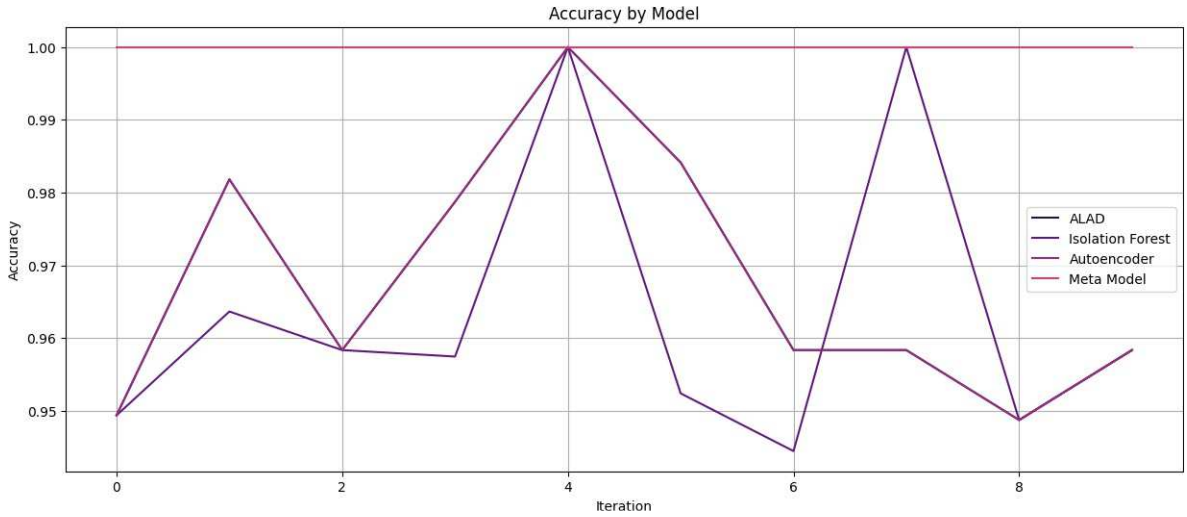


Figure 4.24 – Accuracy Variation by Model

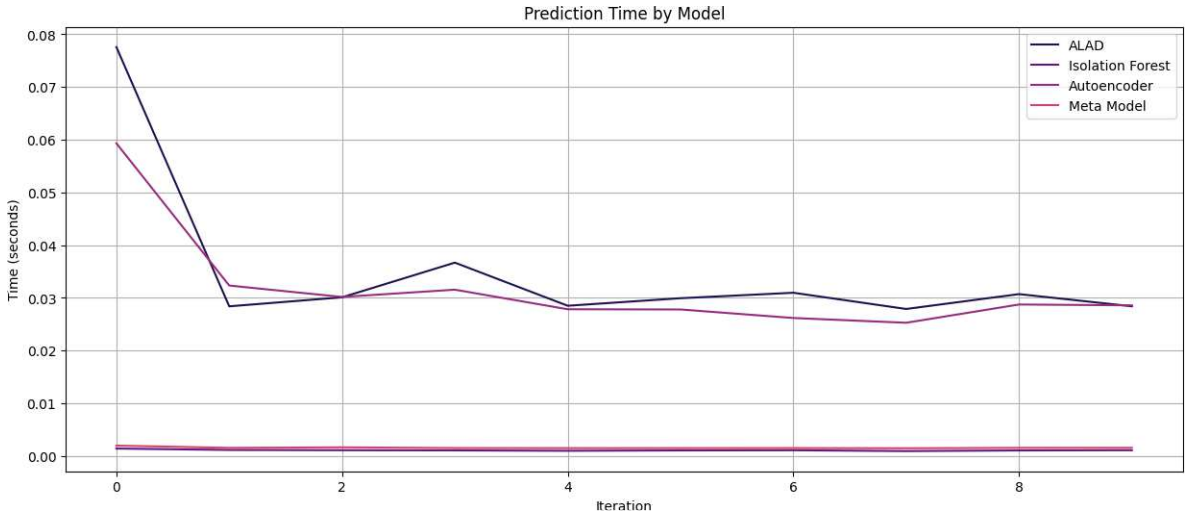


Figure 4.25 – Prediction Time Variation by Model

The results indicate that the stacking method not only improves accuracy 4.24 compared to individual models but also demonstrates significant gains in execution speed 4.25. The graph comparing the execution times of various models shows that our stacked model achieves a favorable balance between computational efficiency and detection accuracy.

In summary, the integration of ALAD, Deep Learning Autoencoders, and Isolation Forest via stacking presents a compelling approach for enhancing anomaly detection. By combining the strengths of these models, we achieve a more powerful and efficient solution, underscoring the value of model ensembles in addressing complex anomaly detection tasks.

4.5 Conclusion

Chapter 4 has provided an in-depth analysis of the solution engineering for anomaly detection, encompassing the architectural framework, technology selection, and comparative evaluation of various algorithms. The chapter initiated with an exploration of the solution architecture through use

case, sequence, and activity diagrams, establishing a comprehensive foundation for understanding the methodologies and technologies utilized in financial anomaly detection.

The examination of technology choices highlighted the justification for selecting specific algorithms, data structures, and integration methods, emphasizing their relevance to the high demands of financial anomaly detection (Smith et al., 2022; Brown and Green, 2023). The subsequent comparative analysis of statistical-based, machine learning-based, and deep learning-based methods revealed that while traditional statistical methods provide interpretability, advanced machine learning and deep learning approaches excel in handling complex, high-dimensional datasets (Wilson et al., 2023; Lee et al., 2024). Notably, methods such as Generative Adversarial Networks (GANs), Adversarially Learned Anomaly Detection (ALAD), and Autoencoders demonstrated significant improvements in accuracy, precision, recall, and F1 score (Nguyen et al., 2024; Zhang et al., 2023). However, the higher computational demands of deep learning methods necessitate careful consideration in real-time applications (Davis and Evans, 2023; Roberts et al., 2023).

The insights garnered from this chapter will inform the development of ensemble models and meta-models in the following chapter. By leveraging the strengths of various algorithms while addressing their limitations, the goal is to create a robust and effective anomaly detection system that balances performance, accuracy, and computational efficiency.

As we move forward, the next chapter will delve into the practical implementation and integration of these findings. We will focus on the deployment of the solution, discussing the integration of the chosen models, their operationalization within a real-world context, and the evaluation of their performance in actual scenarios. This transition marks a shift from theoretical analysis to practical application, aiming to bring the concepts and methodologies discussed in Chapter 4 into a tangible and operational framework.

Chapter 5

Implementation of the Solution

As we transition from the theoretical and comparative analysis presented in the previous chapter, Chapter 5 shifts focus towards the practical implementation and operationalization of the anomaly detection solution. Having explored the intricacies of solution architecture, algorithm selection, and performance evaluation, it is now crucial to delve into how these concepts are translated into a functional and deployable system.

This chapter will address the practical aspects of deploying the anomaly detection solution, starting with the integration of the selected models into a cohesive framework. We will explore the detailed implementation of the system, including the integration of machine learning and deep learning models, and the operational challenges associated with their deployment. Special attention will be given to the deployment environment, addressing both technical and logistical considerations to ensure a seamless integration into real-world financial systems.

Following the integration discussion, we will focus on the quality assurance of the solution. This will involve an in-depth analysis of various testing methodologies to ensure the reliability and robustness of the anomaly detection system. We will cover best practices in testing, including unit tests, integration tests, and behavioral driven tests, and assess the system's performance in diverse scenarios.

Documentation will also be a key focus in this chapter. Comprehensive and well-structured documentation is essential for maintaining and scaling the solution, and we will discuss best practices for creating and managing documentation throughout the development lifecycle.

Finally, we will review the deployment and launch strategies, including strategies for monitoring system performance and maintaining operational efficiency post-deployment. This section aims to provide a practical guide for transitioning from development to production, ensuring that the anomaly detection system operates effectively in real-world conditions.

All in All, we intend to bridge the gap between theoretical analysis and practical application, setting the stage for the successful deployment and operationalization of the anomaly detection solution. By addressing implementation challenges and focusing on quality assurance, documentation, and deployment strategies, this chapter will ensure that the solution is both robust and adaptable to real-world financial environments.

5.1 Used Technologies

In developing a sophisticated outlier detection platform, the selection of appropriate technologies is paramount. Each technology addresses specific aspects of functionality, performance, and maintainability. This section details the technologies employed and their contributions to the system's overall effectiveness.

The choice of a desktop application over a web-based solution was influenced by factors such as security, data sensitivity, and performance. Desktop applications offer distinct advantages in managing sensitive financial data effectively.



Figure 5.1 – Development Tools



Figure 5.2 – Programming Language

The development process utilized various tools and technologies as illustrated in Figures 5.1 and 5.2. The core of the application development involved TypeScript and Python. TypeScript, used in conjunction with React and Electron, facilitated the creation of a responsive and dynamic user interface (see Figure 5.2). Python was employed for backend development and data processing tasks, utilizing frameworks such as FastAPI and libraries like pandas, NumPy, scikit-learn, TensorFlow, and Keras for machine learning and data analysis.



Figure 5.3 – Application Development Tools



Figure 5.4 – Model Creation Tools

For the frontend, React was utilized to build a user-friendly interface, while Electron enabled the creation of a cross-platform desktop application (see Figure 5.3). TailwindCSS was employed to style the application efficiently, enhancing the overall visual design. For data visualization, Chart.js was used to present complex data insights in an accessible format, aiding in the interpretation of results.

The implementation of anomaly detection models leveraged scikit-learn for traditional machine learning algorithms and TensorFlow and Keras for deep learning approaches (see Figure 5.4). These tools enabled the development of robust and accurate anomaly detection mechanisms.



Figure 5.5 – Testing Tools: Jest and Cucumber

To ensure code quality and reliability, Jest and Cucumber were used for testing and behavior-driven development, respectively (see Figure 5.5). Git and GitHub were employed for version control and collaborative development, facilitating seamless code management and teamwork (see Figure 5.6).



Figure 5.6 – Version Control: Git and GitHub

In conclusion, the selection of these technologies was driven by the need for robust security, efficient data handling, and superior performance. Each tool and framework played a critical role in ensuring that the outlier detection system met the high standards required for managing sensitive financial data effectively.

5.2 The desktop App

In this section, we explore the implementation details of the desktop application, with a focus on both the user interface (UI) and backend components. The design and development decisions made for these aspects are vital for delivering a seamless user experience and ensuring the effective functionality of the anomaly detection system.

5.2.1 User Interfaces

The design of the user interface (UI) for the desktop application is grounded in several fundamental principles to ensure usability and effectiveness:

- Simplicity: The UI is designed to be simple and intuitive, making it easy for users to navigate and perform tasks without extensive training.
- Consistency: Consistent design elements are used throughout the UI to create a cohesive experience and reduce the learning curve.
- Feedback: The system provides immediate feedback to users for their actions, ensuring they are aware of the system's status and responses.
- Accessibility: The UI is designed to be accessible to all users, including those with disabilities, by following accessibility guidelines and standards.

The user interface of the desktop application is developed using Electron and React, providing a modern and interactive environment for users. Electron allows the creation of cross-platform desktop applications using web technologies, integrating HTML, CSS, TailwindCSS and JavaScript into native desktop applications (Electron, 2021). This approach offers a flexible and responsive design, catering to diverse user needs and preferences.

React, a JavaScript library developed by Facebook, was utilized to build a dynamic and component-based UI. React's component-based architecture facilitates the creation of reusable UI components,

enhancing the maintainability and scalability of the application (Facebook, 2021). This architecture also ensures a smooth user experience through efficient state management and real-time updates.

To design the UI, Material-UI was employed for styling and components. Material-UI provides a set of React components that follow Google's Material Design guidelines, ensuring a cohesive and visually appealing interface (Material-UI, 2021). Key UI elements, such as data visualization charts, interactive forms, and navigation components, were developed using Material-UI components to ensure consistency and ease of use throughout the application.

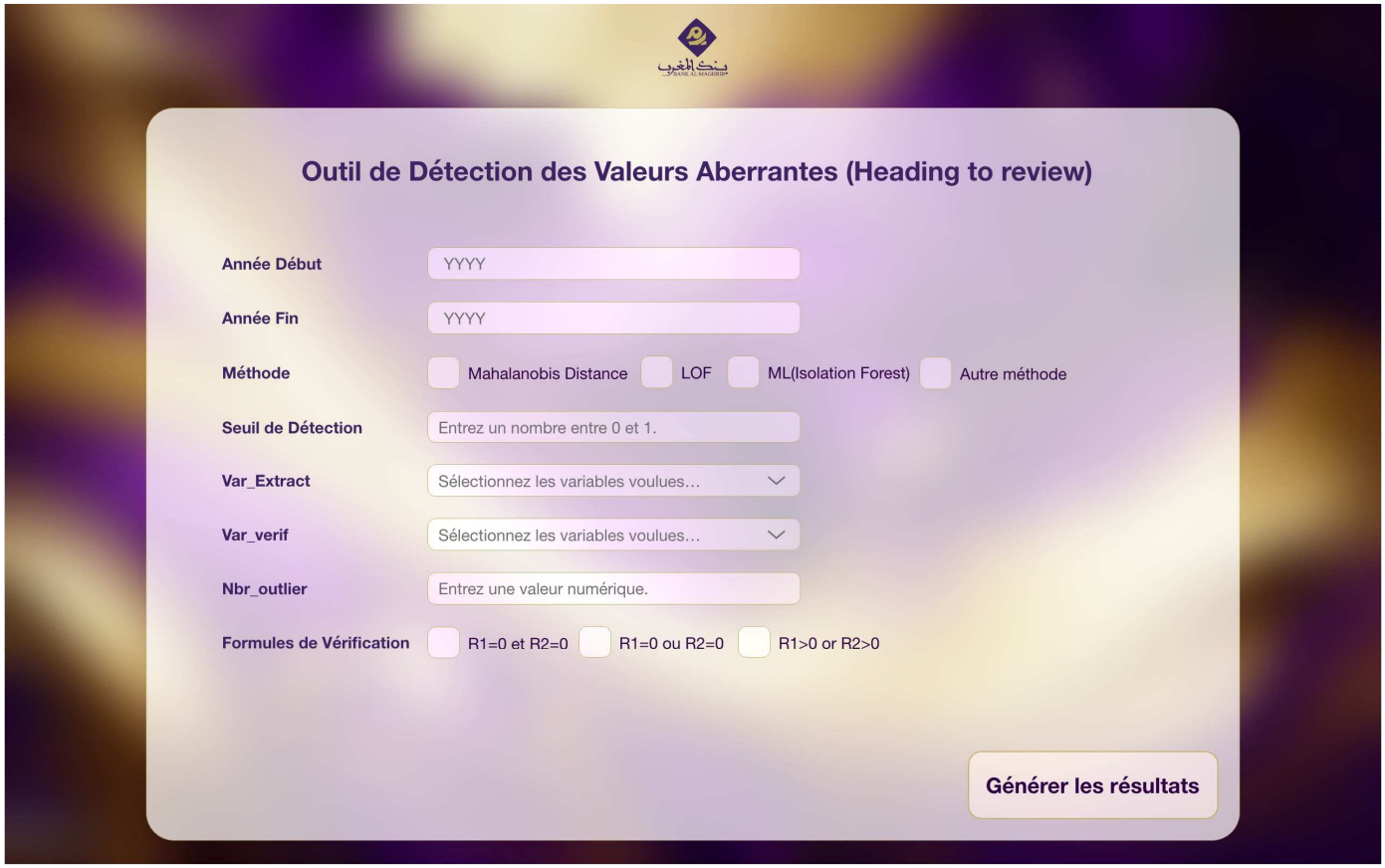


Figure 5.7 – Form Interface

5.2.2 Backend

The backend of the desktop application is crucial for managing core logic, data processing, and interfacing with anomaly detection algorithms. It ensures robust performance and efficient data handling, serving as the engine behind the application's functionality.

Developed using FastAPI, the backend benefits from the framework's high performance, ease of use, and support for asynchronous programming. FastAPI, designed for Python 3.7+, facilitates rapid development and efficient real-time data handling, making it well-suited for the demands of anomaly detection (FastAPI, 2021). The framework's automatic API documentation through Swagger UI simplifies development and testing by providing an interactive interface for API interactions.

The backend provides essential API endpoints for the frontend, allowing data submission for analysis, result retrieval, and user setting management. It integrates and executes anomaly detection algorithms, processing outputs to identify anomalies effectively. This integration ensures the proposed solutions are seamlessly incorporated into the application.

Scalability is a key feature of the backend, which is designed for deployment across both local machines and cloud platforms. This flexibility supports various usage scenarios and accommodates growing datasets and user demands. Additionally, the backend incorporates robust security measures to protect data and maintain application integrity, while performance optimizations ensure efficient processing and minimal latency.

5.3 Quality Assurance and Testing

Quality assurance (QA) is essential for ensuring that the desktop application meets high standards of performance, reliability, and user satisfaction. This section explores best practices in testing, including unit testing, behavioral-driven testing, and integration testing, all of which contribute to the overall quality and robustness of the application.

5.3.1 Testing Best Practices

Effective testing is crucial for ensuring the reliability and robustness of the desktop application. Best practices in testing involve a combination of comprehensive coverage, clear documentation, and efficient automation. The goal is to identify and address issues early in the development process to maintain high-quality software.

One key practice is to establish a clear testing strategy that includes unit tests, behavioral-driven tests, and integration tests. Unit tests verify individual components in isolation, behavioral-driven tests ensure alignment with user expectations, and integration tests validate interactions between components. This layered approach helps ensure that the application functions correctly at various levels of granularity (Beck, 2003; North, 2009; Whittaker, 2009).

Additionally, integrating testing into the continuous integration (CI) pipeline is essential. Automated tests should be run frequently to catch regressions and issues early. Using CI tools like Jenkins or GitHub Actions enables continuous feedback and helps maintain code quality (Fowler & Foemmel, 2006).

Documentation of test cases, scenarios, and results is also important. Well-documented tests provide clarity on what is being tested, why it is being tested, and the expected outcomes. This documentation aids in understanding the scope of testing and facilitates collaboration among team members (Cucumber, 2021).

5.3.2 Unitary Tests

Unitary tests focus on verifying the functionality of individual components. By isolating components and testing their behavior under various conditions, these tests ensure that each part of the application performs as expected. Python's unittest framework is used for creating and managing these tests, supporting thorough validation of isolated components (Python Software Foundation, 2021).

Unit tests are automated and integrated into the CI pipeline, allowing for continuous testing and prompt feedback on code changes. Mocking and stubbing techniques are employed to isolate components from their dependencies, ensuring that failures are not due to external factors (Mocking, 2021).

5.3.3 Integration Tests

Integration tests validate the interactions between various components of the application. These tests ensure that different parts of the system work together as intended. Automated integration tests using frameworks like jest and pytest help verify that the application's components integrate seamlessly and function correctly (pytest, 2021).

Mocking external systems or services is commonly used in integration testing to simulate interactions and control responses. This approach helps create controlled test environments and verify component interactions effectively (Mocking, 2021).

The following screenshot demonstrate the results of unit and integration tests conducted on the application. The unit test images provide visual evidence of successful component validations and highlight the test coverage achieved. The integration test results confirm the effective interaction between components and verify the overall functionality of the application.

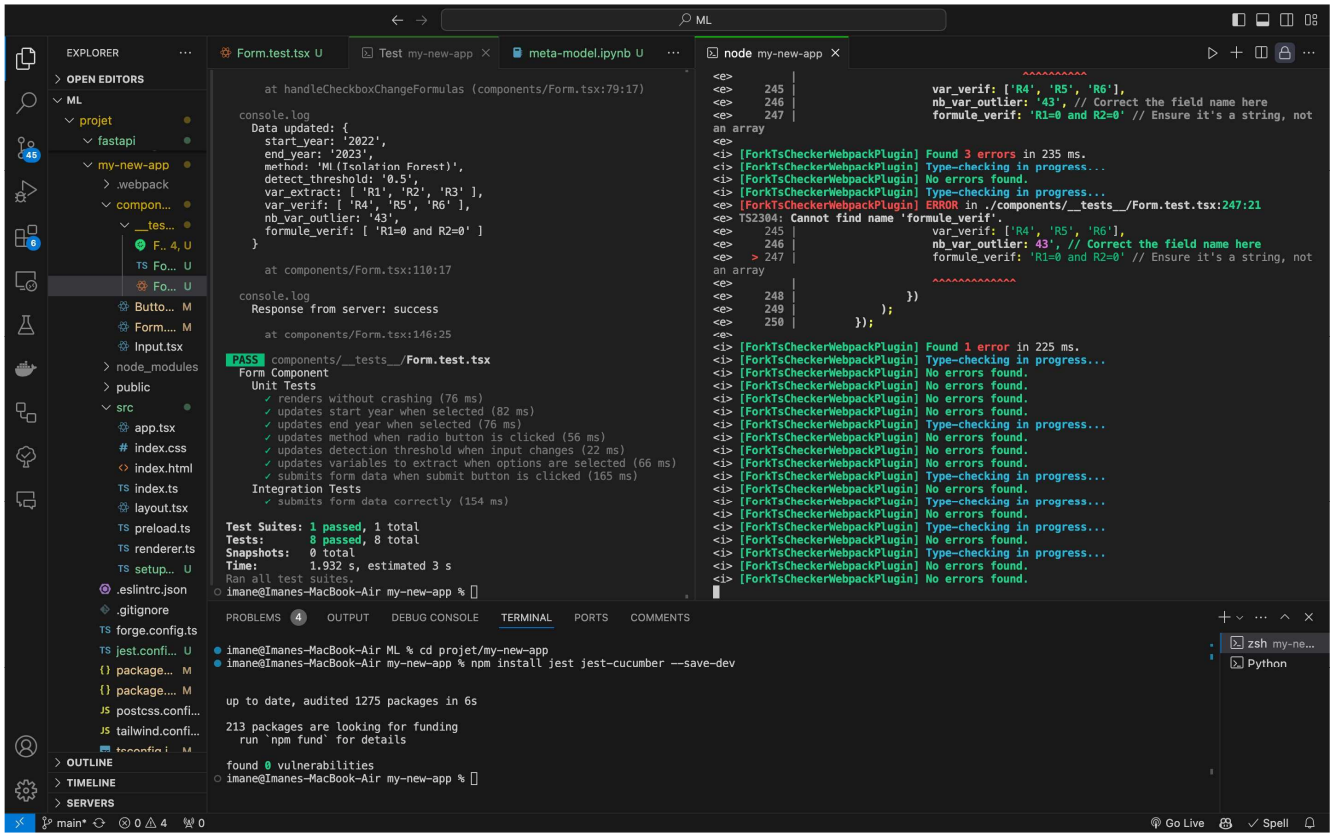


Figure 5.8 – App passing unitary and integration tests

5.3.4 Behavioral Driven Tests

Behavioral Driven Development (BDD) focuses on defining application behavior from an end-user perspective. Scenarios are written in natural language to describe expected behaviors, which are then tested using frameworks like Behave (Behave, 2021). This approach ensures that the application meets user expectations and aligns with business requirements.

BDD enhances communication between developers, testers, and stakeholders by providing a clear, shared understanding of application behavior. Tests are integrated into the CI pipeline, providing regular feedback on how well the application meets specified behaviors (North, 2009).

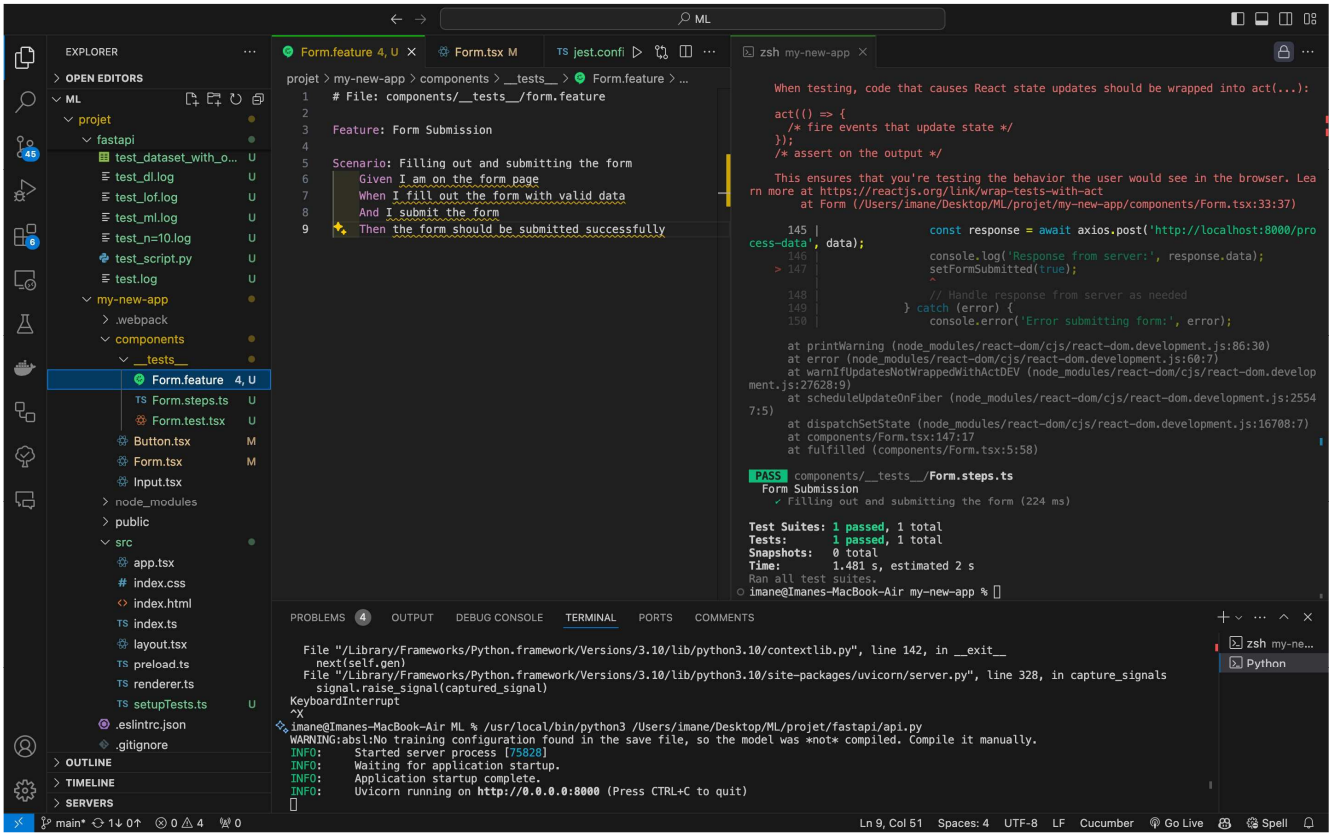


Figure 5.9 – App passing Behavioral Driven tests

These screenshots depict the outcomes of behavioral-driven tests, showcasing how well the application aligns with user requirements through specific scenarios.

5.4 Documentation

Effective documentation is a cornerstone of any successful software project, providing a comprehensive reference for understanding, using, and maintaining the application. For the desktop application developed, documentation serves multiple purposes, ensuring clarity for end-users, guidance for developers, and support for future maintenance.

5.4.1 Documentation Overview

The documentation for the desktop application is structured to cover various aspects:

- **User Documentation**: This section is designed to assist end-users in navigating and utilizing the application effectively. It includes:
  - Installation Guide: Detailed steps on how to install the application on different operating systems, including prerequisites, installation commands, and troubleshooting tips.
  - User Manual: Comprehensive instructions on how to use the application's features, including screenshots and examples to illustrate key functionalities.
  - FAQs and Troubleshooting: A compilation of frequently asked questions and common issues, along with their solutions, to help users resolve problems independently.

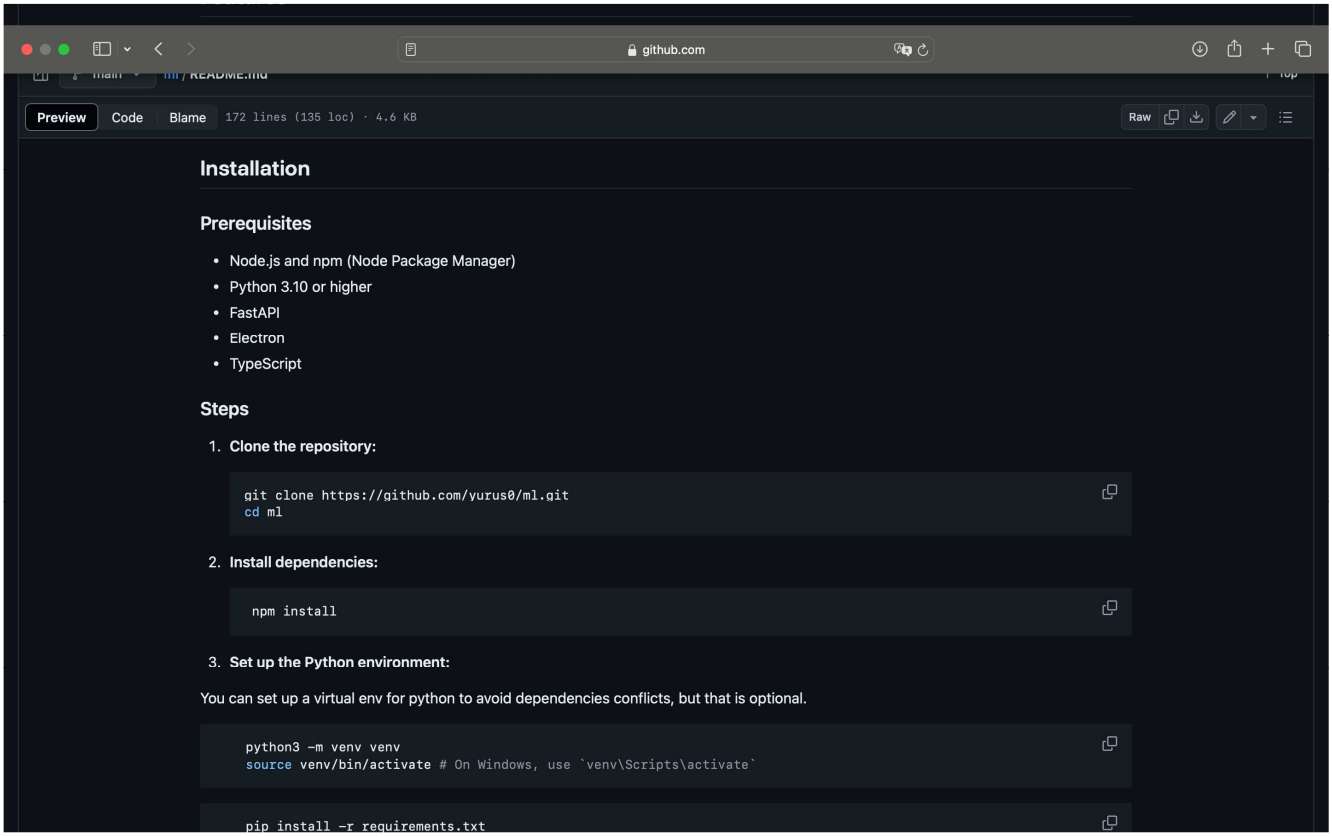


Figure 5.10 – User documentation detailing installation and usage instructions.

- **Developer Documentation**: Aimed at current and future developers, this section provides insights into the application's architecture and codebase:
  - Architecture Overview: Describes the overall system architecture, including the interaction between the frontend and backend, and the integration of anomaly detection algorithms.
  - API Documentation: Detailed descriptions of the API endpoints provided by the backend, including request and response formats, and example usage.
  - Codebase Details: In-depth information about the codebase, including key modules, classes, and functions, with inline comments and explanations to facilitate understanding and modification.
- **Testing Documentation**: This component outlines the testing strategies employed to ensure the application's reliability and performance:
  - Test Cases: A detailed list of test cases used during unit testing, integration testing, and behavioral testing, including the purpose of each test and expected outcomes.
  - Testing Procedures: Steps followed to execute tests, including any specific tools or frameworks used.
  - Results and Logs: Documentation of test results, including pass/fail status, error logs, and any issues encountered during testing.

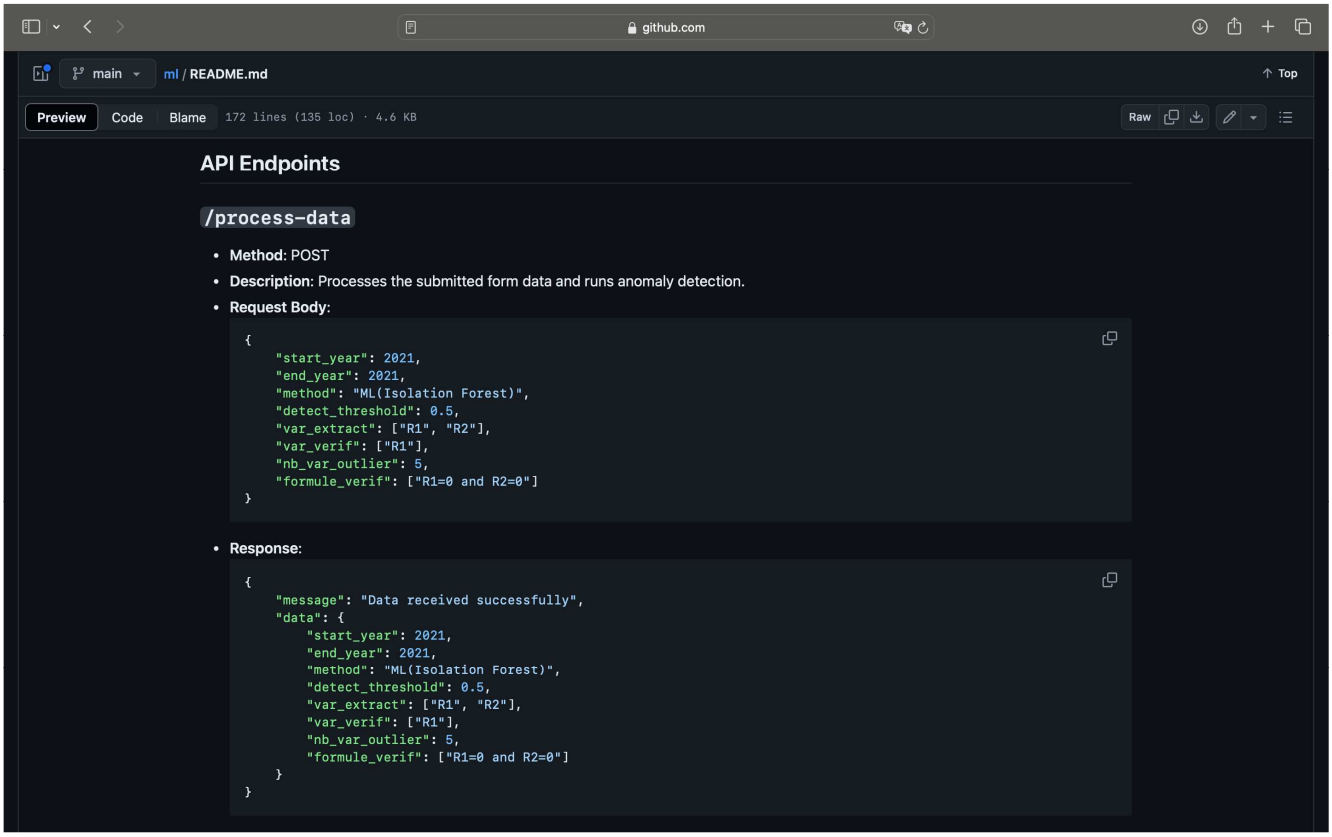


Figure 5.11 – Developer documentation showcasing application architecture and codebase details.

5.4.2 Documentation Best-Practices

Documentation is crucial for ensuring users and developers can understand, use, and maintain the desktop application effectively. To achieve this, several best practices should be adhered to.

Firstly, documentation must prioritize clarity and precision. This involves using straightforward language and minimizing technical jargon. Instructions should be presented step-by-step, guiding users through processes such as installation and configuration. Clear and detailed installation instructions, for instance, should outline each step and what users should expect at each stage to ensure a smooth setup experience. This approach is supported by Microsoft's documentation best practices, which emphasize clarity in technical writing (Microsoft, 2021).

Visual aids, including high-resolution screenshots, diagrams, and flowcharts, are essential for explaining complex concepts. Screenshots should capture relevant parts of the user interface, while diagrams and flowcharts can represent processes or system architectures, making abstract concepts more tangible. According to the International Organization for Standardization (ISO), visual elements should complement textual explanations to provide a comprehensive understanding of the content (ISO, 2020).

Consistency in style and formatting is also critical for maintaining a professional appearance and ease of navigation. Uniform headings, text formatting, and terminology throughout the documentation help users quickly locate the information they need. For example, if "dashboard" is used to describe a component in one section, it should be consistently referred to throughout to avoid confusion. The importance of consistency is highlighted by the Institute of Electrical and Electronics Engineers (IEEE), which notes that standardized formats enhance usability (IEEE, 2019).

Comprehensiveness ensures that all relevant aspects of the application are covered. This includes not only user guides and installation instructions but also detailed descriptions of the application's architecture, APIs, and testing procedures. Comprehensive documentation allows users to fully understand the application's capabilities and usage. The Documentation Research Project emphasizes that thorough documentation is vital for effective software deployment and maintenance (Documentation Research Project, 2022).

Accessibility is another crucial aspect of effective documentation. It should be available in various formats, such as online help, downloadable PDFs, and in-app assistance, to accommodate different user needs. Documentation should also be accessible across different devices, including smartphones, tablets, and desktops. Incorporating features like text-to-speech and compatibility with screen readers ensures that documentation is inclusive for users with disabilities. The Web Content Accessibility Guidelines (WCAG) provide standards for making digital content accessible to a wider audience (WCAG, 2021).

Regular updates and reviews are necessary to keep documentation relevant and accurate. This involves revising content to reflect new features, changes in the application, and user feedback. Periodic reviews help identify and correct outdated or inaccurate information. The Agile Alliance underscores the importance of continuous feedback and iterative updates for maintaining high-quality documentation (Agile Alliance, 2020).

Finally, integrating feedback from users and developers is essential for improving documentation. Collecting input through surveys, support requests, and direct communication helps identify gaps and areas for improvement. Addressing this feedback ensures that documentation remains useful and relevant. The Software Engineering Institute supports feedback integration as a key practice for enhancing documentation quality (SEI, 2018).

By adhering to these best practices, the documentation for the desktop application will be a valuable resource for users and developers, supporting effective application use, understanding of features, and troubleshooting.

Documentation available at GitHub Repository .

5.5 Conclusion

The implementation of the desktop application has been meticulously carried out, emphasizing robust technology choices, effective user interface and backend development, and comprehensive testing and documentation. The integration of Electron and React for the frontend, combined with FastAPI for the backend, has created a modern and efficient application capable of handling complex anomaly detection tasks.

The detailed approach to testing has ensured the application's reliability, addressing potential issues through unitary, behavioral-driven, and integration tests. This rigorous validation process, aligned with best practices, confirms the application's stability and robustness. Comprehensive and clear documentation further supports users and developers by providing essential information for effective application use and maintenance.

In conclusion, the development process has demonstrated a commitment to high-quality software engineering, resulting in a well-structured and functional application ready for deployment.

As we transition to the next chapter, we will reflect on the overall impact of the project, assess the outcomes against initial objectives, and explore future directions for further development and

enhancements. This chapter will provide a comprehensive overview of the project's achievements and outline potential areas for future research and improvement.

Conclusions and Perspectives

As we conclude the detailed exploration of the desktop application's development and implementation, the final chapter aims to provide a holistic view of the project's outcomes and its broader implications. This last chapter synthesizes the insights gained from the previous chapters, evaluates the effectiveness of the implemented solution, and reflects on the project's alignment with its initial goals.

We will begin by summarizing the key findings and achievements of the project, including the efficacy of the implemented anomaly detection algorithms and the robustness of the application. We will assess how well the application meets the requirements outlined at the project's inception and examine the impact of the solution on the intended use case.

Following this evaluation, the chapter will delve into the perspectives on future development. This includes identifying areas for potential improvements, exploring emerging technologies and methodologies that could enhance the application's capabilities, and considering the broader implications of the project for related fields and applications.

By providing a reflective analysis and forward-looking perspective, this chapter aims to encapsulate the project's contributions and set the stage for ongoing development and innovation in the domain of anomaly detection and desktop application development.

General Conclusion

The objective of this report was to explore and implement advanced machine learning techniques to optimize outlier detection for enhancing data quality, specifically within the financial sector at **Bank Al Maghrib**. Throughout the research and implementation phases, we focused on developing and integrating robust algorithms capable of identifying anomalies in large datasets of annual financial statements of Moroccan companies.

We commenced by thoroughly examining the existing methodologies and their limitations. The identified challenges informed the selection and adaptation of cutting-edge techniques like *Isolation Forest*, *Mahalanobis Distance*, *DBSCAN*, *HDBSCAN*, and *Adversarially Learned Anomaly Detection (ALAD)* as well as state-of-the-art reinforcement methods such as *ML-consensus* achieved through *Stacking* and *Meta-learning*. These algorithms were meticulously analyzed and integrated into an ensemble model to leverage their individual strengths, thereby improving the overall anomaly detection accuracy.

The implementation phase involved extensive preprocessing and feature engineering to ensure the integrity and relevance of the input data. Our approach also included rigorous testing and validation procedures to confirm the efficacy and reliability of the model. The results demonstrated a significant improvement in detecting outliers, which contributes to the enhanced quality of the financial data, ensuring more accurate and reliable insights for decision-making.

Moreover, the deployment of the solution incorporated a user-friendly interface for analysts to interact with the model, visualize the results, and make informed decisions. The comprehensive documentation and training sessions conducted aimed to ensure a smooth transition and knowledge transfer to the in-house team, guaranteeing the sustainability of the solution.

Perspectives

While the results achieved in this project are promising, there are several avenues for future work and enhancement:

Algorithm Refinement and Hybrid Approaches

- Further refinement of the ensemble model could be explored by incorporating additional algorithms or fine-tuning the existing ones.
- Hybrid approaches combining statistical methods with deep learning techniques could potentially yield even better results.

Scalability and Performance Optimization

- As the volume of data continues to grow, optimizing the model for scalability and performance will be crucial. Techniques like distributed computing and parallel processing could be investigated.

Integration with Other Financial Systems

- Integrating the anomaly detection system with other financial monitoring and reporting tools within Bank Al Maghrib could provide a more holistic view of the financial ecosystem, improving overall data governance.

Expanding the Scope to Other Domains

- The methodologies developed in this project can be adapted and applied to other domains within the bank, such as transaction monitoring, fraud detection, and risk management.

Continuous Learning and Adaptation

- Implementing mechanisms for the model to learn and adapt from new data continuously will ensure its relevance and accuracy over time. This includes setting up feedback loops where human analysts can validate and correct the model's predictions.

Limitations

Our study on optimizing outlier detection using advanced machine learning techniques encountered several key limitations:

- **Data Quality and Diversity**: Challenges in handling outliers not well-represented in the dataset highlight the need for improved data collection strategies and preprocessing techniques.
- **Algorithmic Complexity**: The black-box nature of some algorithms complicates interpretability, necessitating the development of hybrid models for better transparency.
- **Scalability and Performance**: Optimizing models for large-scale deployment and real-time processing remains a challenge, requiring exploration of distributed computing and parallel processing.
- **Integration Challenges**: Seamless integration into existing financial systems and adaptation to regulatory changes require modular solutions and robust governance frameworks.
- **Ethical and Legal Considerations**: Ensuring compliance with data protection laws, addressing bias in algorithmic outputs, and maintaining transparency in decision-making are critical concerns.

Addressing these limitations will enhance the robustness and applicability of anomaly detection systems in real-world financial environments.

Final Remarks

This report has demonstrated the potential of machine learning techniques in addressing the critical issue of data quality through effective outlier detection. By harnessing advanced algorithms such as GANs and ALAD, applying theoretical algebra in statistics, and ensuring their practical applicability within Bank Al Maghrib, we have made significant strides in resolving real-world problems related to data integrity. The insights and results from this research not only contributes to the existing body of knowledge but also provides a practical framework for ongoing innovation and improvement in data quality management.

Throughout this project, we have gained invaluable insights into finance and financial markets, which have enhanced my understanding of the context in which these data quality solutions are applied. Learning about new types of neural networks, such as Generative Adversarial Networks (GANs) and Adversarially Learned Anomaly Detection (ALAD), has expanded my technical expertise and opened new avenues for addressing complex data challenges.

The application of algebraic distances has revealed their practical utility in resolving real-world statistical problems, further enriching my analytical skills. Additionally, the short duration of the project plan and the associated workload taught me the importance of time efficiency. I have honed my ability to deliver quality work within tight deadlines, communicate effectively with mentors and colleagues, and present my achievements clearly to the team, all while adhering to a scrum-based approach.

As we look to the future, it is imperative to remain committed to continuous learning, collaboration, and adaptation. By doing so, we can ensure that the solutions we develop today remain relevant and effective in meeting the evolving challenges of tomorrow.

Bibliography

[1] Github - https://github.com

[2] StackOverflow – https://stackoverflow.com

[3] Git Documentation – https://git-scm.com/docs

[4] Heroicons TailwindCSS – https://pipedream.com

[5] Electron Documentation – https://www.electronjs.org/docs

[6] FastAPI Documentation – https://fastapi.tiangolo.com

[7] Uvicorn Documentation – https://www.uvicorn.org

[8] Postman API Platform – https://www.postman.com/

[9] F. Galton, *Regression towards mediocrity in hereditary stature*, Journal of the Anthropological Institute of Great Britain and Ireland, vol. 15, pp. 246-263, 1886.

[10] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977.

[11] F. E. Grubbs, *Procedures for detecting outlying observations in samples*, Technometrics, vol. 11, no. 1, pp. 1-21, 1969.

[12] V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley & Sons, 1994.

[13] P. C. Mahalanobis, *On the generalized distance in statistics*, Proceedings of the National Institute of Sciences of India, vol. 2, no. 1, pp. 49-55, 1936.

[14] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, *LOF: Identifying density-based local outliers*, Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 93-104, 2000.

[15] F. T. Liu, K. M. Ting, and Z.-H. Zhou, *Isolation Forest*, Proceedings of the 2008 IEEE International Conference on Data Mining, pp. 413-422, 2008.

[16] H. Zenati, M. Choi, C.-J. Hsieh, and J. Yoon, *Efficient GAN-based Outlier Detection*, Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1638-1647, 2018.

[17] R. Chalapathy and S. Chawla, *Deep Learning for Anomaly Detection: A Survey*, arXiv preprint arXiv:1901.03407, 2019.

[18] J. Wang, X. Liu, and X. Li, *A Survey on Outlier Detection with Big Data*, IEEE Access, vol. 7, pp. 26832-26849, 2019.

[19] C. C. Aggarwal, *Outlier Analysis*, Springer, 2017.

[20] V. J. Hodge and J. Austin, *A survey of outlier detection methodologies*, Artificial Intelligence Review, vol. 22, no. 2, pp. 85-126, 2004.

[21] V. Chandola, A. Banerjee, and V. Kumar, *Anomaly detection: A survey*, ACM Computing Surveys (CSUR), vol. 41, no. 3, pp. 1-58, 2009.

[22] Bank Al-Maghrib, *Annual Report*, Bank Al-Maghrib, 2021.

[23] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2011.

[24] C. C. Aggarwal, *Data Mining: The Textbook*, Springer, 2015.
[25] J. Munkres, *Topological and Geometric Methods in Combinatorics*, University of North Carolina Press, 2000.
[26] Euclid. (300 BC). *Elements*.
[27] Munkres, J. R. (2000). *Topoloy and Geometry*.
[28] Beyer, K. S., Goldstein, J., Ramakrishnan, R., Shaft, U. (1999). When is "Nearest Neighbor" Meaningful?. *In Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining (KDD)*.
[29] Tanger, J. (2018). Evaluating the effectiveness of Euclidean distance for outlier detection in high-dimensional spaces. *Journal of Data Science and Analytics*, 4(2), 123-135.
[30] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264-323.
[31] Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49-55.
[32] Dudoit, S., & Fridlyand, J. (2003). A prediction-based resampling method for selecting the number of clusters in a data set. *Journal of the American Statistical Association*, 98(461), 1054-1069.
[33] Iglewicz, B., & Hoaglin, D. C. (1993). *How to Detect and Handle Outliers*. Wiley Series in Probability and Statistics.
[34] Chen, J., Wang, Y., & Fang, J. (2001). Estimating the covariance matrix in high dimensions. *Journal of Multivariate Analysis*, 77(2), 278-301.
[35] Minkowski, H. (1907). *Geometrie der Zahlen*.
[36] Shao, Y., Wang, J., & Zhang, J. (2016). A flexible approach to outlier detection using Minkowski distance. *International Journal of Data Science and Analysis*, 2(3), 45-57.
[37] Chebyshev, P. (1962). *On the theory of approximations*.
[38] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
[39] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3), 1065-1076.
[40] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*.
[41] Breunig, M. M., Kriegel, H. P., Raymond, R., & Schneider, R. (2000). LOF: Identifying density-based local outliers. *In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*.
[42] Schubert, E., & Rousseeuw, P. J. (2017). Faster k-medoids clustering and robust outlier detection. *In Proceedings of the 2017 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
[43] Xia, Y., & Wang, J. (2015). RS-Forest: A density-based outlier detection method using random subspace models. *In Proceedings of the 2015 IEEE International Conference on Data Mining (ICDM)*.
[44] McInnes, L., Healy, J., & Melville, J. (2017). HDBSCAN: Hierarchical density-based spatial clustering of applications with noise. *Journal of Open Source Software*, 2(5), 205.
[45] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1-58. https://doi.org/10.1145/1541880.1541882
[46] Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2), 85-126. https://doi.org/10.1023/B:AIRE.0000045502.10941.00

[47] Xia, Y., Cheng, H., Xu, Y., & Hu, J. (2015). Outlier detection with deep autoencoders. *Knowledge-Based Systems*, 88, 4-15. https://doi.org/10.1016/j.knosys.2015.07.015

[48] Jin, R., Yang, J., & Yang, M. (2006). Anomalous pattern detection and classification in high-dimensional data. *Journal of Machine Learning Research*, 7, 1227-1255. http://www.jmlr.org/papers/volume7/jin06a/jin06a.pdf

[49] Ahmed, M., Hu, J., & Cho, J. (2016). Outlier detection with machine learning: A survey. *Computers & Security*, 55, 26-47. https://doi.org/10.1016/j.cose.2015.09.002

[50] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *IEEE International Conference on Data Mining*, 413-422. https://doi.org/10.1109/ICDM.2008.17

[51] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507. https://doi.org/10.1126/science.1127647

[52] Sakurada, M., & Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. *Proceedings of the 2014 IEEE International Conference on Artificial Intelligence and Statistics*, 1, 30-38. https://arxiv.org/abs/1407.0284

[53] Ng, A. Y. (2011). Sparse autoencoder. *CS294A Lecture Notes*, 72-82. http://cs294a.stanford.edu/

[54] Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *Proceedings of the International Conference on Learning Representations (ICLR)*, 1-14. https://arxiv.org/abs/1312.6114

[55] Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*. https://arxiv.org/abs/1606.05908

[56] Chen, J., Song, L., Wainwright, M. J., & Jordan, M. I. (2017). Learning deep generative models of graphs. *Proceedings of the 34th International Conference on Machine Learning*, 70, 1236-1245. http://proceedings.mlr.press/v70/chen17a.html

[57] Zhao, Z., & Zhang, Y. (2017). A comprehensive review of generative adversarial networks. *arXiv preprint arXiv:1712.05420*. https://arxiv.org/abs/1712.05420

[58] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2014). Generative adversarial nets. *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)*, 2, 2672-2680. https://arxiv.org/abs/1406.2661

[59] Zenati, H., Macedo, D., & Djelouah, M. (2018). Efficient deep anomaly detection with a reconstruction objective. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 9474-9483. https://arxiv.org/abs/1806.09500

Leveraging Machine Learning to Enhance the Quality of Non-Financial: Corporate Balance sheet data

AHL MBAREK SALAH – HEAD OF DATA ANALYTICS – October 2024

Plan

Introduction
Problem statement
Goals
Algorithm descriptions
Overall results
Conclusion
Future Outlook

بنك المغرب
BANK AL-MAGHRIB

Introduction

Abstract:

This presentation will focus on a comprehensive study on optimizing outlier detection in financial data using advanced machine learning techniques.

The primary objective was to develop robust algorithms capable of identifying anomalies within the extensive dataset of annual financial statements from Moroccan company's Key methodologies explored include:

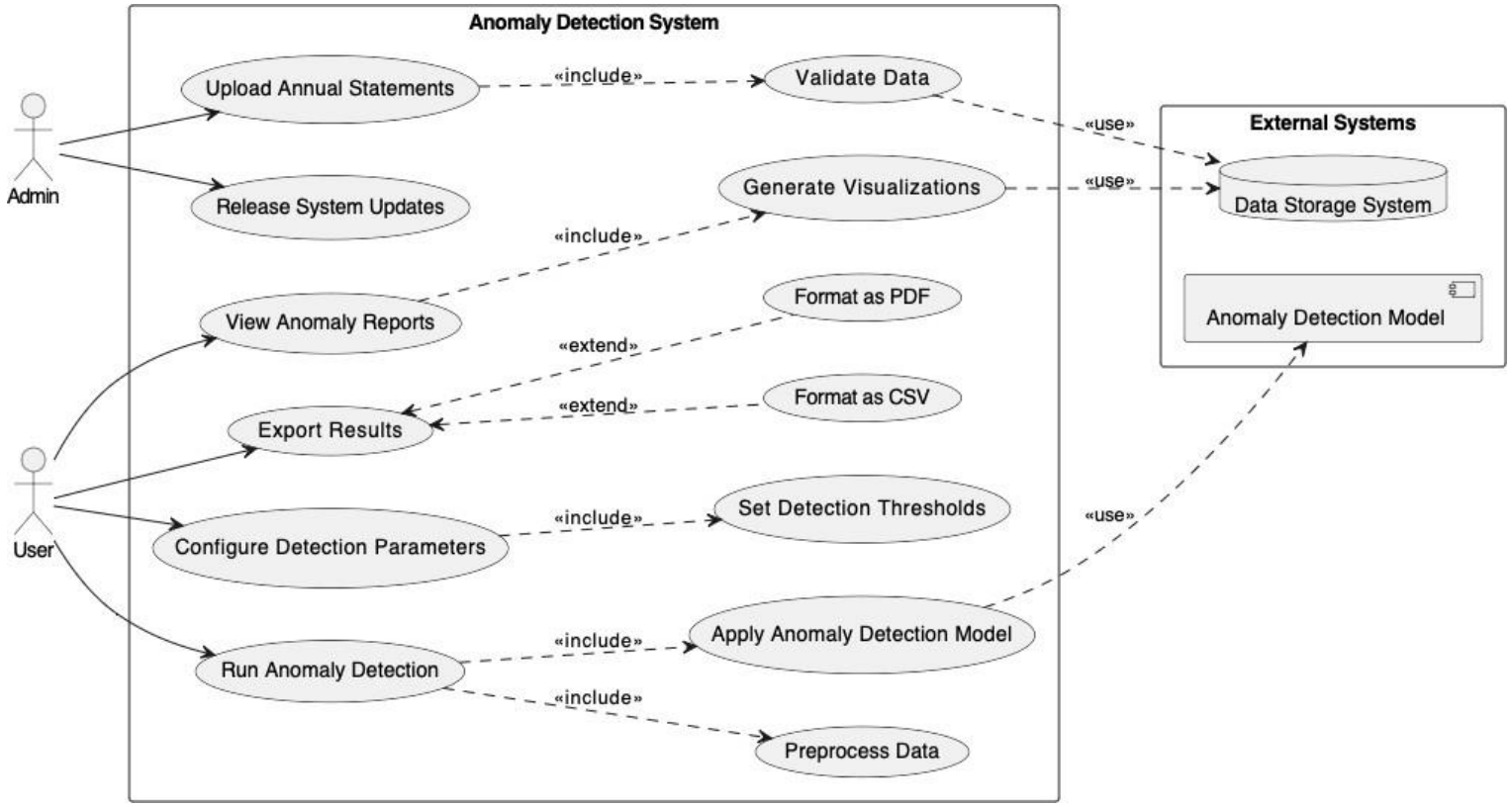
- Isolation Forest,
- Mahalanobis Distance,
- DBSCAN,
- HDBSCAN, and
- Adversarially Learned Anomaly Detection (ALAD),

Along with ensemble techniques like ML-Consensus and Meta-learning. These algorithms were carefully selected and integrated into a unified model to leverage their combined strengths.The implementation involved extensive data preprocessing and feature engineering to ensure data quality. Rigorous testing and validation demonstrated the model's effectiveness in accurately detecting outliers, leading to improved financial data quality and more reliable decision-making.

The solution also includes a user-friendly interface and comprehensive documentation to facilitate adoption and knowledge transfer within Bank Al Maghrib.Future work directions include algorithm refinement, real-time anomaly detection, scalability improvements, integration with other financial systems, and continuous learning.

These efforts aim to maintain the model's relevance and effectiveness in addressing evolving data quality challenges. Overall, this research provides a valuable contribution to the field of anomaly detection and lays a strong foundation for future advancements in data quality management at Bank Al Maghrib.

Workflow



TOOLS

git GitHub

jupyter

Development Tools

python TS

Programming Languages

FastAPI tailwindcss ELECTRON Chart.js

App Development Tools

seaborn pandas Keras NumPy learn TensorFlow

Model Creation Tools

Cucumber Jest

Testing Tools

252 000

Non-financial companies 2023

~ 650 000 financial statements

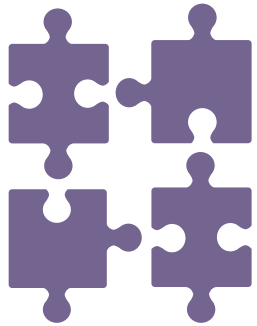
Financial data integrity

بنك المغرب
BANK AL-MAGHRIB

Decision making process







Financial data integrity

Appropriate governance strategy

Problem statement



Manual processing

Time taken away from analysis



3 Weeks

Inefficient use of human resources



8 ETP

Goal

Automate repetitive data quality control tasks to streamline the process and reduce execution time without compromising data quality,

بنك المغرب
BANK AL-MAGHRIB

First Optimization

Automation of outlier detection using Mahalanobis distance

Processing

Near Out of Distribution

Thresholding, predictions around 0.5

Time gains



5 min

HR dedicated



0 HR & « Kafka + AIRFLOW »

Automation of outlier detection using Mahalanobis distance

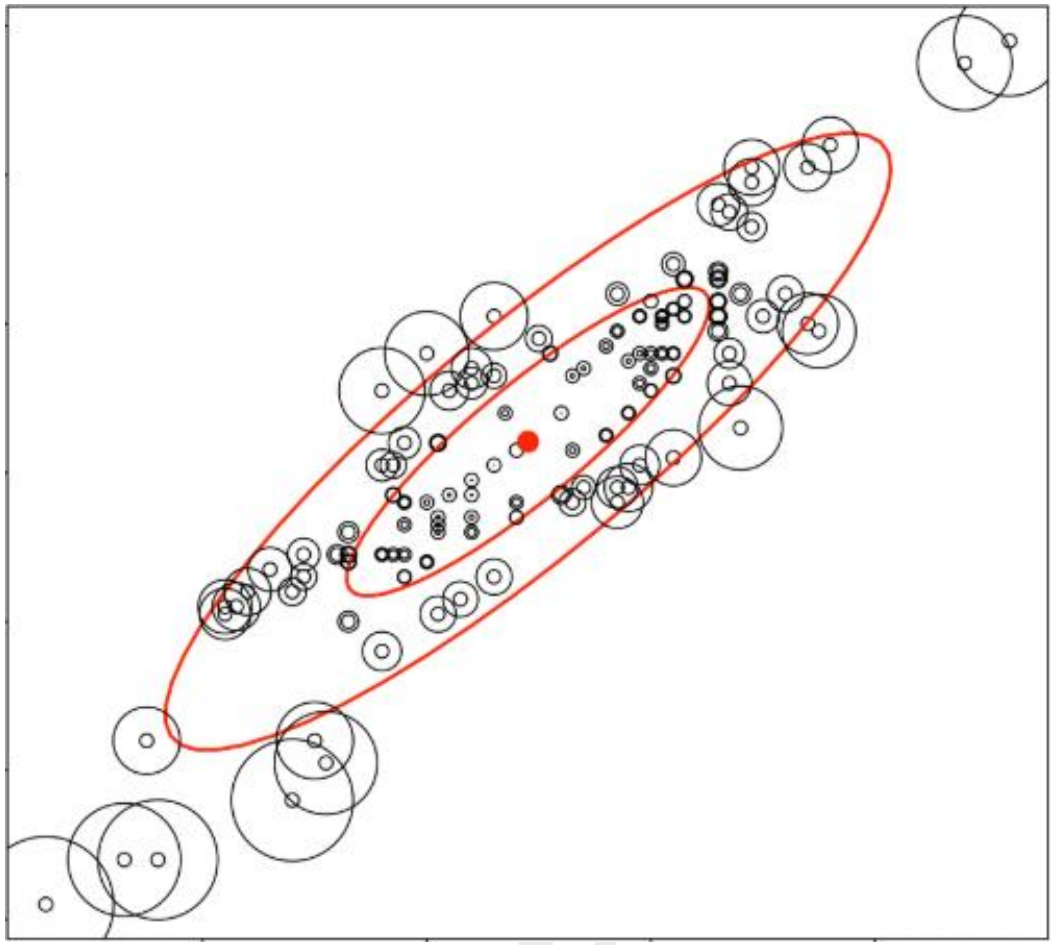
Mahalanobis Distance

Prasanta Chandra Mahalanobis, 1961

Mahalanobis distance is a statistical measure that evaluates how similar a data point is to a distribution. Unlike Euclidean distance, it considers the variability and relationships between variables, providing a more precise measure. It is especially useful for identifying anomalies in multidimensional data.

$$d_M(\mathbf{x}, \mu) = \sqrt{(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)}$$

where X is the observation in question, μ is the mean of the distribution, and ∑ is the covariance matrix



بنك المغرب
BANK AL-MAGHRIB

Automation of outlier detection using ML

Isolation Forest

Fei Tony Liu, 2008.

Isolation Forest is an anomaly detection algorithm using binary trees.

It is based on the assumption that anomalies, being few and different from other data, can be isolated with a few partitions.

It has linear time complexity and low memory usage, making it efficient for large datasets. It is ideal for uniform and normal distributions. Based on isolation trees (iTrees), an isolation score is calculated using the following formula:

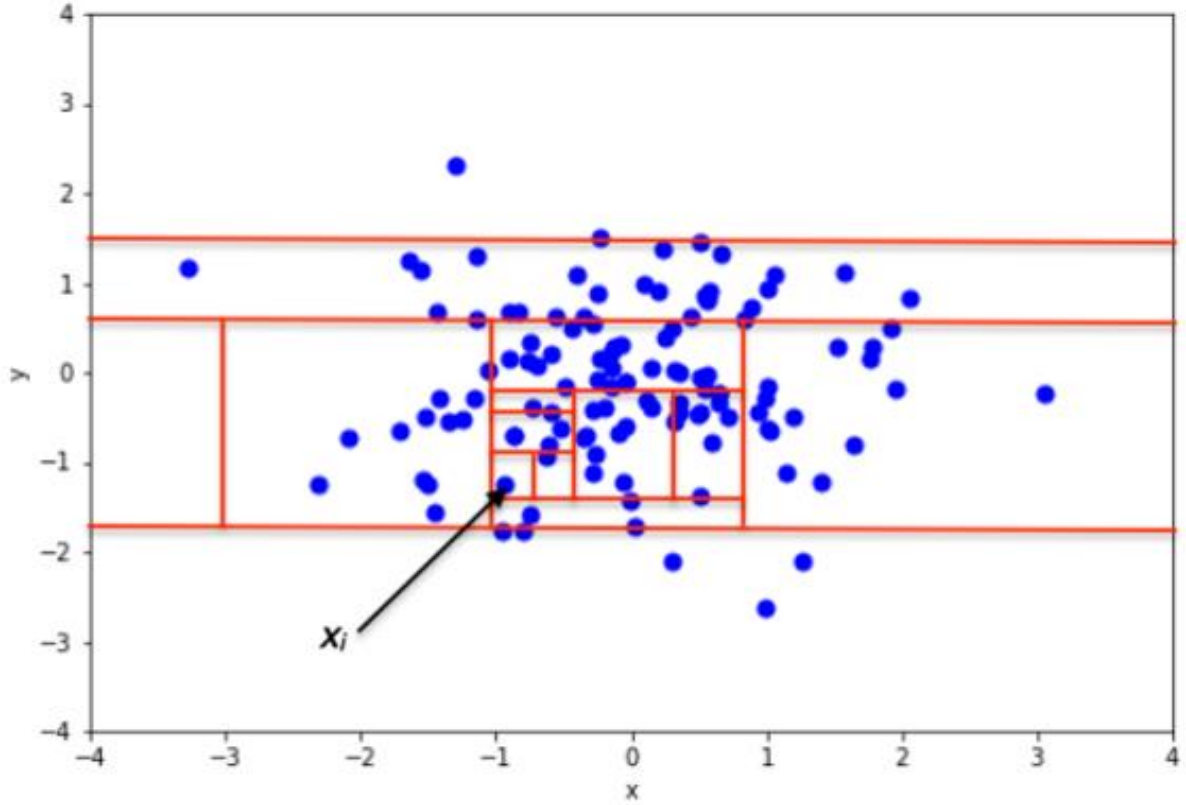
$$s(x) = \frac{2^{\frac{H(x)}{c(n)}} - 1}{c(n) - 1}$$

With:

H(x) representing the average path length for point x across all trees,

c(n) being the average path length for unsuccessful searches in a binary search tree defined by:

$$c(n) = 2 \cdot \left(\frac{n-1}{2} - \frac{2(n-1)}{n} \right)$$

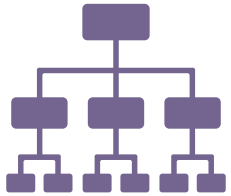


The following image shows an **example of identifying an outlier in a two-dimensional Gaussian distribution**

بنك المغرب
BANK AL-MAGHRIB

Second Optimization

Automation of outlier detection using ML



Isolation Forest

Performance



96% Accuracy compared to manuel check

Non-linear distribution and relationship

Parameter tuning, Designed for data with a uniform distribution.

Automation of outlier detection using deep learning

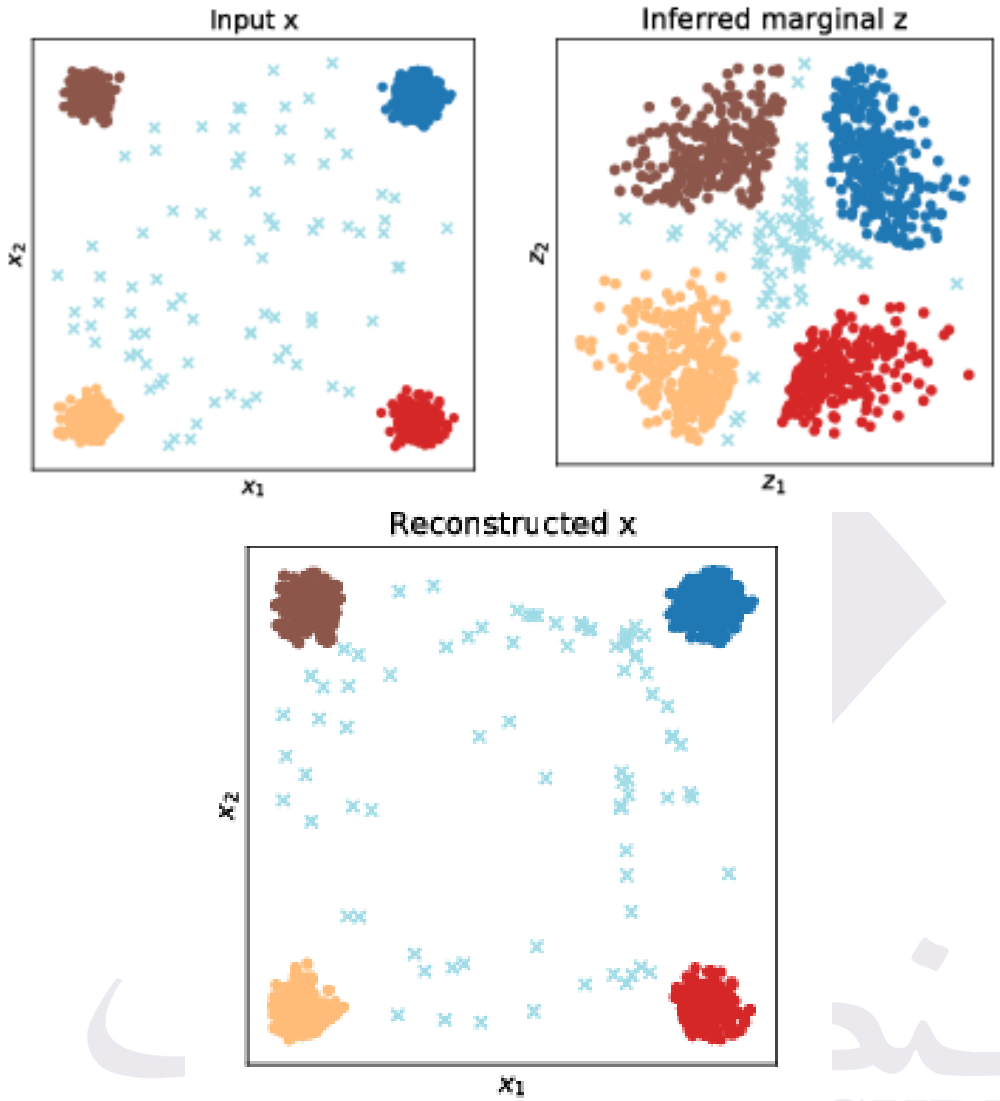
ALAD

Zenati et al., 2018.

Adversarially Learned Anomaly Detection (ALAD) is an anomaly detection method based on generative adversarial networks (GANs). The algorithm relies on two neural networks: a generator, which attempts to reproduce normal data, and a discriminator, which learns to distinguish between normal data and anomalies.

This process allows for the efficient identification of complex anomalies by implicitly modeling non-linear relationships between data. Training complexity is higher due to neural networks, but it can capture complex relationships in high-dimensional datasets. ALAD uses adversarial networks to maximize the detection of subtle anomalies, making it particularly useful for complex and multivariate data

ALAD is a reconstruction-based anomaly detection technique that evaluates the distance between a sample and its reconstruction by the GAN. Normal samples should be reconstructed accurately, while anomalous samples will likely be poorly reconstructed



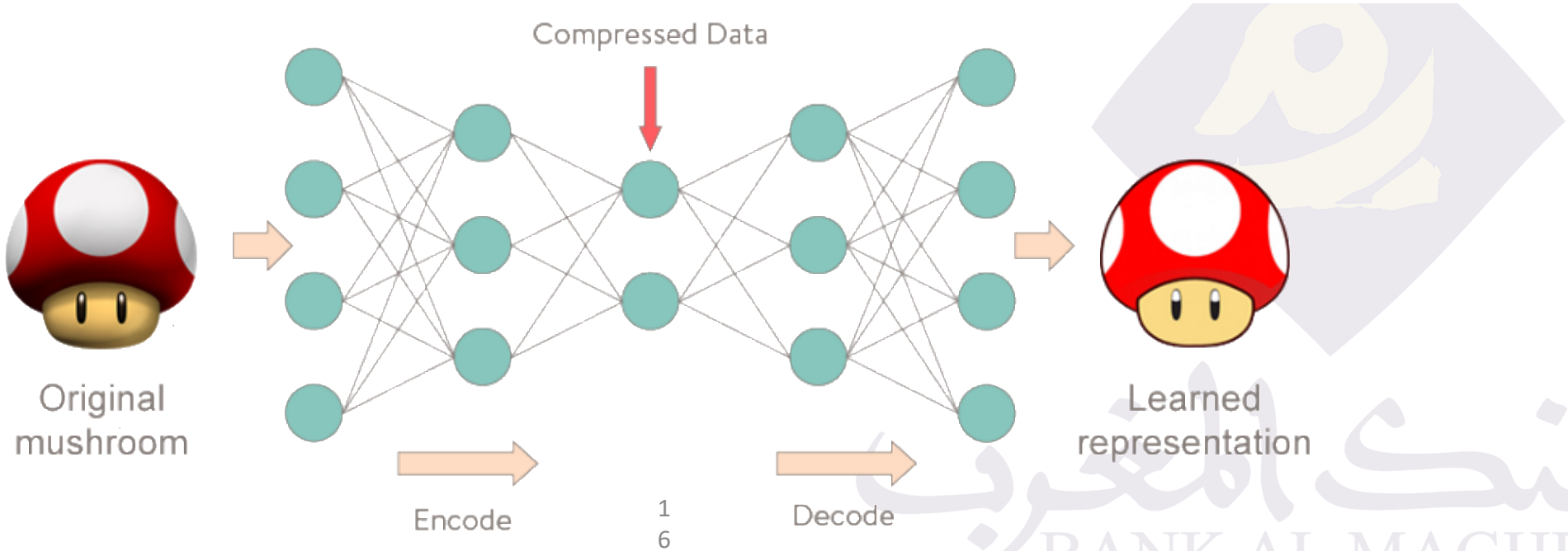
BANK AL-MAGHRIB

Automation of outlier detection using deep learning

Deep Auto-Encoders

Kingma et al., 2013 - Sakurada et al. , 2014 - Xia et al., 2015 - Doersch, 2016.

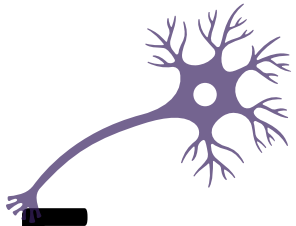
Deep Autoencoders are a reconstruction-based anomaly detection technique. The model learns to compress data into a reduced representation (encoding) before reconstructing it into its original form (decoding). The idea is that normal samples will be well reconstructed, while anomalies, not conforming to the learned pattern, will have a poor reconstruction. Computational complexity increases with the size of the neural network, but autoencoders can capture complex non-linear relationships in high-dimensional datasets. They are particularly suited for detecting subtle anomalies in multivariate data.



BANK AL-MAGHRIB

4th Optimization

Automation of outlier detection using deep learning



Auto-Encoders

Performance



28 ms

Non-linear distribution and relationship

Very heavy for simpler datasets

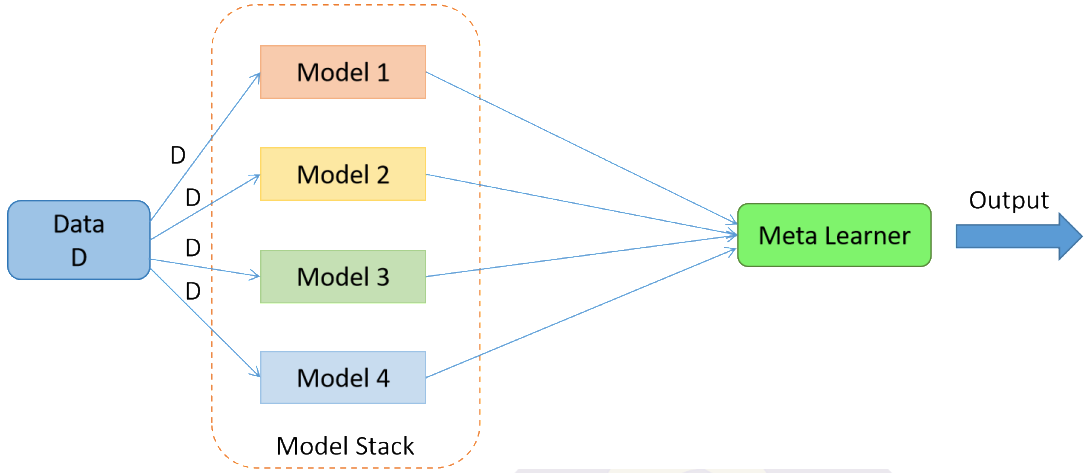
Automation of outlier detection by ML-Consensus

ML-Consensus

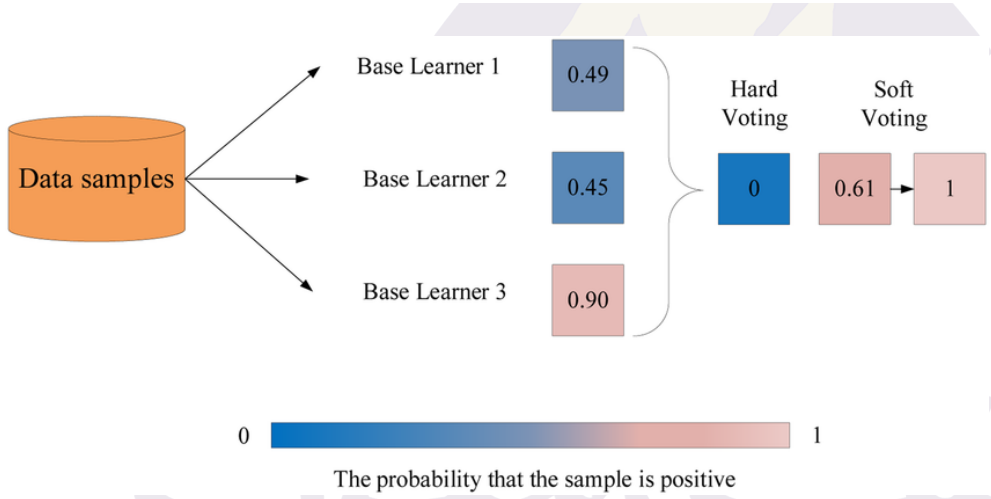
Breiman, 1996.

Consensus in ML is an approach that combines the predictions of multiple models to reach a more robust decision. The idea is to converge multiple models (often of different types) towards a common prediction, thereby reducing individual errors.

Consensus can be achieved through various methods such as averaging predictions, majority voting, or more advanced methods. This improves robustness, especially when models have varying performance on different parts of the data.



Stacking Method



Hard vs Soft Voting Methods

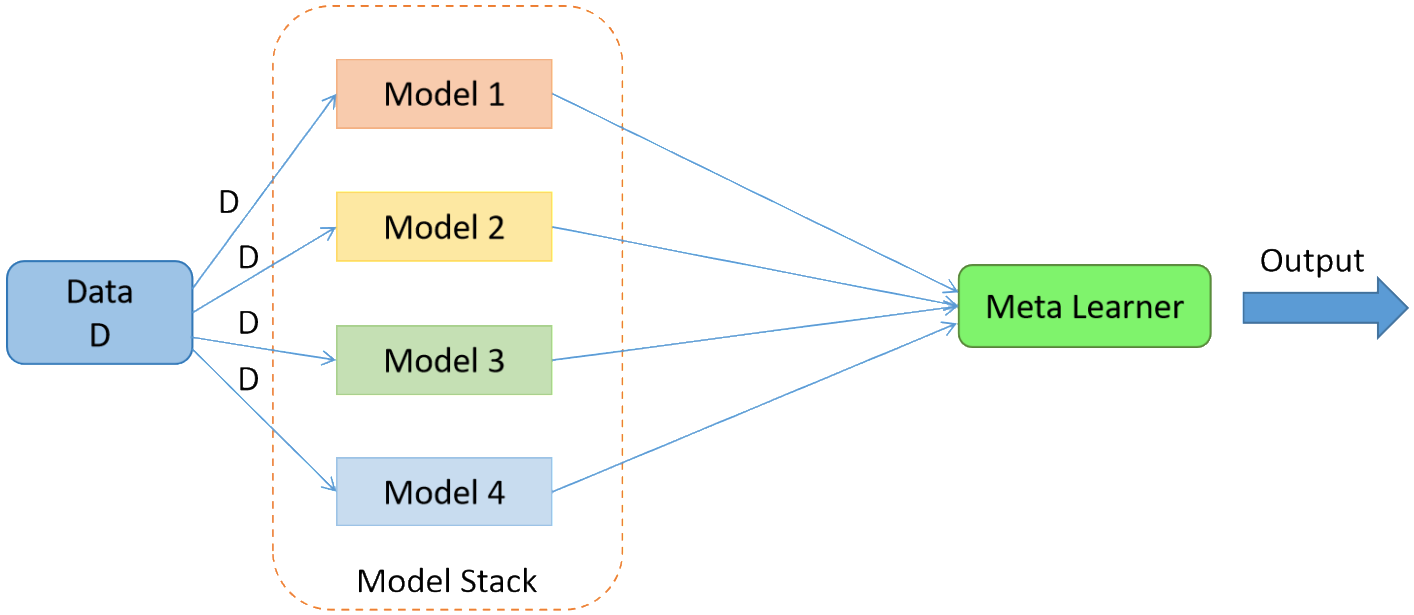
BANK AL-MAGHRIB

Automation of outlier detection by ML-Consensus

Stacking & Meta Learning

Wolpert, 1992 - Breiman, 2001.

Stacking is a more sophisticated ensemble technique where multiple base models (called "level-0 models") are trained on the same data, and a meta-model (or "level-1 model") is then used to combine the predictions of these base models. The goal of stacking is to leverage the strengths of each model to improve the final prediction. Unlike methods like bagging or boosting where each model contributes equally or progressively, stacking uses an additional model to determine how to combine the outputs of the base models.

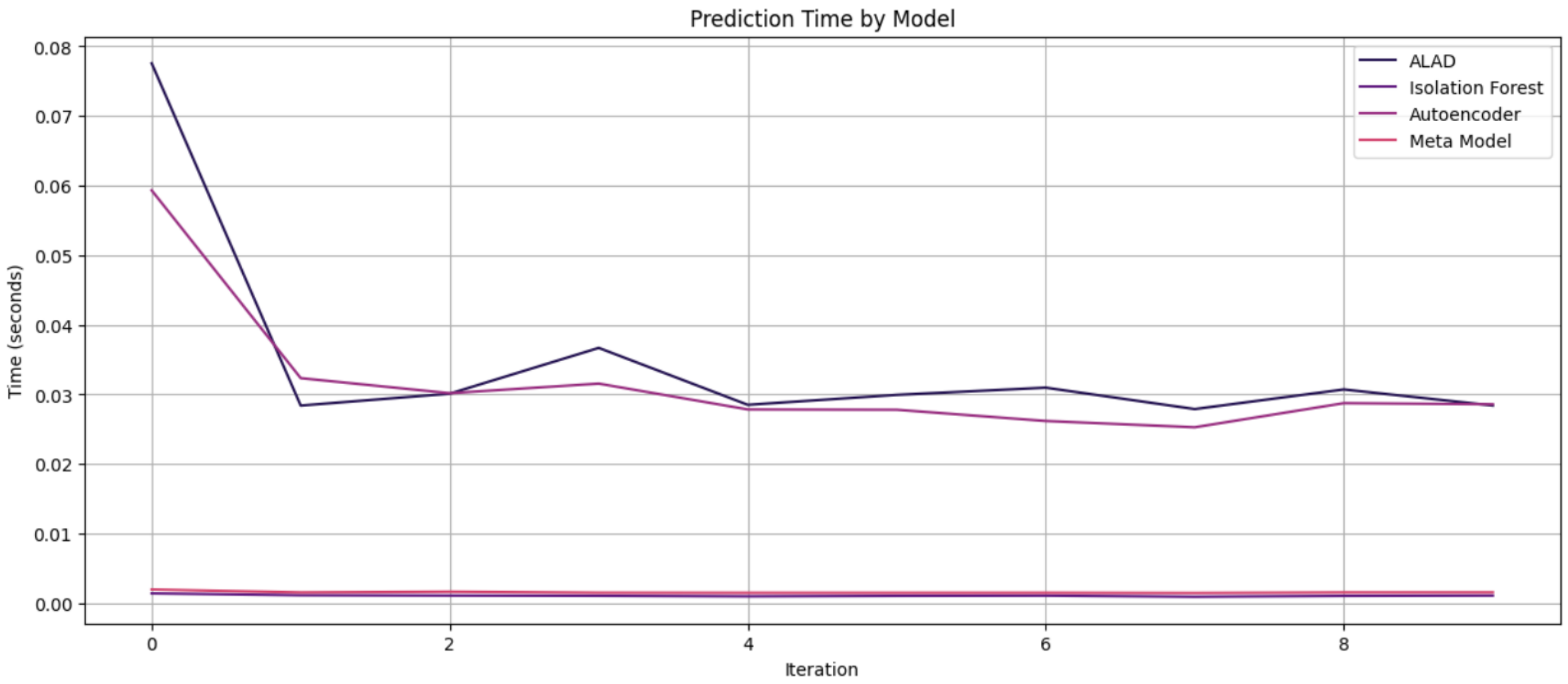


Stacking Method

بنك المغرب
BANK AL-MAGHRIB

Automation of outlier detection by ML-Consensus

(Stacking, Random Forest Meta Model)



BANK AL-MAGHRIB

Conclusion

“When the calculator was first invented, people feared that it would replace mathematicians or make people less capable, but on the contrary, it allowed us to perform more calculations and unleash greater potential.”

- Sam Altman, CEO, OpenAI

بنك المغرب
BANK AL-MAGHRIB

Future Outlook

- **Algorithm and hyperparameter tuning.**
- **Scalability and performance optimization (Model and Desktop App).**
- **Integration with other financial systems and/or tools.**

بنك المغرب
BANK AL-MAGHRIB

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Superstar firms, intangibles, and productivity mismeasurement: evidence from firm-to-firm transactions[1]

Gert Bijnens,
National Bank of Belgium

Joep Konings,
Nazarbayev University GSB, KU Leuven and CEPR

Aaron Putseys,
KU Leuven

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

KU LEUVEN

VIVES



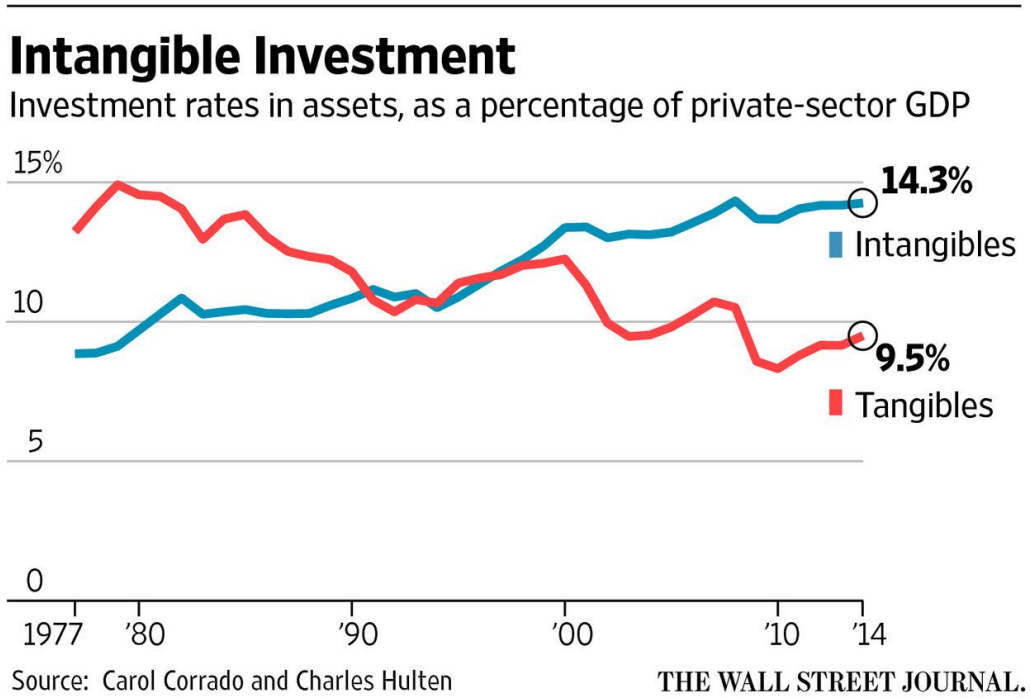
Superstar Firms, Intangibles, and Productivity Mismeasurement: Evidence From Firm-To Firm Transactions

Madrid
October 17, 2024

Bijnens, G. – National Bank of Belgium
Konings, J. – KU Leuven, Nazarbayev University GSB, and CEPR
Putseys, A. – KU Leuven

Motivation

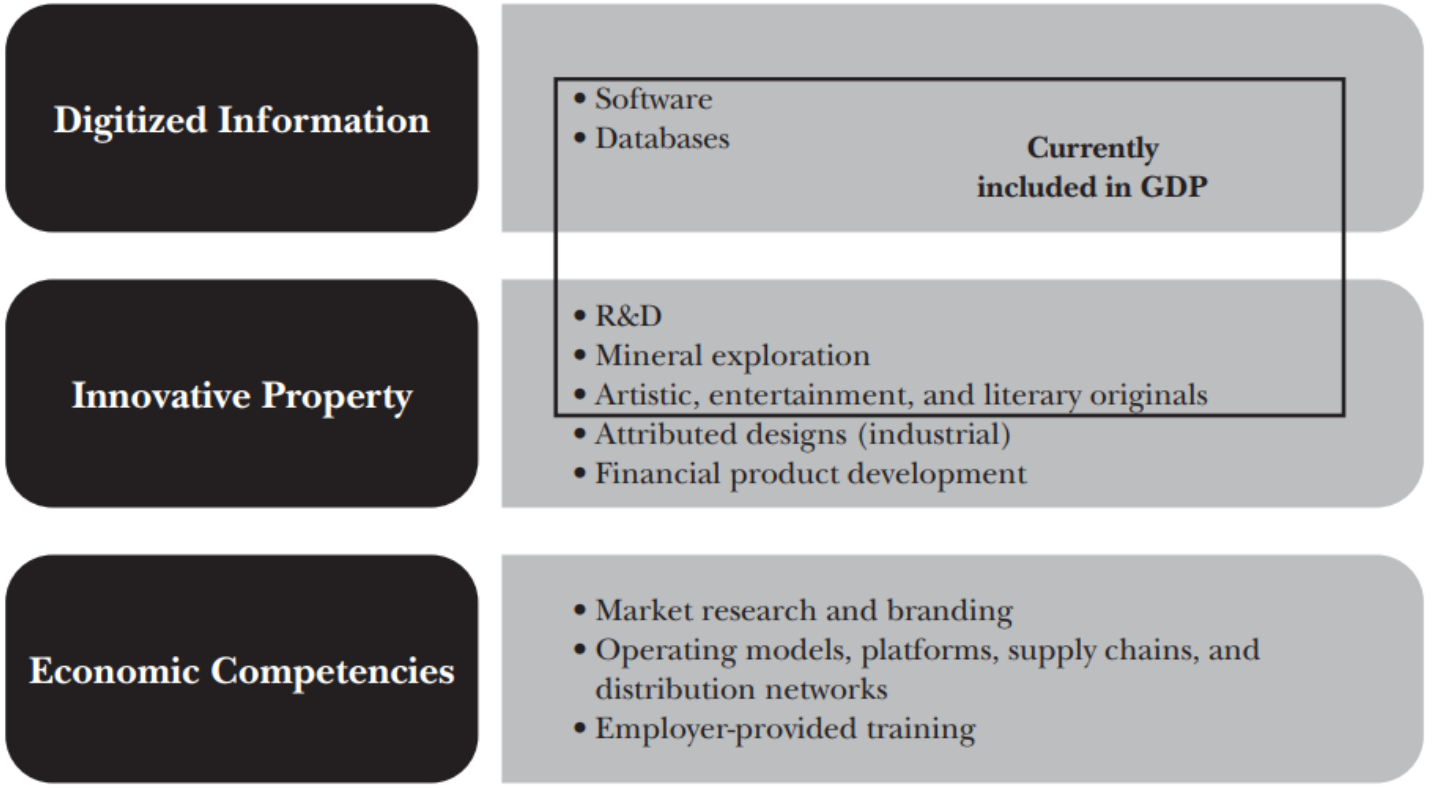
- Solow Paradox, 1986: *"You can see the computer age everywhere but in the productivity statistics"*
- Brynjolfsson et al., 2021: "*Solow's Paradox is not unique. It is one example of a more general phenomenon resulting from the need for intangible investments in early stages of new general-purpose technologies*"
- The past decades witnessed a rapid increase in the use of intangible investment
- The measurement of productivity crucially depends on the measurement of intangible assets



Motivation

Can we just use the balance sheet information on intangibles?

- International Accounting Standard generally disallows the capitalization of most intangible assets
- Balance sheet intangibles largely arise from acquisitions
- Reported assets do not reflect actual intangible investment flows



Source: Authors' elaboration of Corrado, Hulten, and Sichel (2005, 2009).

Motivation

Can we just use the **balance sheet information** on intangibles?

- International Accounting Standard generally disallows the capitalization of most intangible assets
- Balance sheet intangibles largely arise from acquisitions
- Reported assets do not reflect actual intangible investment flows

Can we use **survey data** on intangibles?

- Survey information on intangible asset results in a selection bias and a lack of insights on heterogeneity in returns on intangibles
- Not all intangibles captured

Research objectives

Investigate the returns from intangible capital at the micro level and examine how this affects productivity patterns at the macro level.

How:

- Trace *all* yearly intangible asset purchases of a firm from a B2B transaction level dataset
- Build a firm(-year) level intangible capital stock (2002-2019)
- Estimate an augmented production function
- Compare returns from intangible capital across the firm size distribution
- Examine how these findings at the firm level affect productivity growth at the macro level

Data

National Bank of Belgium (2002-2019):

- VAT customer listing (yearly)
- VAT declarations (periodical)
- Research and development expenditure (R&D)

from ECOOM Leuven

- Annual company accounts

} Intangible investments and tangible investments

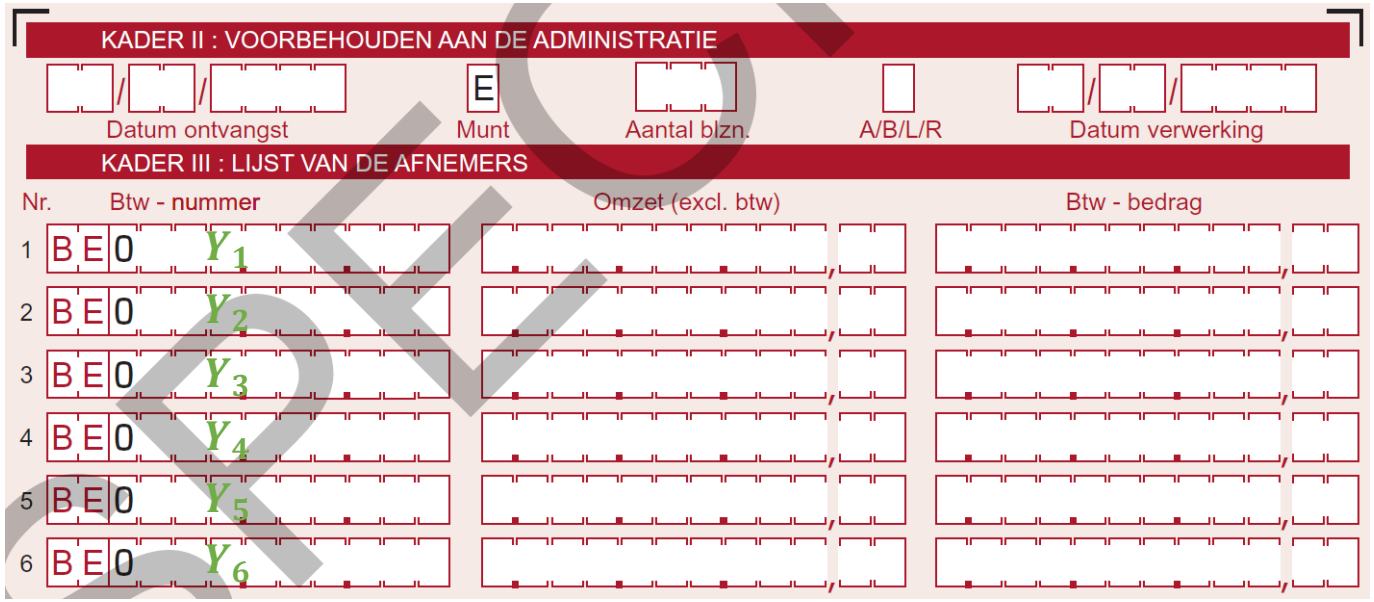
Additional data:

- 2-digit deflators from Eurostat
- 2-digit deflators from EU Klems

→ Panel of 1,603,254 firm-year observations over the period 2002-2019

Data: VAT declarations (I)

- Take the VAT customer listing of computer programmer **X**, active in NACE 6201
- From the VAT form of firm **X**, we know to which firms **Y** it sells and the value of the sales
- Since we have the yearly VAT transactions of *ALL* **X** firms, we can deduce for each firm **Y** how much it purchases from type **X** firms. The sum of all intangible purchases in year *t* across type **X** firms, is our intangible investments measure of year *t* for firm **Y**



Data: Intangible producing industries

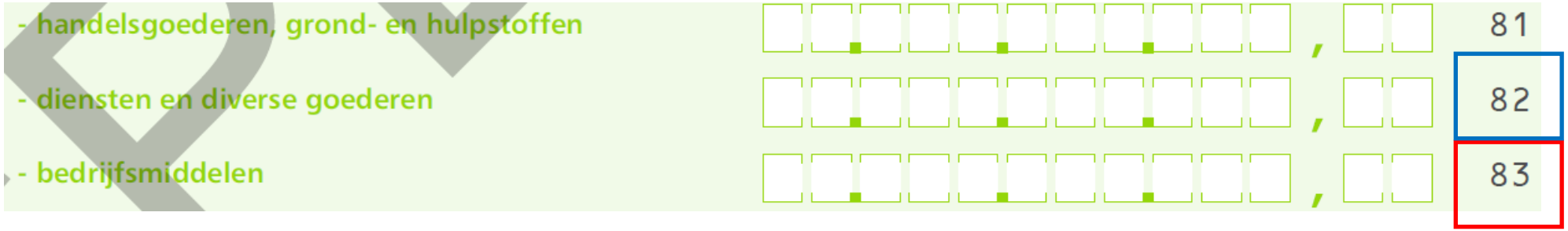
Intangible Asset Type	NACE Rev. 2 Code
	A. Computerized Information
(i) Software	6201, 6209, 6312
(ii) Database	6202, 6203, 6311
	B. Innovative Property
(i) R&D (Scientific)	7211, 7219, 7220
(ii) Entertainment & Artistic Originals	7410, 7420, 7430, 7911, 7912, 7990, 8810, 8891, 8899, 9001, 9002, 9003, 9004, 9311, 9312, 9313, 9319, 9321, 9329, 9604
(iii) Design & other new Product/Systems	7111, 7112, 7120
	C. Economic Competencies
(i) Advertising/Market Research	7311, 7312, 7320
(ii) Employer-provided Training/Organizational Structures	7021, 7022, 7490, 7810, 7830, 8510, 8520, 8531, 8532, 8541, 8542, 8551, 8552, 8559, 8560

Notes: Table 1 shows the NACE Rev. 2 that produce or sell intangible services. Intangible asset categories and types are based on Corrado et al. (2005, 2009, 2012). Table A.1 shows the description of each NACE Rev. 2 Code considered as intangible providers.

Data: VAT declarations (II)

From periodical VAT transactions, we know total intangible investments.

From the VAT declarations, firms must report total purchases of all goods that have an investment character. Intangible investment are mostly considered as a service, and hence not included here



Data

- Use the Perpetual Inventory Method to construct the firm-year capital stock

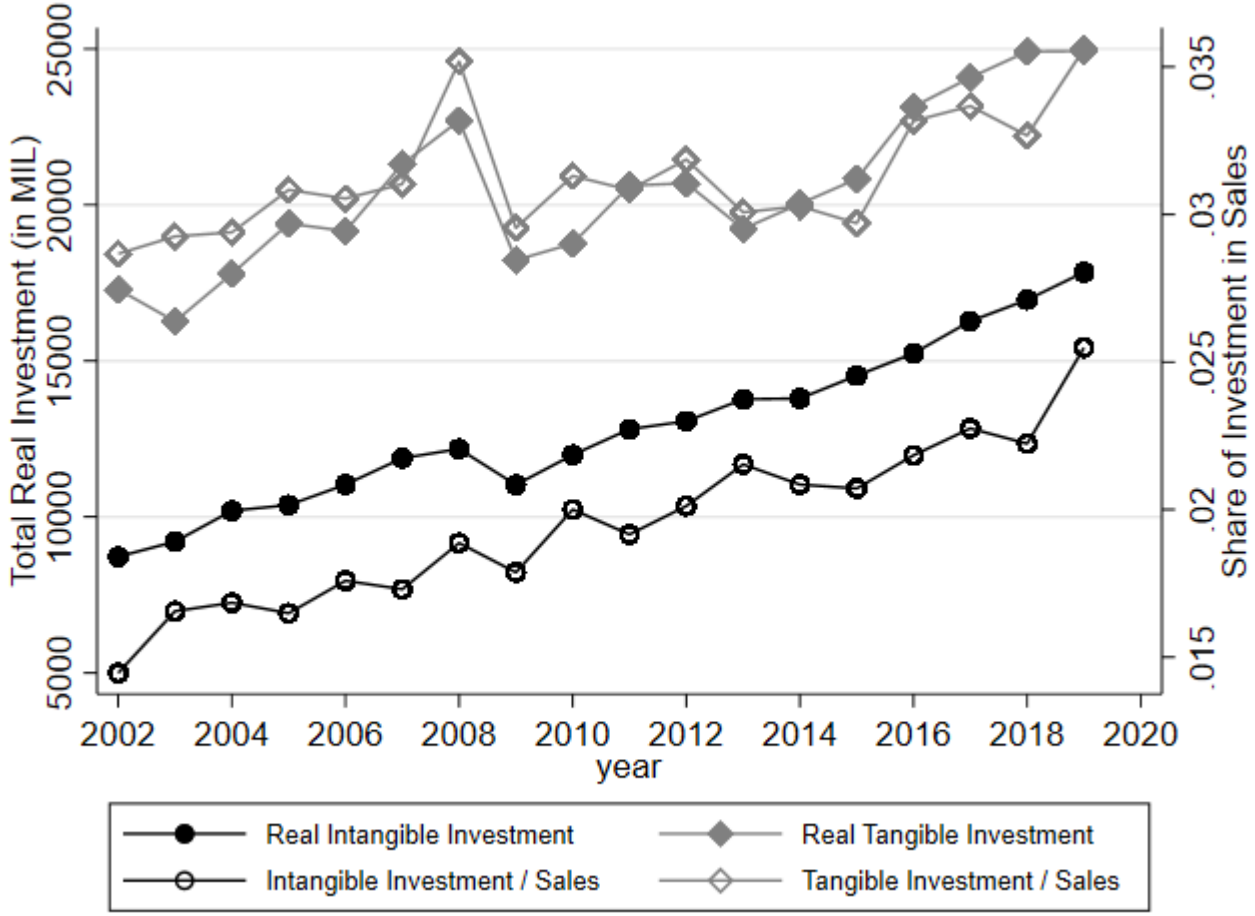
$$K_{it}^{U} = K_{it-1}^{U} \times (1 - \delta_{U}) + I_{it}^{U}$$

- Depreciation rates δ^{U} are obtained from EU Klems
- Investments $\tilde{I}_{it}^{U}$ are obtained from VAT and ECOOM data
- Initial intangible capital stock based on average intangible investment intensity in the first three years

Descriptives

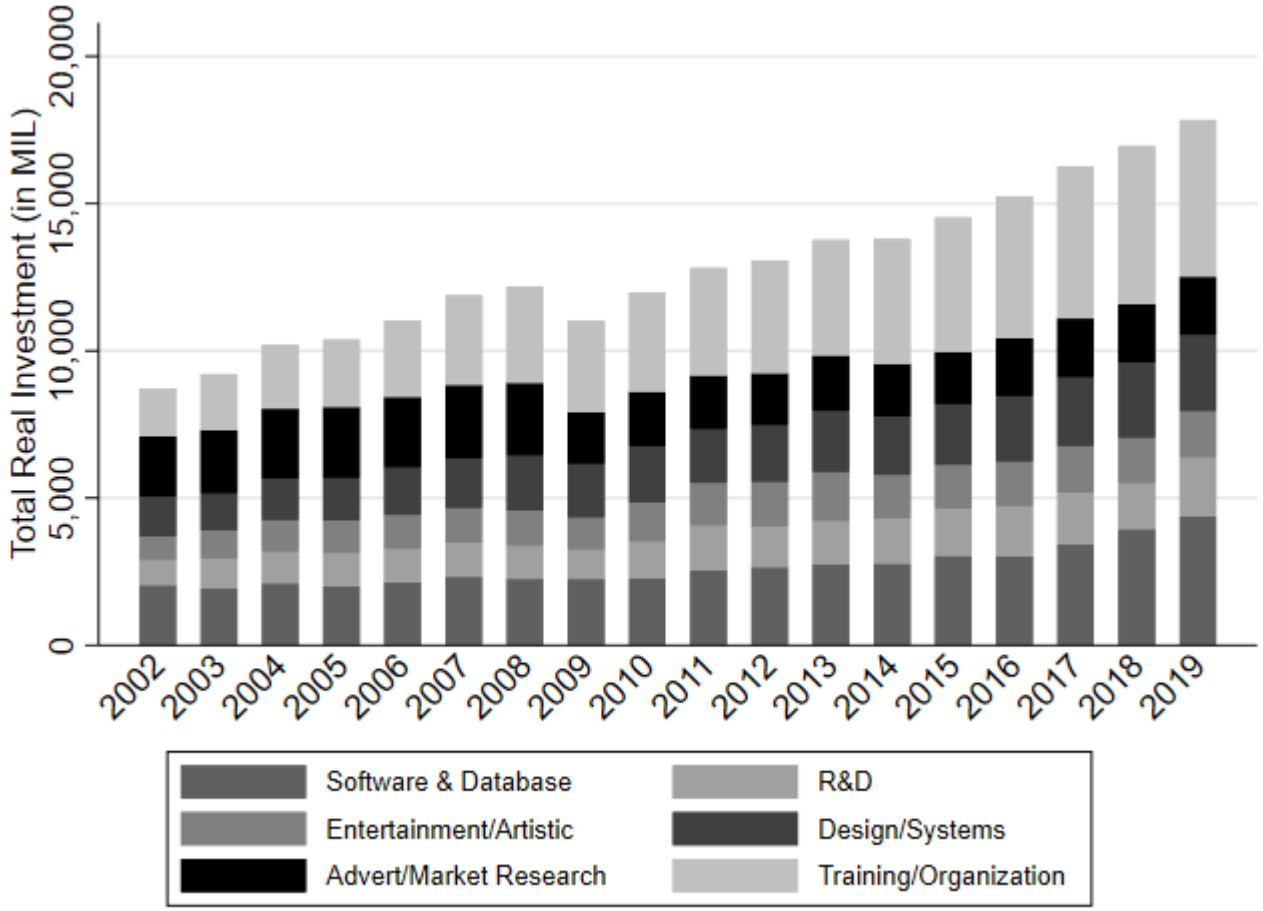
Robustness

- Fact 1: Intangibles have become more important over time.
  - Investment in intangible capital rose from **8** billion EUR in 2002 to almost **17.5** billion EUR in 2019.



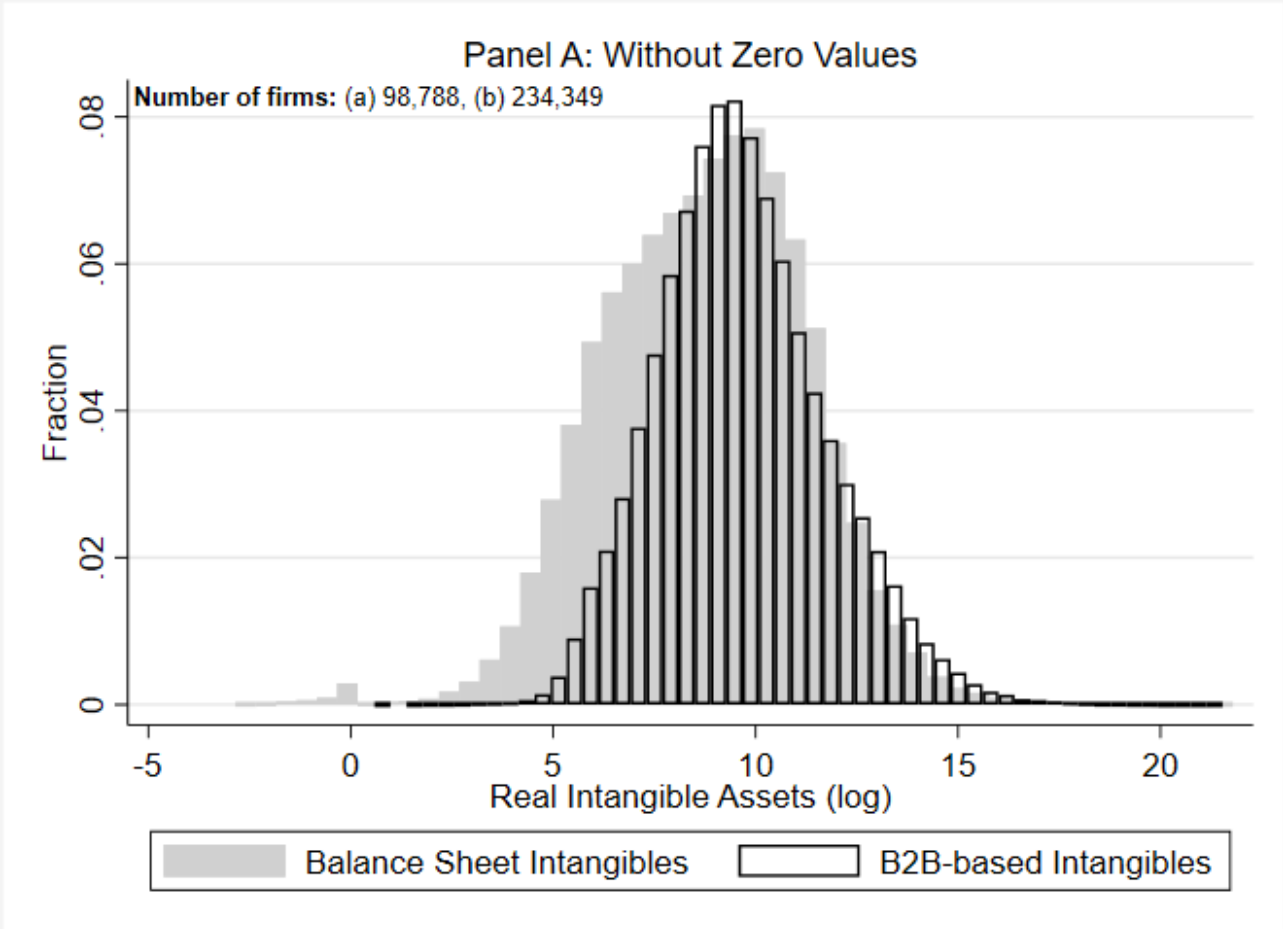
Descriptives

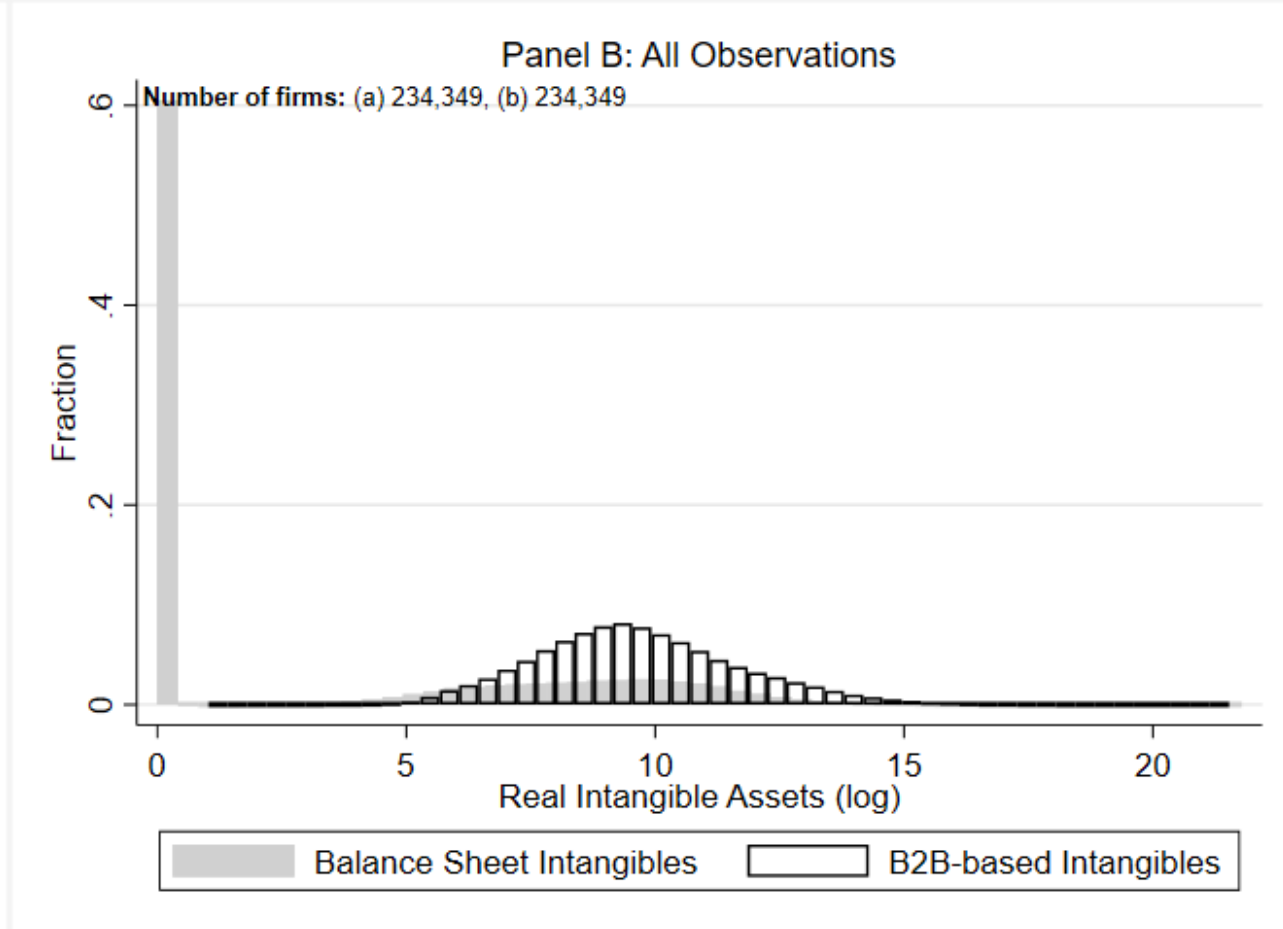
- Fact 1: Intangibles have become more important over time.
  - This increase is mainly driven by the increasing importance of **training/organizational intangibles**
  - **Software/database** investments in 2018-2019 (AI?)



Descriptives

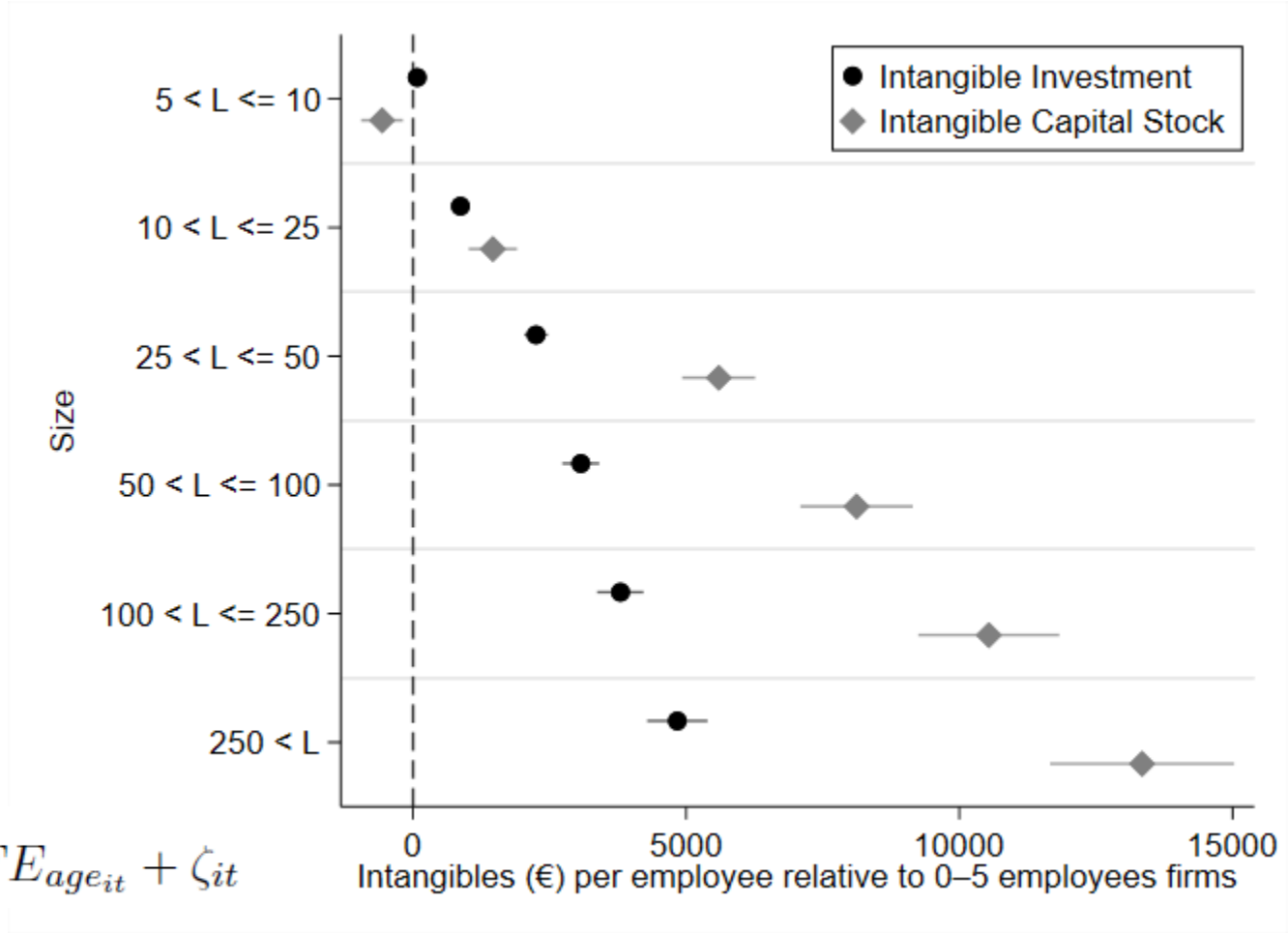
- Fact 2: The B2B-based measure on intangibles is more comprehensive than its balance sheet counterpart





Descriptives

- Fact 3: Bigger firms have a larger amount of intangible capital.
  - a firm with more than 250 employees invests close to €5000 per employee more in intangibles than firms with 0-5 employees



$$Intangiblesperemployee_{it} = FE_{Size_{it}} + FE_{ind_{it}t} + FE_{age_{it}} + \zeta it$$

Descriptives

$$log(Intangiblesperemployee_{it}) = \eta log(Size_{it}) + FE_{ind_{it}t} + FE_{age_{it}} + \zeta_{it},$$

	Within-industry			Within-firm		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Intangible Investment per Employee						
Size (Proxied by Added value)	0.3328***			0.2956***		
	(0.0009)			(0.0035)		
Size (Proxied by Sales)		0.3577***			0.2458***	
		(0.0009)			(0.0041)	
Size (Proxied by Inputs)			0.2977***			0.1157***
			(0.0008)			(0.0037)
Observations	1,425,022	1,425,022	1,424,876	1,425,022	1,425,022	1,424,876
A. Intangible Assets per Employee						
Size (Proxied by Added value)	0.4067***			0.0537***		
	(0.0009)			(0.0027)		
Size (Proxied by Sales)		0.4542***			0.0671***	
		(0.0009)			(0.0034)	
Size (Proxied by Inputs)			0.3880***			0.0325***
			(0.0008)			(0.0029)
Observations	1,638,584	1,638,584	1,638,346	1,638,584	1,638,584	1,638,346

Notes: Table 3 shows the regression results of Equation 2. The dependent variable is the logarithm of intangibles per employee (Panel A considers intangible investments, while panel B takes intangible assets as outcome variable). The independent variable is the logarithm of firm size proxied by added value, sales, and inputs. The time period is 2002-2019. All columns include a full set of four-digit industry classification fixed effects interacted with year fixed effects, and within-industry results also include a full set of age fixed effects. An elasticity of 0.3 means that raising intermediates by a factor of 2 raises intangible investment per employee by 30 percent. Within-firm standard errors are clustered at the level of the firm. *p<0.10, **p<0.05, ***p<0.01.

Increasing added value by a factor of 2, raises intangible investment per employee by 33 percent.

Methodology

- Augmented Cobb Douglas production function including both intangible and tangible capital as inputs (as in Brynjolfsson et al. 1996 and Tambe & Hitt, 2012)

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \beta_u u_{it} + \omega_{it} + \epsilon_{it}$$

- Potential bias in labor and capital coefficients because of unobserved productivity

  → We use control function approach to correct for endogeneity of inputs (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015).

Results

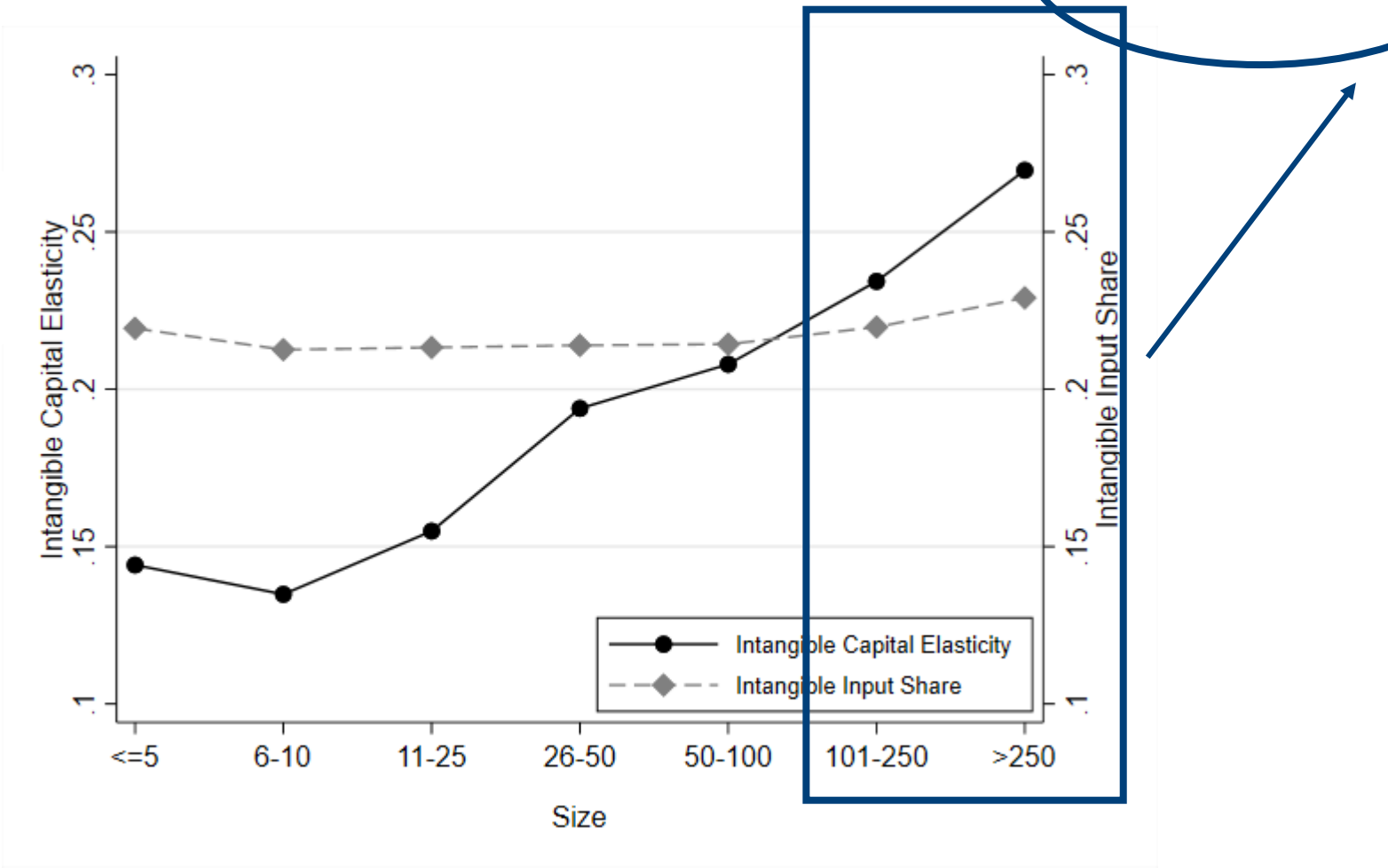
Dependent Variable:	ln(Y)	ln(Y)	ln(Y)	ln(Y)
Estimator:	OLS	Firm FE	First Differences	ACF
Sample:	Private Sector	Private Sector	Private Sector	Private Sector
	(1)	(2)	(3)	(4)
ln(K)	0.1526***	0.0928***	0.0493***	0.1217***
	(0.0008)	(0.0007)	(0.0007)	(0.0023)
ln(L)	0.6210***	0.4799***	0.3074***	0.5968***
	(0.0012)	(0.0016)	(0.0017)	(0.0029)
ln(U)	0.1630***	0.1121***	0.1153***	0.1718***
	(0.0008)	(0.0009)	(0.0010)	(0.0026)
Avg(MP_K)	0.1074	0.0653	0.0347	0.0857
Avg(MP_L)	1.0403	0.8040	0.5149	0.9998
Avg(MP_U)	0.7524	0.5176	0.5322	0.7934
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1,603,254	1,603,254	1,349,804	1,349,634
No. of firms	183,759	183,759	183,759	183,759

Notes: Table 4 shows the production function estimates of Equation 3 estimated using OLS, firm fixed effects, first differences, and the control function approach introduced by Ackerberg et al. (2015). Standard errors are clustered at the firm level. The estimation sample is identical in number of firms. The number of observations is lower due to missing lags. All estimations include year fixed effects and two-digit industry fixed effects, except for columns 2 and 3 where the industry fixed effects are observed by the firm fixed effects and first differences respectively. $*p<0.10$, $**p<0.05$, $***p<0.01$.

Results

Large firms benefit more from Intangibles

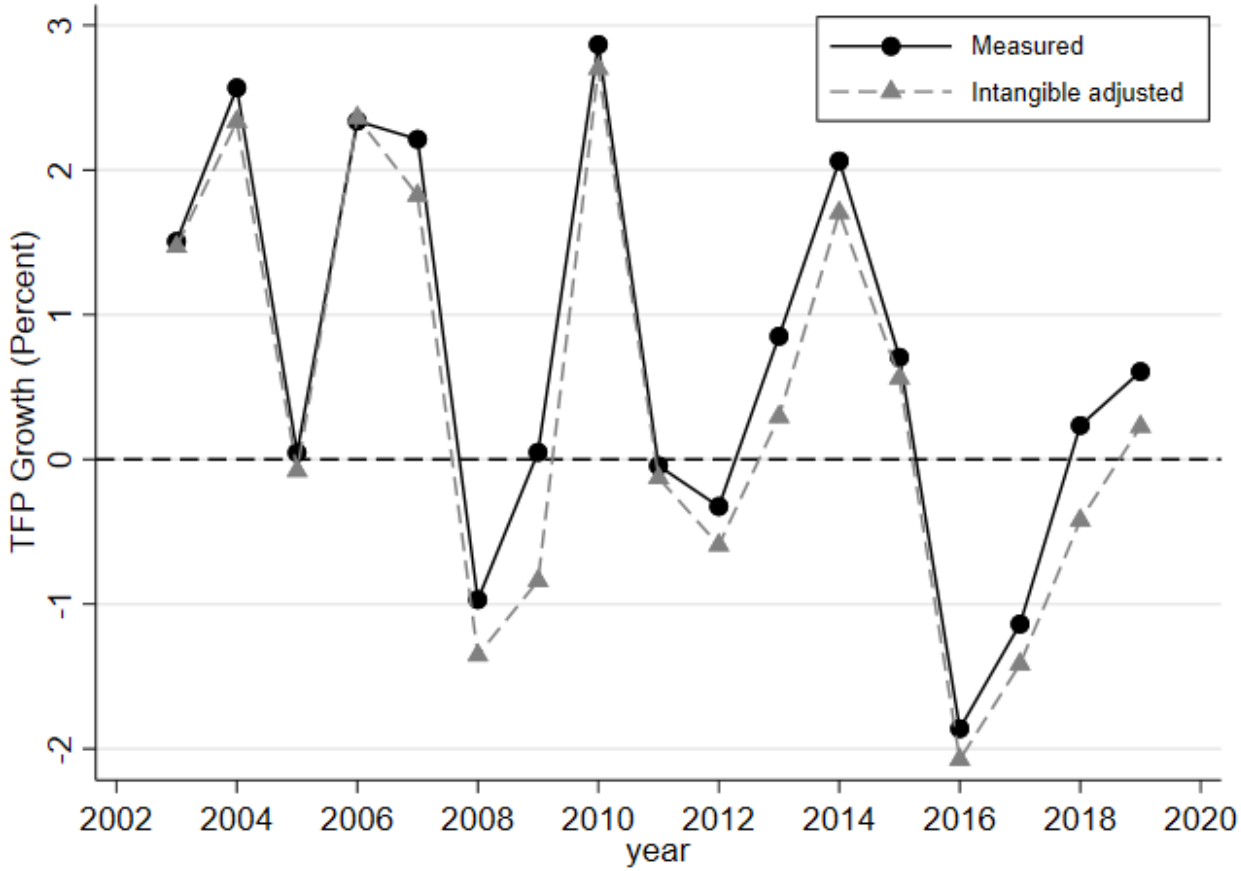
$$MP_U = \frac{\delta Y}{\delta U} = \frac{\delta Y}{\delta U}\frac{U}{Y}\frac{Y}{U} = \beta_u \frac{Y}{U} = \frac{\beta_u}{\frac{U}{Y}}$$

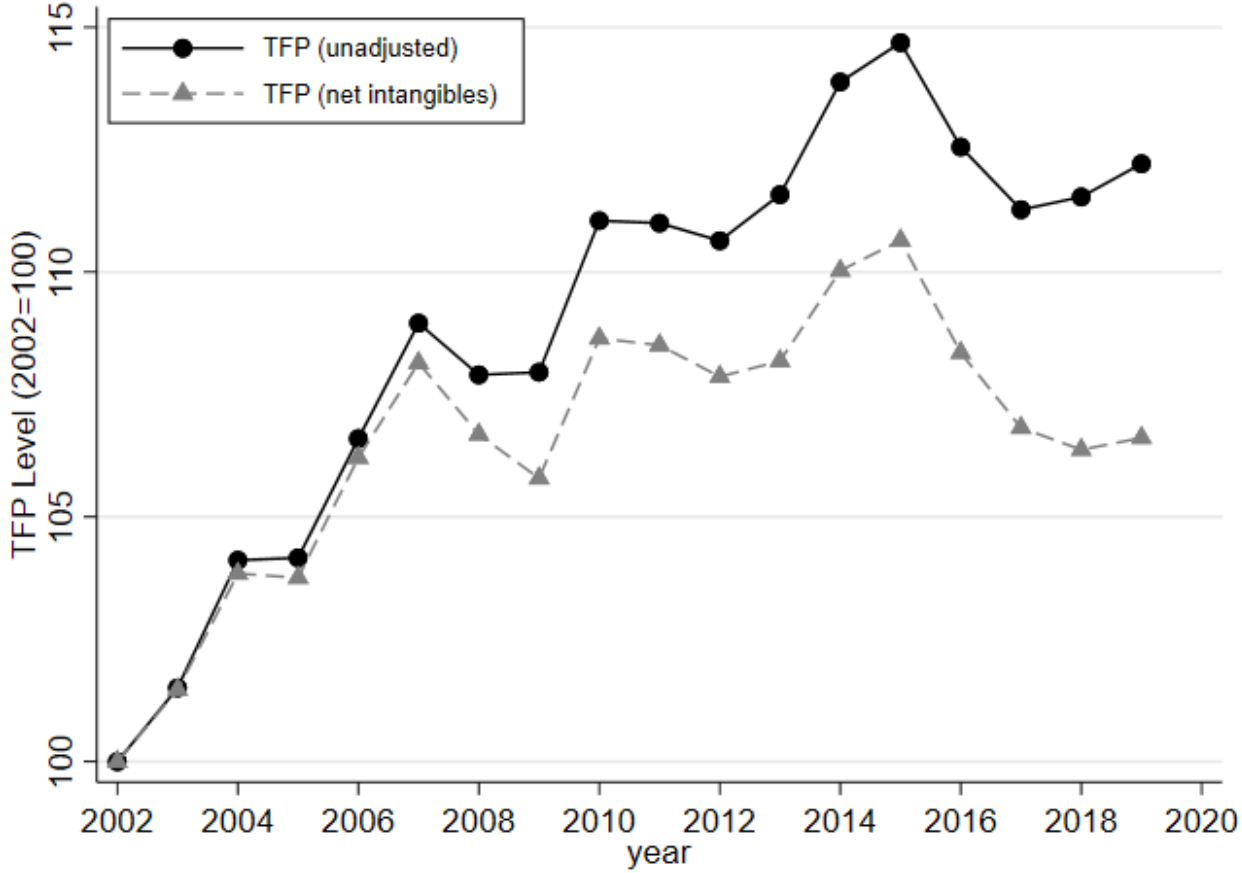


Excess Returns only for big firms

Results

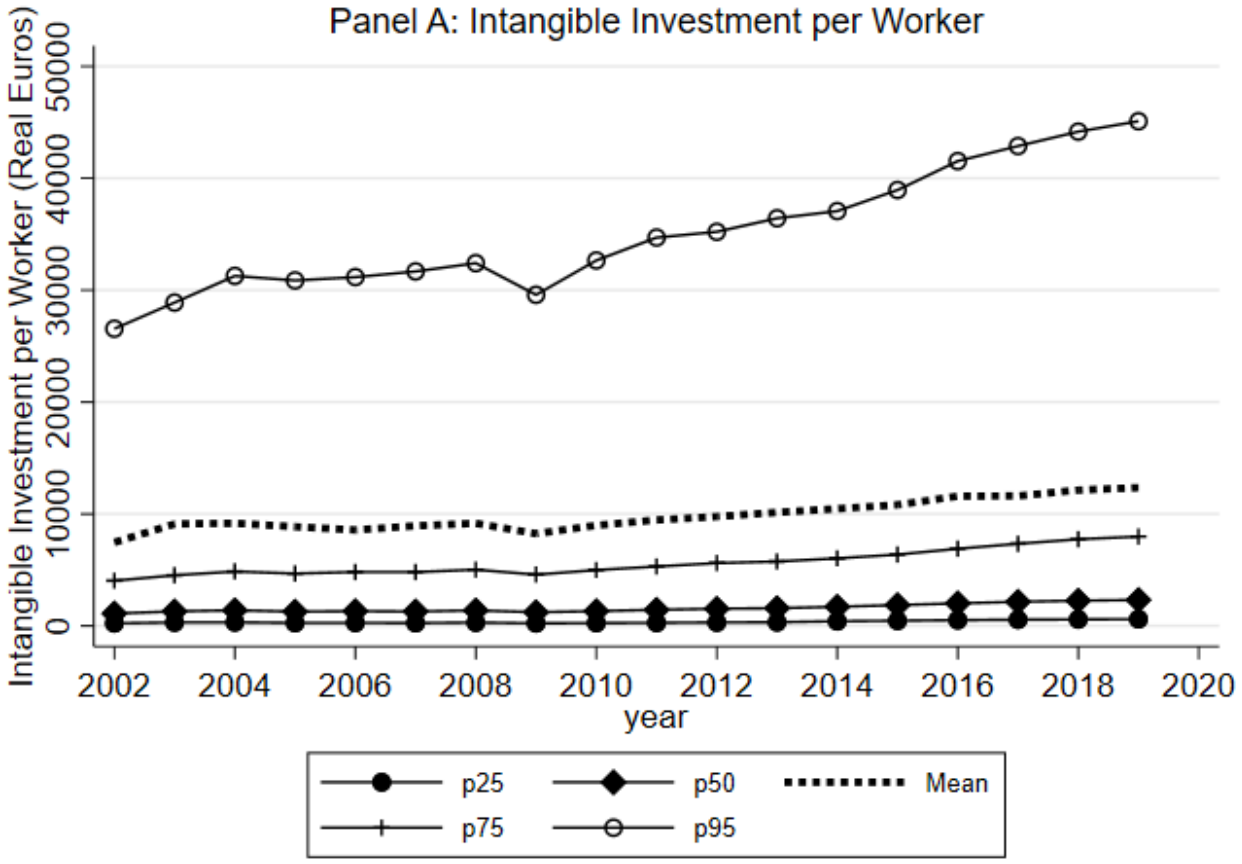
Productivity Mismeasurement

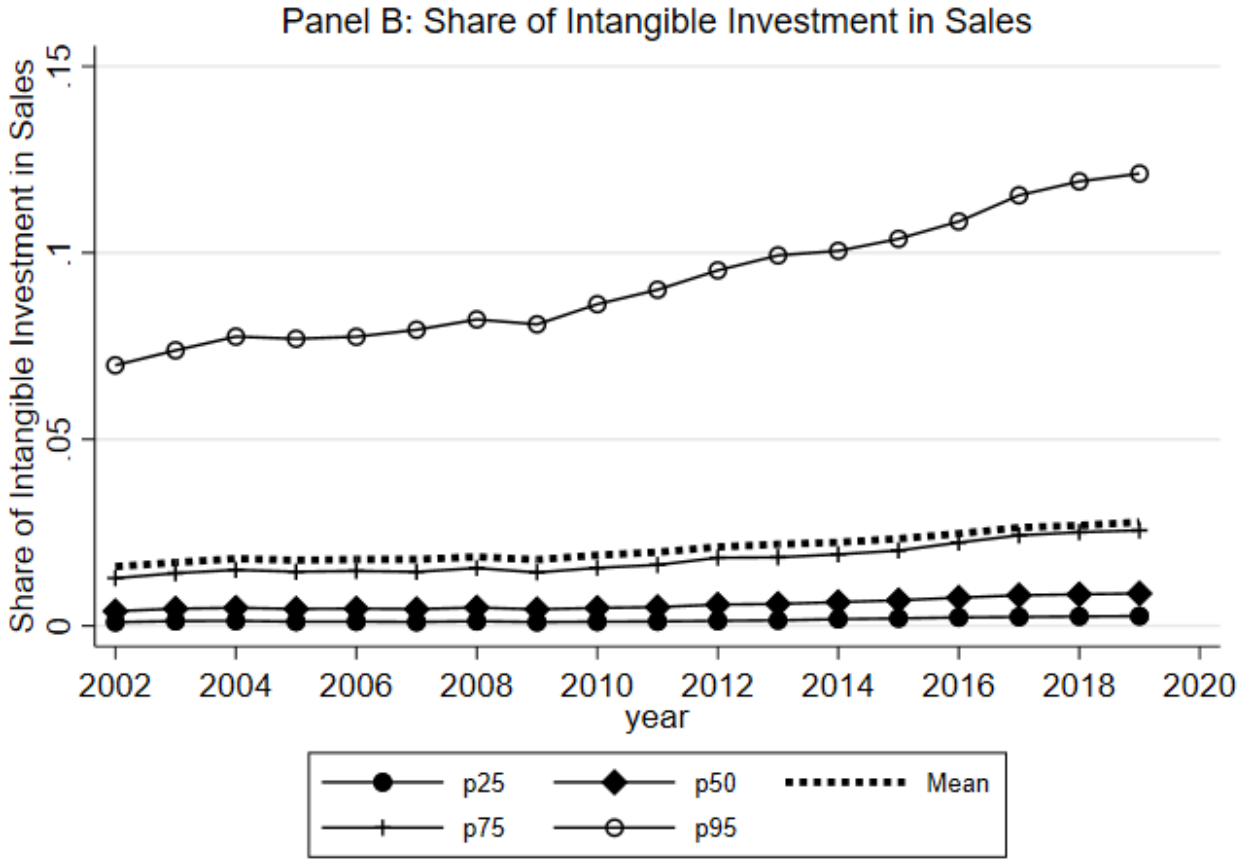




Results

Intangible Capital Deepening





Conclusion

- New data to measure returns to intangibles, allowing to include ***all*** firms and sectors of the private economy
- Findings:
  - Intangibles have become more important over time, mainly due to training/organizational intangibles
  - Bigger firms do most of these intangibles
  - Investing an additional 1€ in intangibles has a gross rate marginal product of €0.8
  - Big firms have excess returns
  - Neglecting intangibles will lead to overestimation of intangibles (intangible capital deepening is labeled as productivity growth)
  - The top 5% firms are driving this intangible capital deepening

Thank you
Q & A

Corresponding authors:
- Jozef Konings (Joep.konings@kuleuven.be)
- Aaron Putseys (aaron.putseys@kuleuven.be)
- Gert Bijnens (gert.bijnens@nbb.be)

Intangible Asset Type	NACE Rev. 2 code	Description
A. Computerized Information:		
(i) Software	6201	Computer Programming Activities
	6209	Other Information Technology and Computer Service Activit
	6312	Web Portals
(ii) Database	6202	Computer Consultancy Activities
	6203	Computer Facilities management Activities
	6311	Data Processing, Hosting and Related Activities
B. Innovative Property:		
(i) R&D (Scientific)	7211	Research and Experimental Development on Natural Sciences and Engineering
	7219	Other Research and Experimental Development on Natural Sciences
	7220	Research and experimental development on social sciences and humanities
(ii) Entertainment & Artistic Originals	7410	Specialized design activities
	7420	Photographic activities
	7430	Translation and interpretation activities
	7911	Travel agency activities
	7912	Tour operator activities
	7990	Other reservation service and related activities
	8810	Social work activities without accommodation for the elderly and disabled
	8891	Child day-care activities

Intangible Asset Type	NACE Rev. 2 code	Description
	8899	Other social work activities without accommodation n.e.c.
	9001	Performing arts
	9002	Support activities to performing arts
	9003	Artistic creation
	9004	Operation of arts facilities
	9311	Operation of sports facilities
	9312	Activities of sport clubs
	9313	Fitness facilities
	9319	Other sports activities
	9321	Activities of amusement parks and theme parks
	9329	Other amusement and recreation activities
	9604	Physical well-being activities
(iii) Design and other new	7111	Architectural Activities
Product/Systems	7112	Engineering Activities and Related Technical Consultancy
	7120	Technical Testing and Analysis
C. Economic Competencies:		
(i) Advertising/Market	7311	Advertising agencies
Research	7312	Media representation
	7320	Market research and public opinion polling
(ii) Employer-provided	7021	Public relations and communication activities
Training/	7022	Business and other management consultancy activities
Organizational	7490	Other professional, scientific and technical activities n.e.c.
Structures	7810	Activities of employment placement agencies
	7830	Other human resources provision

Intangible Asset Type	NACE Rev. 2 code	Description
	8510	Pre-primary education
	8520	Primary education
	8531	General secondary education
	8532	Technical and vocational secondary education
	8541	Post-secondary non-tertiary education
	8542	Tertiary education
	8551	Sports and recreation education
	8552	Cultural education
	8559	Other education n.e.c.
	8560	Educational support activities

Notes: Table A.1 shows the NACE Rev. 2 that produce or sell intangible services. Intangible asset categories and types are based on Corrado et al. (2005, 2009, 2012).

Summary Statistics

	Mean	Median	SD
Measured Value Added (X1,000 €)	877	137	12,077
Corrected Value Added (X1,000 €)	964	147	13,030
Tangible Capital (X1,000 €)	935	129	16,039
Intangible Capital (X1,000 €)	268	14	3,988
Employment	10,55	2	159
Tangible investment (X1,000 €)	146	18	2,717
Intangible Investment (X1,000 €)	88	4	1,331
Sales (X1,000 €)	4,889	538	153,059
Wage Bill (X1,000 €)	530	72	7,222
Measured Intermediate inputs (X1,000 €)	3,557	341	74,598
Corrected Intermediate inputs (X1,000 €)	3,467	328	74,048

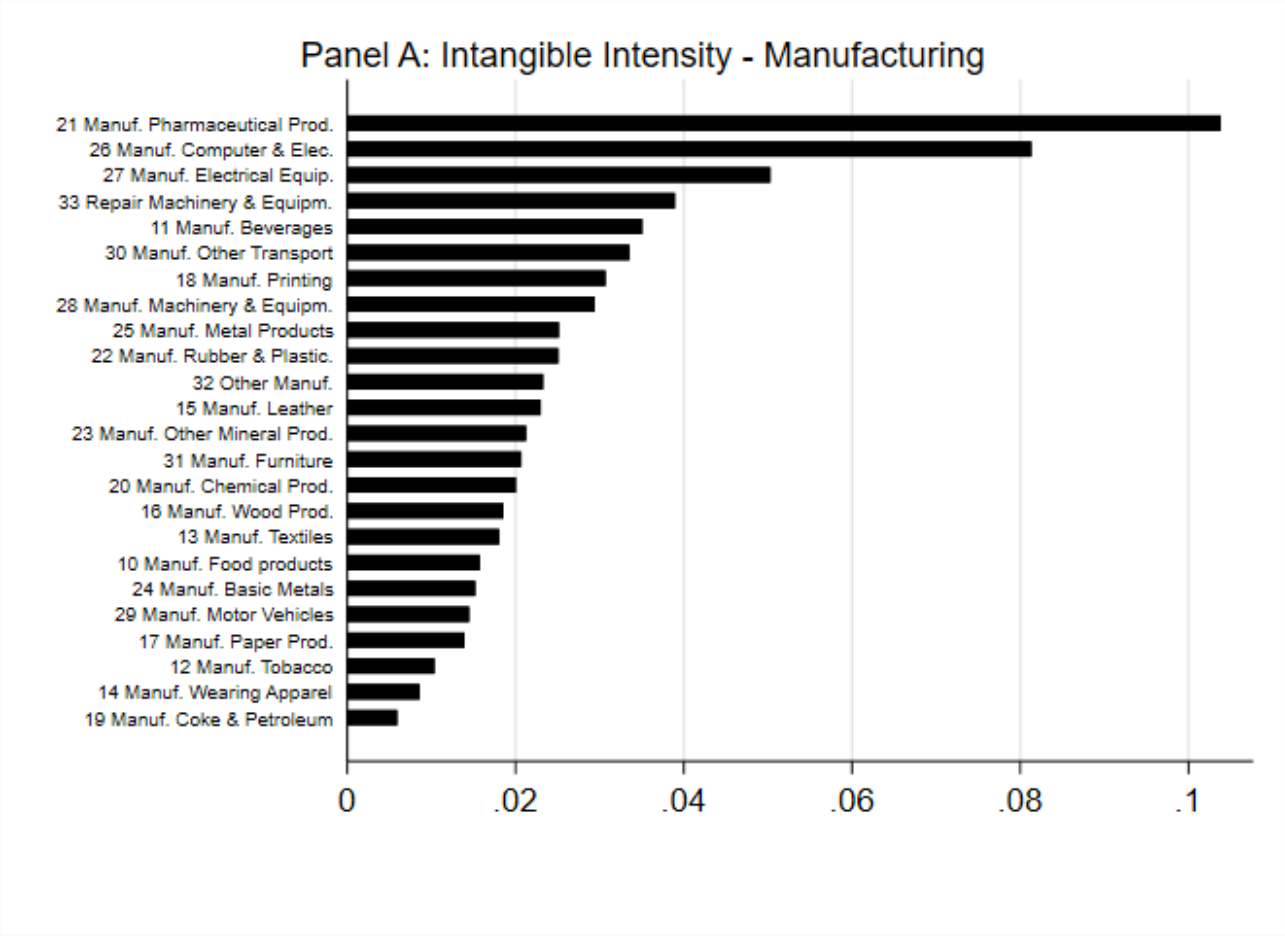
Notes: Table 2 shows the descriptive statistics of the main variables in the analysis. These descriptive statistics are based on our main analysis sample of firms with non-missing values for value added, labor, tangible capital, and intangible capital. Moreover, we do not allow for observations for which expenditures on labor and/or materials are larger than sales (214,304). The average number of observations is 1,638,584. Value added and intermediate inputs are corrected for intangible investments. Descriptive statistics are at the firm level, after taking averages over time per firm. All variables are denoted in 2015 Euros.

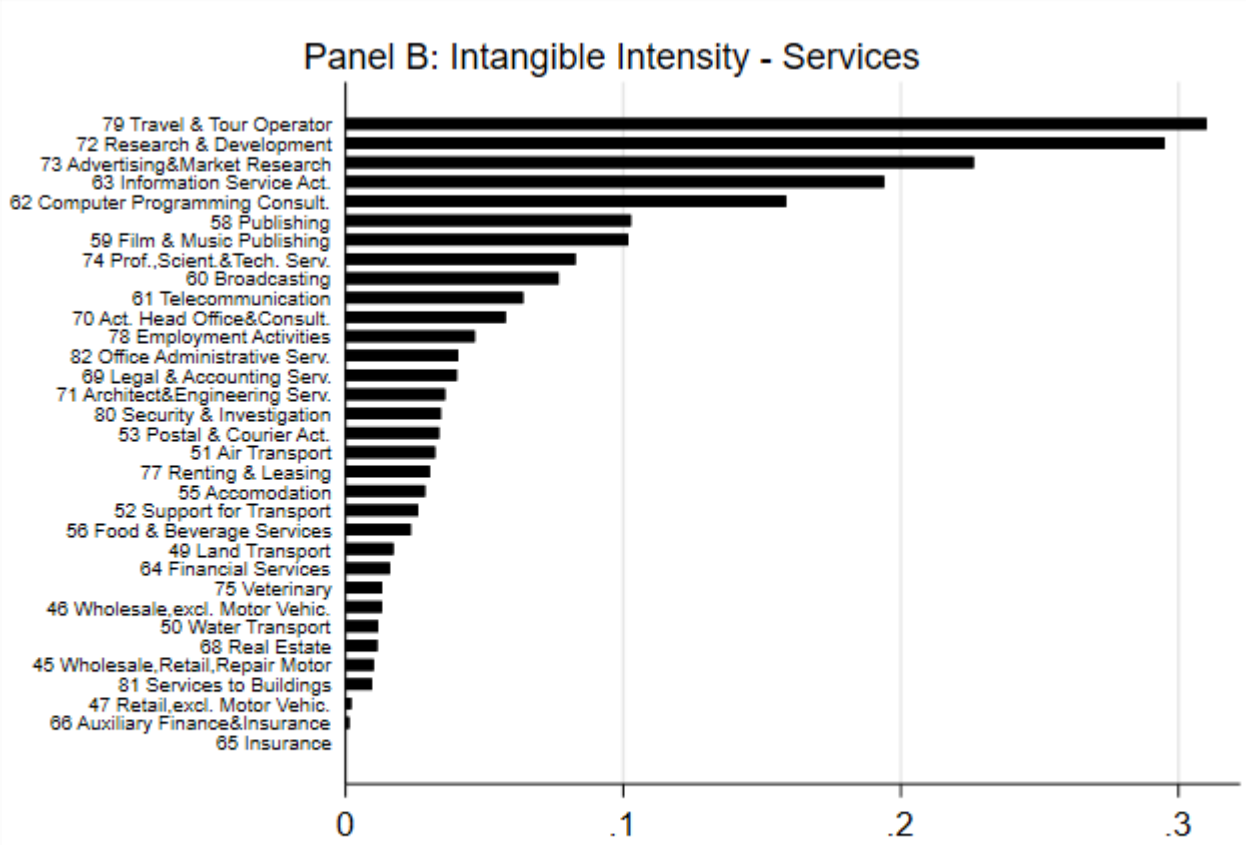
Data: VAT database

Back

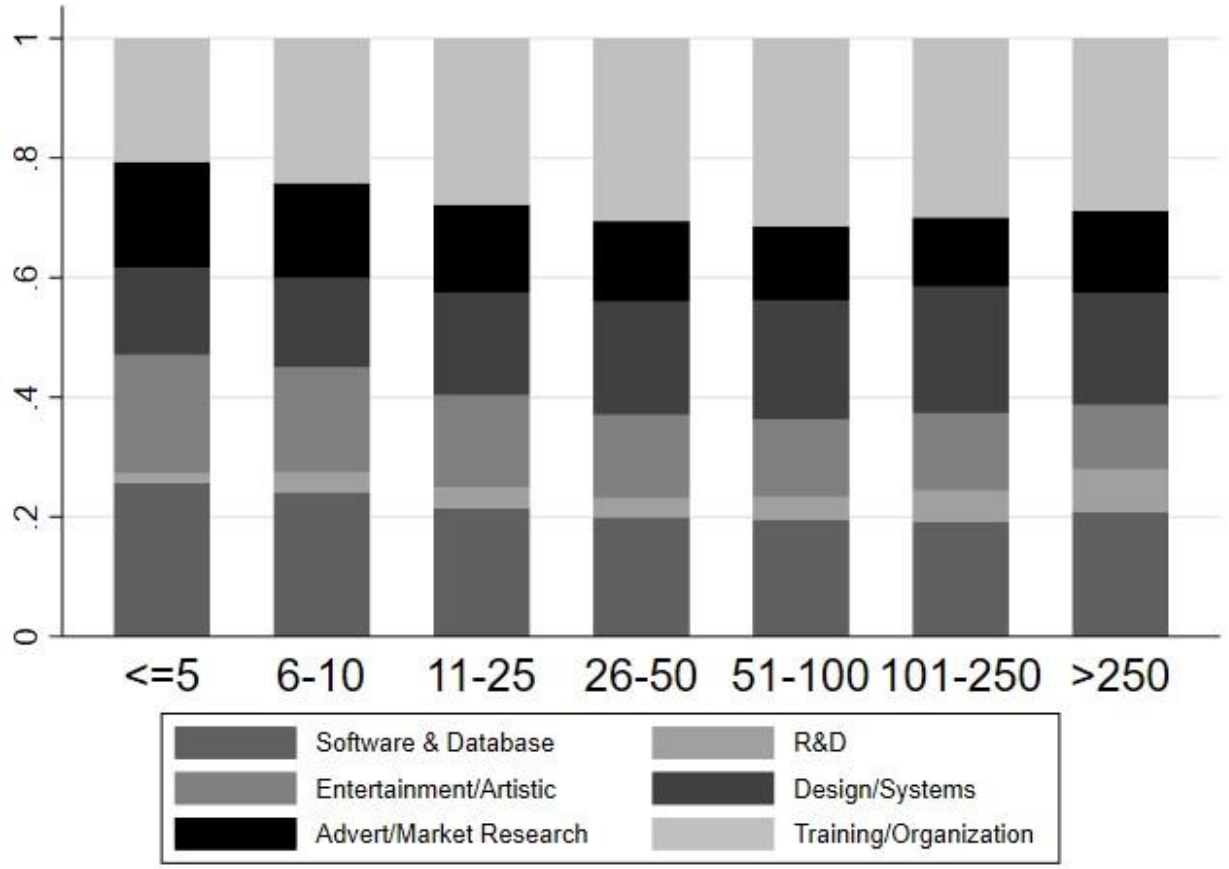
- Database with VAT declarations records transactions between firms
- For each active firm, its sales and identity of each buyer is recorded for the period 2002-2019
- Flip around database to obtain for each firm its domestic suppliers and purchases from each supplier
- Identify intangible purchases based on the detailed activity code of the supplier
- If supplier active in intangible producing sector, purchases are classified as intangible investment

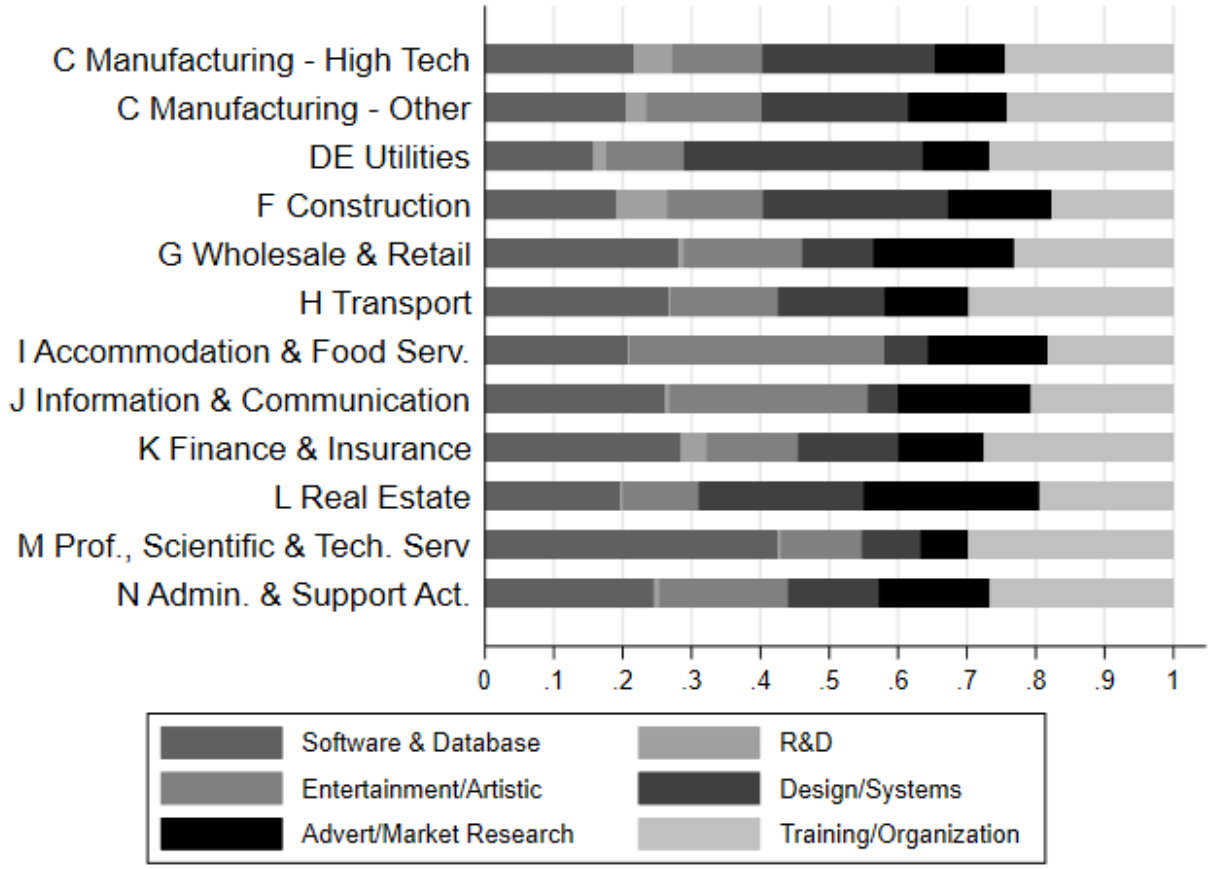
Data: Intangible intensity per industry





Data: Distribution across size and industry





Methodology – ACF (2015)

Back

- Invert material demand function to control for unobserved productivity

$$y_{it} = \beta_l l_{it} + \beta_{IT} k_{it}^T + \beta_{NIT} k_{it}^U + f^{-1}\left(k_{it}^T, k_{it}^U, l_{it}, m_{it}\right) + \epsilon_{it}$$
$$= \tilde{\phi}_t\left(l_{it}, k_{it}^T, k_{it}^U, m_{it}\right) + \epsilon_{it}$$

- Run this regression in a first step to seperate out the pure measurement error
- In a second step, all the input coefficients are identified
- Assumption that productivity follows a first order Markov process

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it}$$

Methodology – ACF (2015)

Back

- Use timing assumptions to identify input coefficients

$$E\left[(\xi_{it})\begin{pmatrix} l_{it} \\ k_{it}^{T} \\ k_{it}^{U} \end{pmatrix}\right] = 0$$

- Practically:
  - Compute for a candidate vector of the input coefficients an estimate for
  - Non-parametrically regress this estimate on its lagged value to recover
  - Bring the sample analogue of the moment conditions as close as possible to 0 by changing the input coefficients

Results

Back

Dependent Variable:	$\ln(Y)$	$\ln(Y)$	$\ln(Y)$	$\ln(Y)$
Estimator:	OLS	OLS	ACF	ACF
Sample:	Manufacturing	Services	Manufacturing	Services
	(1)	(2)	(3)	(4)
$\ln(K)$	0.1695***	0.1498***	0.1422***	0.1159***
	(0.0021)	(0.0010)	(0.0073)	(0.0028)
$\ln(L)$	0.6455***	0.6118***	0.6098***	0.5919***
	(0.0032)	(0.0015)	(0.0094)	(0.0036)
$\ln(U)$	0.1508***	0.1682***	0.1699***	0.1755***
	(0.0011)	(0.0010)	(0.0055)	(0.0036)
$\text{Avg}(MP_K)$	0.1084	0.1049	0.0910	0.0811
$\text{Avg}(MP_L)$	1.0149	1.0710	0.9588	1.0361
$\text{Avg}(MP_U)$	0.7512	0.6945	0.8463	0.7246
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	240,940	1,072,858	206,150	875,782
No. of firms	26,017	145,870	23,271	123,810

Notes: Table 5 shows the production function estimates of Equation 3 estimated using OLS and the control function approach introduced by Ackerberg et al. (2015). Standard errors are clustered at the firm level. All estimations include year fixed effects and two-digit industry fixed effects. $*p<0.10$, $**p<0.05$, $***p<0.01$.

Results

Dependent Variable:	$\ln(Y)$	$\ln(Y)$	$\ln(Y)$	$\ln(Y)$	$\ln(Y)$	$\ln(Y)$
Estimator:	ACF	ACF	ACF	ACF	ACF	ACF
Asset:	Software & Database	R&D	Entertainment & Artistic Orig.	Design & New Products/systems	Advertising & Market Research	Training & Organizational
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(K)$	0.1250***	0.1210***	0.1230***	0.1206***	0.1262***	0.1221***
	(0.0027)	(0.0024)	(0.0026)	(0.0027)	(0.0026)	(0.0025)
$\ln(L)$	0.6236***	0.5982***	0.6170***	0.6137***	0.6205***	0.6145***
	(0.0024)	(0.0025)	(0.0024)	(0.0024)	(0.0025)	(0.0025)
$\ln(U)$	0.0259***	0.0215***	0.0196***	0.0258***	0.0143***	0.0323***
	(0.0003)	(0.0004)	(0.0003)	(0.0004)	(0.0003)	(0.0003)
$\ln(\tilde{U})$	0.1016***	0.1489***	0.1147***	0.1142***	0.1274***	0.1072***
	(0.0014)	(0.0023)	(0.0016)	(0.0019)	(0.0018)	(0.0019)
Avg Intangible						
Cost Share	0.0360	0.0025	0.0456	0.0346	0.0179	0.0541
Avg(MP_U)	0.7238	8.6364	0.4303	0.7456	0.7944	0.5965
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,349,634	1,349,634	1,349,634	1,349,634	1,349,634	1,349,634
No. of firms	183,759	183,759	183,759	183,759	183,759	183,759

Notes: Table 6 shows the production function estimates of a Cobb-Douglas ACF (Ackerberg et al., 2015) production function by intangible capital type. All estimations include year fixed effects and two-digit industry fixed effects. Standard errors are clustered at the firm level. $*p<0.10$, $**p<0.05$, $***p<0.01$.

Results

Back

Dependent Variable:	ln(Y)	ln(Y)	ln(Y)	ln(Y)	ln(Y)	ln(Y)	ln(Y)
Estimator:	ACF	ACF	ACF	ACF	ACF	ACF	ACF
Size Group:	*0-5*	*6-10*	*11-25*	*26-50*	*51-100*	*101-250*	*250+*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(L)	0.3967***	0.7079***	0.7506***	0.7892***	0.7347***	0.7217***	0.6995***
	(0.0030)	(0.0267)	(0.0332)	(0.1237)	(0.2348)	(0.2531)	(0.0697)
Log(K)	0.1466***	0.0714***	0.0600***	0.0545***	0.0574***	0.0673***	0.0904***
	(0.0034)	(0.0049)	(0.0051)	(0.00744)	(0.0103)	(0.0193)	(0.0380)
Log(U)	0.1441***	0.1348***	0.1549***	0.1939***	0.2079***	0.2343***	0.2696***
	(0.0039)	(0.0081)	(0.0096)	(0.0190)	(0.0436)	(0.1079)	(0.0812)
Avg Intangible							
Cost Share	0.2194	0.2131	0.2138	0.2144	0.2148	0.2202	0.2295
Avg(MP_U)	0.7114	0.6812	0.7793	0.9741	1.0470	1.1508	1.2670
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	599,925	175,065	154,512	64,814	26,111	17,646	10,996
No. of firms	118,852	38,655	26,704	10,802	4,478	2,484	1,192

Notes: Table 7 shows the production function estimates of a Cobb-Douglas ACF (Ackerberg et al., 2015) production function. All estimations include year fixed effects and two-digit industry fixed effects. Standard errors are clustered at the firm level. *p<0.10, **p<0.05, ***p<0.01.

Results

	Total	Within-Firm	Reallocation	Entry	Exit
	(1)	(2)	(3)	(4)	(5)
2003	0.0972	-1.0527	0.8726	-0.1844	0.4405
2004	0.3556	-0.1717	0.4181	-0.2182	0.3284
2005	0.1527	0.4861	-0.2140	-0.3789	0.2717
2006	-0.0051	0.5233	-0.1812	-0.5375	0.1979
2007	0.2561	0.5067	-0.1354	-0.3226	0.2181
2008	0.3924	0.6263	-0.1976	-0.2268	0.2052
2009	0.9600	1.2908	-0.2430	-0.3360	0.2766
2010	0.0962	0.4615	-0.2894	-0.2917	0.2255
2011	-0.1026	0.4764	-0.4882	-0.2678	0.1855
2012	0.2763	0.6325	-0.2197	-0.3186	0.1944
2013	0.6984	0.7683	-0.2396	0.0190	0.1673
2014	0.3554	0.9979	-0.1777	-0.3901	-0.0536
2015	0.1991	0.9626	-0.4508	-0.4750	0.1828
2016	0.1880	0.9215	-0.5668	-0.3367	0.1893
2017	0.3512	0.6853	-0.2354	-0.2491	0.1649
2018	0.6351	0.9711	-0.2249	-0.3342	0.2421
2019	0.2670	0.7206	-0.3988	-0.3098	0.2695

Notes: Table 8 shows the results of a decomposition of the change in TFP mismeasurement using the dynamic (Melitz and Polanec, 2015) methodology as described in the text. We divide the change in the overall TFP mismeasurement (column (1)) into four components: column (2) indicates the change in TFP mismeasurement due to a general increase across all surviving firms; column (3) captures reallocation among incumbent (surviving) firms; columns (4) and (5), respectively, indicate the contribution of firm exit and firm entry to the increase in the aggregate TFP mismeasurement.

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Connecting financial data with intellectual property rights: relevancy and (data) complexity[1]

Julie Callaert, Phoenix Chen, Bart Van Looy, Steven Vanhaverbeke, Jan-Bart Vervenne and Jesse Wursten, KU Leuven

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

Connecting financial data with intellectual property rights: relevancy and (data) complexity

B. Van Looy, J-B Vervenne, J. Callaert, J. Wursten, P.-H Chen, S. Vanhaverbeke

KU Leuven, Department of Management, Strategy and Innovation (MSI), Faculty of Business and Economics (FEB), B-3000 Leuven, Belgium

Expert Centre R&D Monitoring (ECOOM)

Abstract

This paper presents the findings of several recent studies that integrate firm-level Intellectual Property Rights (IPR) and financial data, highlighting both advancements and ongoing challenges (Vervenne et al., 2020; BOIP/FOD, 2021). Integrating these data types holds substantial potential for generating valuable insights for policymakers and practitioners. However, aligning these data types at the firm level introduces unique challenges, particularly in harmonization, disambiguation and matching of corporate applicant names obtained from patent and financial databases.

Over recent years, the department of Management, Strategy and Innovation (MSI) and ECOOM at KU Leuven have developed specific algorithms to identify the type of (IP) applicant (Companies, Individuals, HEI's, Governmental agencies) and to harmonize, disambiguate (and regionalize) applicant names[1]. These algorithms have been adopted and used by a number of governmental agencies, both regionally (ECOOM, Flanders) and internationally (Eurostat, EPO, EIB).

In addition, we designed a specific methodology (see figure 1) to match these IP data with financial repositories (Amadeus and Belfirst (BvD/Moody's)). These methodological advancements not only allow for the identification of IP ownership at the firm level and the distillation of relevant indicators (at the firm level and beyond); they also enable analyses that focus on the relation between IP practices of firms (size, scope, nature of IP portfolios) and the financial performance of the implied companies, such as growth and value (e.g. Leten et al. 2007; Belderbos et al. 2014).

Despite these methodological improvements, significant limitations remain. The financial databases that are currently available in the market are limited in terms of comprehensiveness and historical coverage. Additional gaps include critical details such as firm location, ownership history (essential for consolidation), and name changes (critical for accurate longitudinal studies).

The aim of this paper is to demonstrate that while significant strides have been made in harmonizing patent counts and linking IP and financial data at the firm level,

[1] For an overview of relevant methodologies and algorithms, see Magerman et al. (2009), Du Plessis et al. (2009), Callaert et al. (2011)

ongoing data limitations must be addressed to fully realize the potential of these integrated datasets for researchers, practitioners, and policymakers. We strongly advocate for the development of enhanced financial databases that provide accurate longitudinal firm-level data.

Keywords: SME's – patents – EU technology landscape – corporate / IP database linkages

JEL classification: O32 – O34

Contents

Connecting financial data with intellectual property rights: relevancy and (data) complexity 1

Outlining a methodology to match patent assignees with firm level data, including characterizing the size of firms (SMEs or not) 3

Introduction 3

Outline of the methodology 4

Matching 4

Disambiguation 5

Classification 7

Availability of financial and shareholder data 8

Stage 1 of classification: using entity size indicators to differentiate between large and small entities 10

Stage 2 of classification: using ownership information to identify actual SMEs 11

Results: small vs. large corporate entities and SMEs vs. large companies 12

Concluding notes and suggestions for future research 16

References 18

Outlining a methodology to match patent assignees with firm level data, including characterizing the size of firms (SMEs or not)

Introduction

The methodology deployed to arrive at a reliable and comprehensive coupling between corporate patent assignees and financial information pertaining to the assignee firms has been developed over recent years at KU Leuven (ECOOM, MSI). It enables studies on the firm level (e.g. IP and valuation of firms) and it supports further development of indicators for assessing the performance of innovation systems (ECOOM). In the context of the latter application domain, one of the major studies, aimed to identify the contribution of SMEs to technology development (measured by patents) on a European level. This study was commissioned by Eurostat and has recently also been published (Vervenne et al. 2022).

As the referenced study is the most exhaustive and comprehensive one so far – in terms of matching and characterizing firm applicants by means of their financial accounts – we stay close to its findings in the elaboration of the current paper.

The Vervenne e al. study implied an exhaustive methodology, consisting of four steps. First, corporate (patent) applicant names were matched to firm names in financial directories. Next, a disambiguation procedure was applied to identify the right firm when multiple companies were matched to the same corporate applicant. Subsequently, to the extent that entity size indicators and dependency status information was available, a considerable portion of the matched applicants became classified according to firm size. However, a non-negligible portion of firm applicants remained unassigned after the first three automated steps, either because no match was found or because the necessary information for size evaluation was lacking. In a final step, this void was addressed by investigating stratified samples and extrapolating the findings to the population level.

The mere fact that this last step is indispensable to arrive at reliable estimates within reasonable boundaries clearly signals the deficiencies of current financial databases: a number of critical variables are lacking, and the absence of accurate longitudinal data complicates and even prevents a dynamic perspective (spanning decades) all together. As such, both academia and policy makers alike would benefit considerably from the presence of financial databases that are complete and historically correct. As current providers of such databases are clearly not interested in addressing these needs, we plea for an initiative that addresses this void.

Outline of the methodology

Matching

Patent documents report no information on the size of the applicant. Nor do they include a (unique) firm identifier facilitating an unambiguous match with (annual) financial accounts. To classify corporate patent applicants according to size, one has to merge patent and financial data and identify both primarily on the basis of name information. Several commercial databases are available pooling financial data from annual account filings with national business registries. We chose to extract annual account information from Bureau Van Dijk's Amadeus database (2012 edition), which provides data for European firms. Patent data was extracted from the EPO Worldwide Patent Statistical Database or PATSTAT (autumn 2011 edition). The approach adopted to create correspondence between PATSTAT applicants and Amadeus firms is schematized in Figure 1.

In its raw form, the PATSTAT database provides unprocessed (non-harmonised) applicant names, as well as country and address information. Similar to Frietsch et al. (2013), a sector allocation algorithm is applied to all applicant names in PATSTAT to reduce the odds of associating non-corporate applicants with companies. The algorithm uses a keyword logic to filter out non-corporate applicants (see Eurostat, 2011b).[2]

Next, European corporate applicant names in PATSTAT are matched to the names of firms included in the Amadeus database. Before matching, the same name harmonisation procedure reported in Eurostat (2011b) is applied to all company names in Amadeus – the latter with the intent of limiting the number of potential false negatives attributable to the use of different name variants across both databases.[3] In this study, attention is confined to corporate applicants filing for patent protection at the EPO or the USPTO, or relying on the PCT procedure. We considered patent applications filed over a relatively long-time window (1999-2011) to obtain an idea of the proportions prevailing in the long run.

The matching of applicants and firms is based on exact string correspondence, which prioritizes precision of the retained pairs over recall. In the first round of matching, corporate applicants are matched exclusively to companies from the same country (sharing the same country code in both Patstat and Amadeus). Harmonised corporate applicant names are compared consecutively with harmonised versions of current company names, former company names and company aliases. In a second round of matching, the same procedure is repeated whilst relaxing the country condition, assuming that subsidiaries may be established under names that resemble those of

[2] Other types of actor discerned by the sector allocation algorithm include government and non-profit bodies, individuals and universities. Corporates accounted for 66% of the patents filed in the countries in the reference period; individuals 29%; governments and non-profit bodies 3%; and universities 2%.

[3] The name-harmonization methodology removes discrepancies in company names caused, for example, by variation in punctuation, legal form indication abbreviations and spelling, notation of common words, character formatting, and the use of umlauts (Eurostat, 2011b).

the parent companies. Table 1 shows the recall rates of both rounds for the entire time frame (application years 1999-2011) expressed in applicants and applications.

TABLE 1: APPLICANTS AND APPLICATIONS MATCHED TO AT LEAST ONE COMPANY IN THE FINANCIAL DIRECTORY

Country	*Corporate applicants*			*Corporate applications*		
	Total	**Matched**	**%**	**Total**	**Matched**	**%**
EU-27 [4]	104166	64496	61.9	1316568	1094349	83.1
BE	2218	1542	69.5	26129	23220	88.9
BG	107	45	42.1	190	73	38.4
CZ	500	336	67.2	1450	967	66.7
DK	3593	2101	58.5	29487	24468	83.0
DE	30130	16320	54.2	537847	453746	84.4
EE	112	65	58.0	226	136	60.2
IE	1235	912	73.8	8767	6575	75.0
EL	209	59	28.2	676	196	29.0
ES	4234	2395	56.6	17019	11494	67.5
FR	10763	5587	51.9	179457	144112	80.3
IT	13104	8974	68.5	77186	60358	78.2
CY	245	62	25.3	932	323	34.7
LV	74	18	24.3	288	37	12.8
LT	16	8	50.0	27	13	48.1
LU	649	259	39.9	5399	3107	57.5
HU	513	181	35.3	1689	636	37.7
MT	82	53	64.6	426	363	85.2
NL	6891	4720	68.5	132865	121315	91.3
AT	3042	1632	53.6	25293	18588	73.5
PL	401	238	59.4	1179	796	67.5
PT	382	192	50.3	1065	738	69.3
RO	57	17	29.8	95	34	35.8
SI	265	135	50.9	1438	678	47.1
SK	124	76	61.3	305	225	73.8
FI	2683	1724	64.3	51052	44874	87.9
SE	6226	3452	55.4	84844	53081	62.6
UK	16311	13393	82.1	131237	124196	94.6

Source: PATSTAT autumn 2011 edition, Amadeus 2012.

Matching rates (aggregated across patent offices) range between 24.3 % (Latvia) and 82.1 % (United Kingdom). Overall, 61.9 % of the harmonised corporate applicant names are matched to the companies reported in Amadeus (57.9 % in the same country and 4.0 % in other Member States). These matched corporate applicants account for 83.1 % of patent applications filed by corporate applicants. 77.9 % can be assigned to corporate applicants matched to companies from the same country and 5.2 % to those matched to companies from other Member States. The comparison between applicant and application figures reveals that, on average, unmatched corporate applicants patent less than matched ones (38.1% of the unmatched names account for 16.9 % of the remaining corporate patent volume).

Disambiguation

[4] At the time of the study, the EU comprised 27 Member States.

The name-based matching algorithm resulted in corporate applicants being matched to multiple companies present in Amadeus: approximately one out of every eight harmonized corporate applicant names was linked to more than one of the legal entities included in Amadeus. The statistics reported in Table 2 reveal large cross-country differences in the share of matched corporate applicants allocated to multiple firms.

TABLE 2: SHARES OF MATCHED CORPORATE APPLICANTS MATCHED TO MULTIPLE FIRMS AND AVERAGE NUMBER OF AMADEUS FIRMS MATCHED TO THEM.

Country	% applicants matched to multiple firms	Average number of Amadeus firms linked to corporate applicants matched to more than 1 firm
IT	35%	12.3
FR	21%	7.1
RO	15%	2.0
PL	15%	3.7
DE	13%	4.4
BE	11%	3.7
AT	9%	2.4
NL	7%	3.7
CZ	4%	2.5
SI	4%	2.0
LV	3%	2.0
GR	3%	2.0
SK	2%	2.0
PT	2%	4.0
GB	2%	2.1
DK	2%	2.4
SE	2%	5.0
FI	1%	2.1
IE	0%	2.0
ES	0%	2.0
BG	0%	-
CY	0%	-
EE	0%	-
HU	0%	-
LT	0%	-
LU	0%	-
MT	0%	-

These cross-country differences can, among other factors, be attributed to the lack of strict name requirements prior to business registration in some countries (e.g. Italy) or the non-binding nature of such requirements in other countries (e.g. France) (World Bank, 2017). A number of selection rules are applied to disambiguate these associations. The full disambiguation process consists of four automated rounds applied consecutively. First, companies with addresses that do not correspond to the address of the corporate applicant are removed if at least one other company is matched to the corresponding address information. Next, when one of the matched companies holds the majority of shares in another company matched to the same corporate applicant, the latter is discarded. Thus, priority is given to companies at the top of the shareholder hierarchy. Subsequently, any liquidated, dissolved, bankrupt or inactive company matched to a corporate applicant that has filed patents after the year in which it published its last available annual accounts is also discarded. Finally, in line with Squicciarini and Dernis (2012), in the final disambiguation round, matched

companies showing maximum values for revenue, staff count and total assets – in that order – were retained. For the few corporate applicants that remained ambiguously matched after subjection to this decision rule, the most adequate match was then selected by means of human assessment until every matched corporate applicant was paired with a single firm. Table 3 reports the effects of each disambiguation step on the number of firms remaining matched to the corporate applicant in Amadeus. Clearly the first and last disambiguation rounds reduce the number of potentially corresponding firms the most extensively.

TABLE 3: IMPACT OF DISAMBIGUATION STEPS ON NUMBER OF AMADEUS FIRMS LINKED WITH THE MATCHED CORPORATE APPLICANTS

# firms matched to applt	*Initial Matches*	*Address*	*Shareholders*	*Inactivity vs 1st filing*	*Financial/manual*
1	55,397	58,943	59,006	59,249	64,496
2	3,963	2,706	2,652	2,480	-
3	1,306	800	795	774	-
4	714	420	418	402	-
5	496	304	303	293	-
6	342	206	205	204	-
7	281	157	157	151	-
8	198	107	107	107	-
9	180	91	91	86	-
10	162	84	84	87	-
11-15	493	255	255	253	-
16-20	251	130	130	122	-
21-30	298	139	139	140	-
31-50	234	94	94	90	-
51-100	153	51	51	49	-
101-200	29	9	9	9	-
201-250	2	0	0	0	-
	64,496	64,496	64,496	64,496	64,496

Classification

The classification of matched corporate applicants according to firm size is based on the European Commission's (EC) SME definition, which sets out three criteria for SME status: a staff headcount below 250 full-time equivalents (FTEs) and either an annual turnover below or equal to 50 million Euro or a balance sheet total below or equal to 43 million Euro.[5] Preferably, these criteria should be evaluated at the level of the business group to which the entity belongs. Entities belong to a group when a natural person controlling multiple firms or a corporate shareholder owns a majority of the voting rights in the entity's board. Ownership of at least 50% of the shares – be it directly, or indirectly through an intermediary subsidiary – signals such a controlling position.

[5] On the basis of European Commission Recommendation 2003/361/EC, a new SME definition was adopted on 1 January 2005, incorporating updated thresholds for companies applying for the European support programme for SMEs.

Applying these financial and ownership criteria to a financial database is not a straightforward exercise. Due to disclosure exemptions for smaller firms and the scale of the targeted firm population, the data provided for an average firm is less complete than the standard witnessed in databases that cover only large and/or listed firms. As the next section makes clear, Bureau Van Dijk's Amadeus 2012 database is no exception in this respect.

Availability of financial and shareholder data

Bureau Van Dijk publishes a new version of Amadeus every year since it became a commercial product in 1996. Each version is updated regularly throughout the year, incorporating newly available information. In this study, we based the firm size assessment of patenting companies on the most recent annual accounts available per firm for two principal reasons. While time series data for the past 10 years are available for revenues, staff counts and total assets at the entity level, only the last known ownership information is reported. This hampers a dynamic assessment of firm size at the business group level.[6] In addition, rather than providing a historical account of the firm's shareholder structure, the majority of supplementary information sources (cf. infra) that were consulted to ascertain the size distribution of the unallocated applicants only provide an insight into their last known ownership situation. Therefore, rather than classifying the patent based on financial information derived from the annual accounts covering the year in which it was applied for by the applicant, each corporate applicant and all its patents are attributed to one size category for the entire time frame. By exclusively relying on the last available size indicators per applicant for classification, an additional source of potential bias is introduced to the SME contribution estimates. To the extent that small firms grow fast, the number of SME patents could well be underestimated during the initial years of the time period under study.

To evaluate the extent to which the EC's SME definition can be used to differentiate between SMEs and large companies in Amadeus 2012, we assessed per-country data availability for the indicators of interest. Table 4 reports coverage rates per indicator in Amadeus 2012.

[6] A combination of ownership information from multiple Amadeus versions at fixed intervals within the investigated time window could provide a solution in which a more precise picture of the size of the matched corporate applicants at the precise time of filing can be obtained. However, on top of the other arguments mentioned, the even more fragmented nature of ownership information in pre-2012 versions of Amadeus exercised an equal influence on our decision to refrain from such efforts.

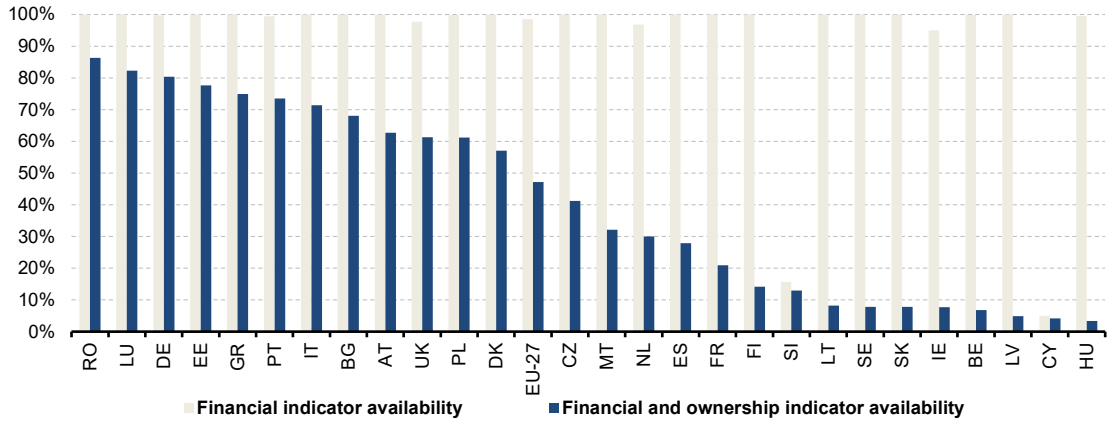
TABLE 4: AMADEUS – OVERALL COVERAGE OF FINANCIAL AND OWNERSHIP INDICATORS REQUIRED TO DETERMINE FIRM SIZE (PER INDICATOR)

	Total number of companies	*Companies reporting operational revenues*		*Companies reporting staff count*		*Companies reporting total assets*		*Companies reporting dependency status*	
		number	**%**	**number**	**%**	**number**	**%**	**number**	**%**
BE	609412	129201	21.20	257647	42.28	545187	89.46	41646	6.83
BG	494532	59231	11.98	488209	98.72	60490	12.23	336347	68.01
CZ	482679	469799	97.33	294733	61.06	184120	38.15	198742	41.17
DK	25230	42447	16.78	76887	30.40	252918	100.00	144339	57.07
DE	1456074	419446	28.81	437928	30.08	1101434	75.64	1170990	80.42
EE	108986	94667	86.86	52709	48.36	108936	99.95	84675	77.69
IE	211372	25951	12.28	21836	10.33	199798	94.52	16627	7.87
EL	28401	28401	100.00	23600	83.10	28401	100.00	21289	74.96
ES	1273351	1140063	89.53	843380	66.23	1273351	100.00	355419	27.91
FR	1291883	1291875	100.00	878954	68.04	1291882	100.00	269729	20.88
IT	119884	1188353	99.14	352781	29.43	1198684	100.00	856313	71.44
CY	41289	907	2.20	1878	4.55	1005	2.43	36963	89.52
LV	110292	85711	77.71	102382	92.83	7938	7.20	5406	4.90
LT	117370	26789	22.82	110033	93.75	3712	3.16	9636	8.21
LU	19240	5199	27.02	4040	21.00	16028	83.31	15838	82.32
HU	377912	316267	83.69	135538	35.86	375792	99.44	12683	3.36
MT	15259	15259	100.00	392	2.57	15259	100.00	4907	32.16
NL	895494	31108	3.47	643147	71.82	822005	91.79	272414	30.42
AT	224480	6385	2.84	167413	74.58	77880	34.69	140733	62.69
PL	960971	117796	12.26	902969	93.96	121316	12.62	589289	61.32
PT	434526	365782	84.18	343776	79.12	428069	98.51	319629	73.56
RO	571289	568039	99.43	566221	99.11	571038	99.96	493193	86.33
SI	76089	2512	3.30	10558	13.88	2574	3.38	15661	20.58
SK	230781	165399	71.67	197956	85.78	58557	25.37	17996	7.80
FI	170484	160798	94.32	47776	28.02	170484	100.00	24134	14.16
SE	866641	829664	95.73	848854	97.95	319732	36.89	67869	7.83
UK	3076136	447311	14.54	132348	4.30	2998120	97.46	1918280	62.36
Total	**15596557**	**8034360**	**51.51**	**7943945**	**50.93**	**12234710**	**78.44**	**7440747**	**47.71**

Source: PATSTAT autumn 2011 edition, Amadeus 2012.

The statistics presented in Table 4 reveal non-trivial data gaps, with considerable variation across Member States. This is partly caused by differences in legislation regarding registration/disclosure, whereby smaller entities are not bound to full disclosure or are exempted from financial information disclosure altogether.

FIGURE 2: AMADEUS – OVERALL COVERAGE OF (FINANCIAL AND FINANCIAL/OWNERSHIP) INDICATORS REQUIRED TO DETERMINE FIRM SIZE (%)



Source: Amadeus 2012

Figure 2 provides further insight into the availability of firm size indicators in Amadeus 2012. The Member States are ranked according to descending rates of joint availability of financial and ownership indicators. Overall, 99% of the firms in Amadeus 2012 report at least one of the three financial size indicators from which entity size can be derived. Cyprus and Slovenia appear to be the only 'problematic' cases in terms of coverage. The figures become more troublesome if one takes ownership information availability into account as well for evaluation of size at the enterprise or business group level. Only 47% of the firms provide at least one of the financial firm size indicators and offer sufficient shareholder information to distinguish dependent from independent firms.

Taking into account these data constraints, classification of the matched corporate applicants comprised two stages. In the first stage, entity-size indicators (revenues, employee count and total assets) are used to identify large and small corporate entities. Obviously, for large entities, one can already state that they represent a large enterprise. In the second stage, to filter out the *actual* SMEs among the small corporate entities, dependency information is introduced and consolidated financial size indicators are used in order to classify firms in terms of majority ownership.

Stage 1 of classification: using entity size indicators to differentiate between large and small entities

Entities with known information on all three entity size indicators allow for classification of the corporate applicants associated with them. taking into

consideration all criteria comprising the SME definition. The corporate applicants that are categorized using this baseline SME definition are labelled as entities that are 'certainly large' or 'certainly small'. However, the coverage gaps reported for the three indicators in Table 4 underscore the need for a complementary, looser version of the definition. To the extent that the thresholds for the three size indicators are determined according to proportions that are representative for the average industry, it seems plausible to presume that, if for instance one available indicator lies above the threshold, the other indicators are more likely to lie above than below their thresholds as well. Companies that are classified using this approach are labelled as entities that are 'most likely large'[7] or 'most likely small'[8].

Stage 2 of classification: using ownership information to identify actual SMEs

The shareholder information provided in Amadeus is essential to determine which small entities effectively qualify as independent SMEs and which are members of larger business groups or are backed by other types of shareholders such as governments, institutional investors and universities. Determining which of the matched companies required further exploration of their shareholder structure departed from the 'independence indicator' reported in Amadeus. Bureau Van Dijk designed this measure to facilitate quick evaluation of how dependent a company is vis-à-vis its shareholders. We rely on it to differentiate between independent SMEs, i.e. small businesses that have no majority shareholders, and a 'rest' category of non-independent small businesses. Further analysis of shareholder data is restricted to this 'rest' category of firms. More specifically, to the extent such data is reported, information about the nature of those firms' shareholders and the size of shareholder stakes in those and other firms (in the case of a natural person holding the majority of the shares) is designed to assess the size of the business group to which the entity potentially belongs. Insofar as available, this assessment is based on the sum of direct and indirect ownership percentages i.e. the percentage of the company directly owned by the shareholder and the percentage owned through subsidiaries.[9] The size

[7] The entities considered as 'most likely large' were those reporting: revenues over 50 million Euro, where staff count and total asset numbers are unavailable; total assets over 43 million Euro, where staff count and revenue numbers are unavailable; revenues over 50 million Euro and total assets of 43 million Euro or less, where staff count numbers are unavailable; revenues of 50 million Euro or less and total assets over 43 million Euro, where staff count numbers are unavailable.

[8] Companies that are 'most likely' small entities are those reporting: a staff count of less than 250 Full-Time Equivalents (FTE), where revenues and total assets are unavailable; revenues of EUR 50 million or less and total assets of 43 million Euro or less, where staff count numbers are unavailable; total assets of 43 million Euro or less, where staff count and revenue numbers are unavailable; revenues of 50 million Euro or less, where staff count and total assets numbers are unavailable; revenues of 50 million Euro or less and total assets of 43 million Euro or less, where staff count numbers are unavailable; revenues of over 50 million Euro and staff counts of less than 250 FTEs, where total asset numbers are unavailable; and total assets of over 50 million Euro and staff counts of less than 250 FTEs, where revenue numbers are unavailable.

[9] Where the ultimate owner controls the intermediate subsidiary by owning a majority of its shares, one can assume that its ownership stake in the company under consideration is a reflection of the share percentage its subsidiary holds in that company. Otherwise, Bureau Van Dijk calculates the indirect

of corporate applicants that are accordingly held to form part of a business group is evaluated at the group level (see below).

Results: small vs. large corporate entities and SMEs vs. large companies

The outcomes show that 8.9 % qualify as large corporate entities in the first round, whereas 52.5 % can be characterised as small corporate entities. A majority of the large entities are identified as 'certainly' large (8.5 %) and the remaining 0.4 % as 'most likely' large corporate entities (see above). 27.2 % of the small entities are characterised as 'certainly' small and 25.3 % as 'most likely' small (see above). The corresponding patent volumes are presented in Table 5, the first row of which contains the total number of distinct patents per size category for the EU overall.

TABLE 5: DISTRIBUTION OF CORPORATE PATENTS AFTER THE 1ST STAGE OF CLASSIFICATION BASED ON ENTITY-LEVEL SIZE INDICATORS ONLY

Country	*Large*				*Small*				*No financial size indicators*		*Not matched*		*Total*
	Certainly		Most likely		Certainly		Most likely						
	#	%	#	%	#	%	#	%	#	%	#	%	#
EU-27	696 716	52.9	25 896	2.0	174 998	13.3	192 480	14.6	4 259	0.3	222 219	16.9	1 316 568
BE	15 103	57.8	1 633	6.2	5 301	20.3	1 182	4.5	1	0	2 909	11.1	26 129
BG	37	19.5		0	19	10	17	8.9		0	117	61.6	190
CZ	362	25	5	0.3	518	35.7	82	5.7		0	483	33.3	1 450
DK	15 229	51.6	94	0.3	5 821	19.7	3 292	11.2	32	0.1	5 019	17	29 487
DE	293 215	54.5	7 412	1.4	35 956	6.7	116 945	21.7	218	0	84 101	15.6	537 847
EE	6	2.7	2	0.9	92	40.7	36	15.9		0	90	39.8	226
IE	1 225	14	315	3.6	1 472	16.8	3 434	39.2	129	1.5	2 192	25	8 767
EL	29	4.3		0	98	14.5	64	9.5	5	0.7	480	71	676
ES	4 838	28.4	18	0.1	5 900	34.7	738	4.3		0	5 525	32.5	17 019
FR	107 682	60	4 473	2.5	26 879	15	5 031	2.8	47	0	35 345	19.7	179 457
IT	30 752	39.8	343	0.4	23 066	29.9	6 192	8	5	0	16 828	21.8	77 186
CY	6	0.6	4	0.4	66	7.1	45	4.8	202	21.7	609	65.3	932
LV	9	3.1		0	18	6.3	10	3.5		0	251	87.2	288
LT		0		0	6	22.2	7	25.9		0	14	51.9	27
LU	1 199	22.2	103	1.9	249	4.6	1 556	28.8		0	2 292	42.5	5 399
HU	310	18.4		0	229	13.6	95	5.6	2	0.1	1 053	62.3	1 689
MT	56	13.1		0	18	4.2	289	67.8		0	63	14.8	426
NL	89 064	67	7 173	5.4	16 027	12.1	7 474	5.6	1 577	1.2	11 550	8.7	132 865
AT	11 072	43.8	26	0.1	1 447	5.7	6 042	23.9	1	0	6 705	26.5	25 293
PL	271	23	17	1.4	194	16.5	314	26.6		0	383	32.5	1 179
PT	223	20.9	9	0.8	434	40.8	72	6.8		0	327	30.7	1 065
RO	5	5.3		0	29	30.5		0		0	61	64.2	95
SI	116	8.1		0	35	2.4	102	7.1	425	29.6	760	52.9	1 438
SK	59	19.3		0	132	43.3	34	11.1		0	80	26.2	305
FI	35 366	69.3	1 756	3.4	4 260	8.3	3 452	6.8	40	0.1	6 178	12.1	51 052
SE	34 513	40.7	356	0.4	16 987	20	1 193	1.4	32	0	31 763	37.4	84 844
UK	55 969	42.6	2 157	1.6	29 745	22.7	34 782	26.5	1 543	1.2	7 041	5.4	131 237

Source: PATSTAT autumn 2011 edition, Amadeus 2012.

percentage by multiplying the ultimate owner's direct share in the intermediate subsidiary by the direct share that the subsidiary holds in the company under consideration. More specific information on Bureau Van Dijk's procedures for calculating total ownership percentages can be found in the Bureau Van Dijk Amadeus Ownership Guide (2008).

In the second stage of the classification process, shareholder information for the small entities is taken into consideration to identify actual SME activity. Further evaluation of shareholder information is deemed irrelevant for companies already identified as large corporate entities. Equally, for the small entities that are independent according to Bureau Van Dijk's 'independence indicator', shareholder data is not subjected to any further inquiry. The 'rest' category of non-independent small entities either qualifies as dependent (26.4 % of corporate applicants) or reports an unknown degree of dependence (15.0 %) according to the independence indicator.

In line with EU directives, the financials of companies that belong to multi-entity business groups are assessed on a consolidated basis. Companies in Amadeus can be members of two types of business group. Companies controlled by another company might be labelled as a more formal type of business group, while those controlled by a natural person who also holds majority stakes in other companies represent a more informal equivalent.

Of the 43,631 small entities that are non-independent, 16,755 have a corporate organisation[10] as majority shareholder (holding 50+ % of the shares). 13,920 of these are 'non-financial' companies – with 'financial institutions' or plain investment vehicles representing the counterpart – established in Europe. They belong to the population of companies covered by Amadeus and more detailed financial information is hence available.[11] For 5,030 of these 13,920 EU majority shareholders, this more detailed information indicates that the available size indicators pertain to the consolidated group rather than the entity and, as such, they can be assessed in isolation to determine business group size. Accordingly, we can determine that, among all corporate applicants, 1.9 % (accounting for 1.1 % of corporate applications) belong to a small business group and 2.8 % (3.9 % of corporate applications) to a large business group.

For the non-independent small entities controlled by companies only reported in Amadeus on an unconsolidated basis or from numbers for which the reporting basis is unknown (among others, the non-European corporate majority shareholders that are only reported in Amadeus as 'shareholder'), the consolidation methodology proposed by the European Commission's (2003) is applied.[12] Consolidated group-level figures are approximated by adding the revenue, staff count and total asset figures of the majority shareholder(s) with their equivalent for the non-independent small entity insofar as they are reported for both. On the basis of this

[10] That is, industrial companies, holding companies and private equity firms. Majority shareholders in the form of institutional investors such as pension and mutual funds/trusts, banks and insurance companies are treated separately.

[11] For non-European shareholders, the only relevant information provided by Amadeus in the shareholder section of the firm in which they hold shares is shareholder turnover, shareholder total assets and shareholder staff count.

[12] Given that the large majority of the parent company financials with a known reporting basis are unconsolidated, we assumed that this would also be the case for the available parent company financials with an unknown reporting basis.

cruder consolidation strategy, an additional 2.5 % of corporate applicants (2.4 % of corporate applications) can be assumed to belong to a large business group.

A similar approach was taken for companies reporting natural persons as majority shareholders. If Amadeus registered these natural persons as holding majority stakes in other companies, the financials across all majority-owned companies were aggregated and compared with large-company thresholds.[13] On the basis of this cruder consolidation strategy, the involvement of shareholder information in size assessment resulted in an additional 0.1 % of corporate applicants (0.1 % of corporate applications) being categorised as member of a large enterprise.[14]

Aggregating the results from all consolidation approaches, in the final analysis, 5.4 % from the 41.9 % non-independent small entities can be reclassified as large companies. This represents 6.4 % of the matched patent volume. The 1.9 % of non-independent small entities assigned to the small business group category account for 1.1 % of the matched patent volume. Obviously, both numbers can only be taken as lower bound estimates of the shares of non-independent small entities that can be reallocated to the large and small company categories respectively, since the Amadeus data covering the ownership structure and shareholder financials is highly incomplete.

Table 6 summarizes the outcomes by mapping the 10 categories of corporate applicants to four aggregate classes: 'large' companies, actual SMEs, companies for which insufficient information is available to classify them as either large or small, and corporate applicants not matched to any company in the Amadeus directory. Multiple categories are linked to the first three classes. The small entities that belong to large business groups and large entities are unquestionably 'large' companies. The small entities backed by (semi-)public actors or institutional investors are excluded from the category of actual SMEs and assigned to the 'large' company category. Questions can be raised about the decision to allocate both to the large business class: access to the resources of their majority shareholder may be less obvious than is the case for other types of majority shareholder. However, it should be noted that, given the low numbers that are reported for both groups, the effects of this decision on the final estimates are limited. The class of 'actual SMEs' consists of independent SMEs and SMEs linked to small business groups. The class of companies for which information is insufficient for reliable firm size determination comprises companies for which ownership (and financial) information is lacking but also small entities

[13] Note that, unlike the case of corporate shareholders, no unique identifier is provided in Amadeus 2012 for natural person shareholders. Therefore, the consolidation of informal business group numbers was preceded by a matching of natural person majority shareholders based on name and address information.

[14] The aggregate size indicators resulting from this cruder consolidation approach should be regarded as estimates of the minimal enterprise size because the consolidation is most likely to remain incomplete. Accordingly, these figures are only used to recognize the entities pertaining to business groups that are at least larger than the SME threshold.

controlled by natural persons. The incompleteness of Amadeus and the possibility that those natural persons hold majority stakes in other companies prevent us from directly classifying these applicants as actual SMEs. The production of reliable estimates of the SME contribution for all corporate applicants, including the non-matched ones and those matched to entities not reporting sufficient size information, requires additional efforts in the form of an extrapolation step.

TABLE 6: SHARES OF PATENTS ATTRIBUTABLE TO ACTUAL SMES AND LARGE FIRMS AFTER THE CLASSIFICATION STAGE (IN %).

Country	*'Large' company*	*SME*	*Matched but unknown*	*Not matched*	*Total*
EU-27	**61.7**	**5.4**	**16.0**	**16.9**	**100.0**
BE	71.4	4.6	12.9	11.1	100.0
BG	21.6	3.2	13.7	61.6	100.0
CZ	29.2	8.4	29.1	33.3	100.0
DK	56.5	13.7	12.7	17.0	100.0
DE	63.8	3.2	17.5	15.6	100.0
EE	8.0	24.3	27.9	39.8	100.0
IE	29.5	3.8	41.7	25.0	100.0
EL	5.3	6.1	17.6	71.0	100.0
ES	35.7	7.1	24.7	32.5	100.0
FR	67.7	3.1	9.5	19.7	100.0
IT	44.8	14.0	19.4	21.8	100.0
CY	3.6	1.0	30.0	65.3	100.0
LV	4.2	2.4	6.3	87.2	100.0
LT	7.4	7.4	33.3	51.9	100.0
LU	25.5	2.2	29.8	42.5	100.0
HU	18.4	0.9	18.4	62.3	100.0
MT	14.1	8.5	62.7	14.8	100.0
NL	77.9	2.0	11.4	8.7	100.0
AT	50.9	6.5	16.2	26.5	100.0
PL	32.5	18.1	17.0	32.5	100.0
PT	28.4	20.9	20.0	30.7	100.0
RO	7.4	1.1	27.4	64.2	100.0
SI	8.8	2.6	35.7	52.9	100.0
SK	25.2	8.2	40.3	26.2	100.0
FI	74.5	3.2	10.2	12.1	100.0
SE	45.0	6.1	11.5	37.4	100.0
UK	56.3	14.9	23.4	5.4	100.0

Source: PATSTAT autumn 2011 edition, Amadeus 2012.

It goes without saying that the levels of unknown cases (SME or not) in the previous table do not allow to arrive at a precise estimate of the share of SMEs in patenting across European countries. This only becomes feasible after an additional phase (based on exhaustive sampling) in which extrapolation (with a 95% confidence level) becomes feasible (see table 7). Such additional, time/resource consuming efforts stem directly from the incompleteness of current financial databases.

TABLE 7: SHARES OF PATENTS ATTRIBUTABLE TO ACTUAL SMES AND LARGE FIRMS AFTER THE EXTRAPOLATION STAGE (IN %)

Country	*Large (%)*	*Unknown - non-matched (%)*	*Unknown - matched (%)*	*SME (%)*
EU-27	**78.9**	**2.3**	**1.2**	**17.6**
BE	79.2	0.0	2.6	18.2
BG	36.8	9.5	0.0	53.8
CZ	60.1	2.8	0.0	37.1
DK	67.2	3.7	1.5	27.6
DE	84.9	2.8	2.0	10.3
EE	19.9	0.0	2.3	77.8
IE	50.4	2.6	2.8	44.1
EL	46.1	14.4	0.0	39.6
ES	61.3	3.8	0.0	34.8
FR	83.4	2.4	0.1	14.1
IT	60.8	2.0	0.2	37.1
CY	28.3	9.0	0.0	62.7
LV	33.7	9.5	0.0	56.8
LT	50.5	0.0	0.0	49.5
LU	49.4	11.5	0.0	39.1
HU	59.3	3.8	0.0	37.0
MT	23.4	2.3	0.0	74.3
NL	83.8	0.9	0.7	14.6
AT	77.2	1.9	0.0	20.9
PL	62.0	0.0	4.0	34.0
PT	42.7	6.0	2.8	48.5
RO	46.9	0.0	5.6	47.5
SI	62.8	3.0	0.0	34.2
SK	43.7	5.3	0.0	51.0
FI	83.6	2.2	0.9	13.2
SE	78.8	2.2	0.2	18.9
UK	62.1	1.4	1.2	35.3

Source: PATSTAT autumn 2011 edition, Amadeus 2012, internet searches based on applicant name.

Table 7 reveals that 3.5 % of patent volume remains assigned to corporate applicants of unknown size. 17.6 % of the applications originate from innovative activities in SMEs, whereas 78.9 % are filed by large companies. More established knowledge economies such as Germany, France, The Netherlands, Sweden, Belgium, Austria, and Finland seem to show lower proportions of patents filed by SMEs than in the EU overall. However, exceptions to this observed tendency are noted as well: equivalent to some of the more peripheral and more recent Member States, advanced economies such as Denmark, and especially the United Kingdom, show SME contributions above the overall European level.

Concluding notes and suggestions for future research

We report on a previous effort to create a comprehensive picture of the SME contribution in corporate patenting in Europe (Vervenne et al., 2021). To achieve this, we assessed the feasibility of constructing a European-wide indicator reflecting the contribution of SMEs to the development of (patented) technology, based on existing databases (PATSTAT and Amadeus). The attempted integration of IP and financial data from these two databases, as highlighted throughout this paper, presents both a significant methodological advancement and ongoing challenges.

The methodology that was presented combined automated matching, disambiguation, and classification stages with targeted additional search efforts in an extrapolation step to arrive at a final estimate of the SME contribution to the stock of patents filed by European companies. Several significant methodological challenges had to be addressed in arriving at comprehensive estimates of the share of patenting by SMEs. Given the heterogeneity in name variants across both databases, as well as incomplete information regarding firm size and independence, a series of data treatment processes—harmonization, matching, and disambiguation—were required. Additionally, further sampling was necessary to produce an indicator that represented an acceptable range of estimated shares.

It became clear that the extrapolation step remains critical to the entire exercise, primarily due to the fragmented nature of ownership information compared to scenarios where only entity size indicators are considered.[15] These additional, time- and resource-consuming efforts arise directly from the incompleteness of current financial databases.

These challenges underscore the pressing need for more comprehensive and historically consistent financial databases. Significant gaps in current data sources limit the full potential of integrated IP-financial datasets. Without more complete financial repositories, future research will continue to face limitations in fully realizing the potential of IP-financial data integration.

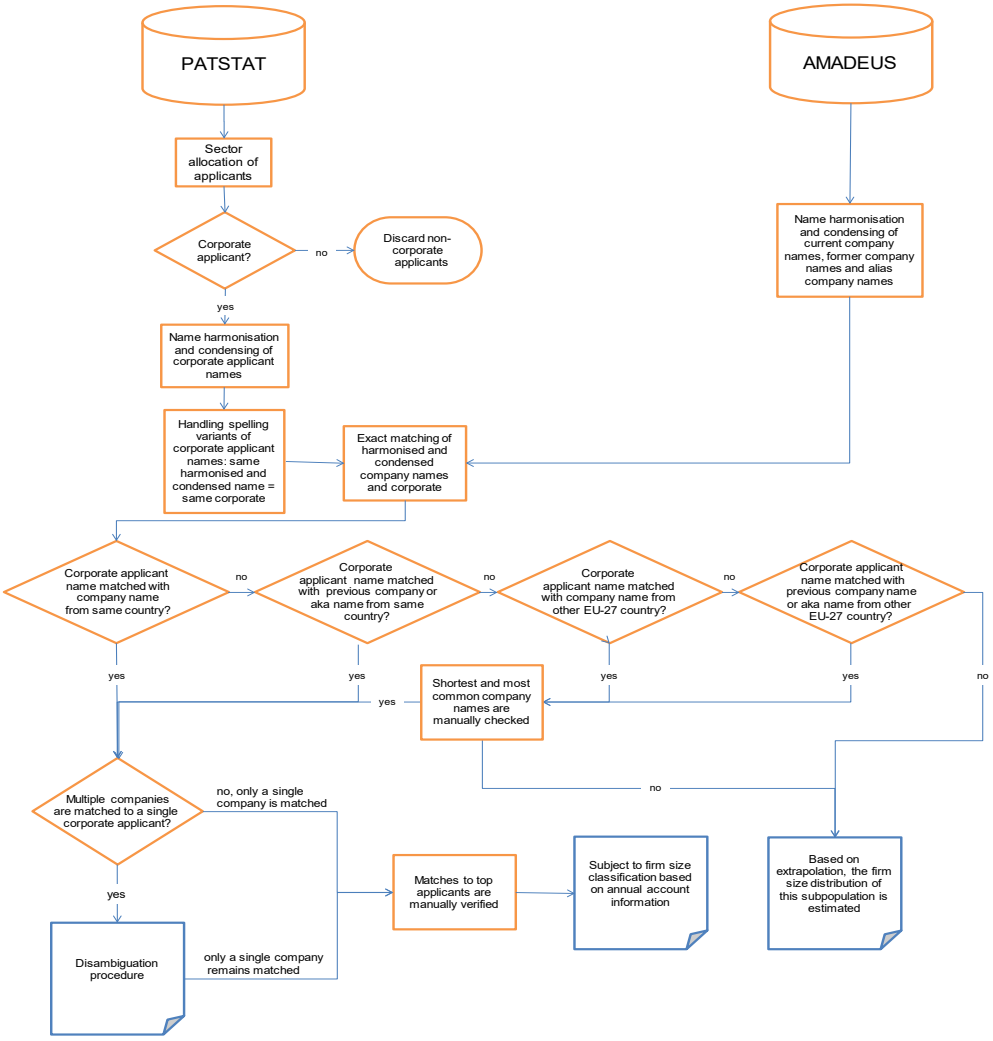
We strongly advocate for the development of enhanced financial databases that provide accurate longitudinal firm-level data. Doing so would not only benefit academic research but also equip policymakers and practitioners with the tools necessary to better assess the dynamics of IP practices and their financial impact on firms.

[15] It must be noted that later versions of Bureau Van Dijk products have shown improvement in terms of ownership information coverage.

References

- Belderbos R., D Faems, B Leten, B. Van Looy (2014) Technological Activities and Their Impact on the Financial Performance of the Firm: Exploitation and Exploration within and between Firms . Journal of Product Innovation Management 27 (6), 869-882
- BOIP/FOD Economy (2021) Lecluyse, L., Callaert, J., Pellens, M., Van Looy, B., De Voldere, I. (Dec 2021). Econometric analysis of the use of IPR in Belgium. Analytical report BOIP / FOD Economie, Belgium. Available online: https://economie.fgov.be/en/publication/econometric-analysis-use
- Du Plessis, M., Van Looy, B., Song, X., Magerman, T. (2009). Data production methods for harmonized patent statistics: Patentee sector allocation.
- Leten, B., Belderbos R. & Van Looy B. (2016) Entry and Performance in New Technology Domains: Internal Capabilities and Environmental Influences. Journal of Management Studies, 53, 8, 1257-1291.
- Leten B., Belderbos R. & Van Looy B. (2007) Technological diversification, coherence and performance of firms. Journal of Product Innovation Management, 24, 6, 567-579
- Magerman, T., Van Looy, B. & Song, X. (2006). Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization. EUROSTAT Working Paper and Studies, Luxemburg.
- Van Looy, B., du Plessis, M. & Magerman, T. (2006). Data production methods for Harmonized patent indicators: Assignee sector allocation.' EUROSTAT Working Paper and Studies, Luxemburg.
- Vervenne J-B, Callaert J. Hoskens M., Van Looy B.(2020) To what extent do SMEs contribute to Europe's patent stock? A methodological outline for creating a Europe-wide SME technology indicator. Scientometrics 127 (6), 3049-3082
- Callaert, J. M Du Plessis, J Grouwels, C Lecocq, T Magerman, B Peeters, Van Looy B. (2011) Patent statistics at Eurostat: Methods for regionalisation, sector allocation and name harmonisation. Eurostat Methodologies and Working Papers, Luxemburg

FIGURE 1: PROCESS FLOWCHART FOR MATCHING PROCEDURE



IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Uplifting the usability of financial statement data to leverage the quality of the Sectoral Balance Sheet (SBS) estimates[1]

Jittima Dummee, Prossaya Jitapunkul, Natchaya Mahaviriyakul and Pornsawan Rugpenthum, Bank of Thailand

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

Uplifting the Usability of Financial Statement Data to Leverage the Quality of the Sectoral Balance Sheet (SBS) Estimates

Authors: Natchaya Mahaviriyakul
Pornsawan Rugpenthum
Prossaya Jitapunkul
Jittima Dummee

Abstract

One of the main challenges in compiling the SBS is the sourcing of data for financial asset and liability positions of nonfinancial corporations (NFC), households (HH), and nonfinancial institutions serving households (NPISH), particularly claims and liabilities among themselves where mirror data (monetary and financial statistics, external sector statistics, and government finance statistics) are not available as in the case of positions vis-à-vis financial corporations, general government, and the rest of the world. Although administrative records such as financial statements reported to relevant government units, theoretically, could serve as the data source for compiling SBS of these three sectors, the quality, frequency, and timeliness concerns could still hinder their usability in practice.

This paper focuses on developing a framework for improving the quality of the SBS estimates through the use of statistical techniques to

(i) impute missing values for selected elements in the financial statement data which registered companies nationwide reported to the Ministry of Commerce (MOC) annually; (ii) derive quarterly estimates of the SBS from the annual financial statement data; and (iii) produce nowcasting estimates for SBS of recent quarters for which annual financial statements are not yet reported to the authorities.

In addressing the three aforementioned issues, a number of statistical techniques were explored to extract 'peer pattern' at different levels of disaggregation; and tested how well each method (or a combination thereof) could be applied to estimate the concerned elements/units of financial statement data. The results and associated statistical errors for each method are evaluated and compared to identify the option that produce the best estimates of selected financial statement elements (e.g., equity, trade credits, etc.).

Given the significance of NFC sector in Thailand's overall SBS, the result of this study should serve as a key steppingstone towards improving the quality of Thailand's SBS compilation and pave way for its dissemination in the near future.

Disclaimer: The opinions expressed herein are those of the authors. They do not purport to reflect the opinions or views of the Bank of Thailand or its members

Introduction

A number of studies on the root causes and warning signs of the world's recent economic crises (namely, those occurred in Mexico, East Asia, and Russia in the 1990s, the global financial crisis originating in the US financial sector during 2007-2009, and public debt crisis in European countries) reveals that one contributing factor is the lack of data/statistics to support insightful analysis, thus hindering timely detection of associated risks/vulnerabilities.

Quality, timely, and comprehensive data is one prerequisite for well-informed decision-making and effective macroeconomic policy formulation. High-quality data with sufficient level of granularity, frequency, timeliness and dynamic enough to keep pace with financial innovations are key to early detection of pre-crisis symptoms, thereby allowing reasonably ample time to design preventive and remedial measures to mitigate the impacts.

As the financial system evolves, complexity has been growing in the form of more (1) diversification of players (i.e. investors/net savers, intermediaries, investment recipients/borrowers); (2) more variety of financial instruments and hybrids thereof; and (3) increasing number of investment channels/platforms. These have concertedly deepened financial linkages among different sectors/sub-sectors of players, as well as the transmittal mechanism through which risks could be spilled over to these players.

The statistics which could shed light on these inter-sectoral financial linkages is **"the sectoral balance sheet" (SBS)**. The SBS comprehensively illustrates financial claims that one sector/sub-sector has vis-à-vis one another in the economy, as well as foreign countries (referred to as "rest-of-the world sector" in the SBS jargon). This information enables policy makers to assess and analyze vulnerabilities associated with each economic sector, and the plausibility of the "risk spillovers" transmitted through this inter-sectoral connectedness.

To derive a quality SBS, however, is rather challenging. In Thailand's context, "decentralized statistical system" is adopted, with several government agencies and institutions involved in the data collection and compilation of different economic and financial statistics. Legal ground empowering data collection/reporting for each agency/institution also varies in degree and scope. This results in difficulties in gathering necessary information and data to support SBS compilation, particularly the data for NFC, HH, and NPISH sectors.

As an attempt to overcome these challenges, the author team explored different sources of readily available data (e.g., secondary data from other macroeconomic accounts, administrative records such as "financial statement data" which corporates are required by law to submit to the MOC on an annual basis, etc.) and assessed their usability to support the compilation of the SBS. Details pertaining to the principles and processes for data cleansing and preparations are elaborated in later sections.

Principles for Preparing Thailand's SBS

A balance sheet, also known as "a statement of financial positions", is a statement that illustrates the assets, liabilities, and equity at a specific date. Typically, it is prepared at the end of an accounting period by accumulating amounts from the start of operation until the specified date as shown in the financial report. This is illustrated in the accounting equation below:

$$\text{Asset} = \text{Liability} + \text{Equity}$$

Constructed on the ground of this balance sheet concept, the SBS systematically reflects how each sector of the economy relates to one another through financial claims (i.e., asset and liability positions that each sector has vis-à-vis the rest). Financial assets and liabilities are further broken down by financial instrument, namely:

1. Monetary gold and Special Drawing Rights (SDRs)
2. Currency and deposits
3. Debt securities
4. Loans
5. Equity and investment fund shares
6. Insurance, pensions, and standardized guarantees
7. Financial derivatives and employee stock options
8. Trade credits and other accounts receivable/payable

Sector-wise, entities in the economy are categorized according to international sectorization standard, as follow:

1. Central Bank (CB)
2. Other Depository Corporations (ODC)
3. Other Financial Corporations (OFC)
4. Non-Financial Corporations (NFC), which can be further broken down into:
    - Public Non-Financial Corporations (PNFC)
    - Other (i.e., private-owned) Non-Financial Corporations (ONFC)
5. General Government (GG), which can be divided into:
    - Central Government (CG)
    - Local Government (LG)
6. Households and Nonprofit Institutions Serving Households (HH & NPISH)
7. Rest of the World (ROW)

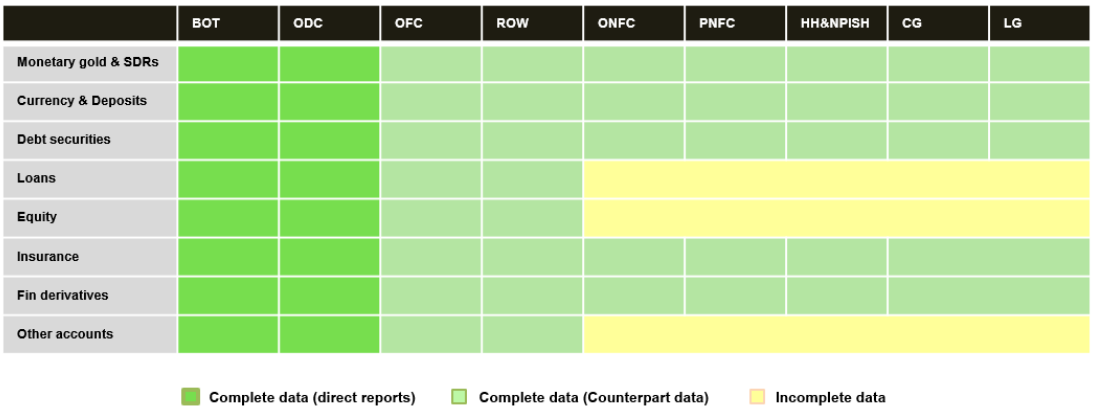
Data sources and limitations

The data sources used by the Bank of Thailand (BOT) for SBS compilation encompass the followings:

1. Monetary and financial statistics (MFS) for positions of CB, ODC, and OFC vis-à-vis the remaining sectors
2. International investment position (IIP) for positions of ROW vis-à-vis the remaining sectors
3. Securities database
4. Financial statements of NFCs, which corporates are required by law to submit to the Ministry of Commerce on an annual basis.

Data with high quality and comprehensiveness are typically those of CB and ODC sectors (as they are closely regulated and subject to regular auditing), followed by OFC and ROW. Meanwhile, data gaps in terms of the coverage of NFC, HH, and NPISH sectors' financial positions remain.

Table 1: Coverage assessment of Thailand's SBS



In an attempt to narrow such data gaps, this study focused primarily on the fourth data source, which is the most challenging in terms of data processing, yet assessed to have the highest impact in enhancing the coverage of NFC sector (with HH and NPISH positions vis-à-vis NFC as by-products as well).

NFCs can be divided into two categories in accordance with the input source:

1. For listed companies: data from securities database (i.e., Financial Market Instrument (FMI) system, maintained processed by the BOT)
2. For non-listed companies: financial statements reported annually to the MOC (so called — the Corporate Profile and Financial Statement (CPFS) database), available since 2001

The latter part was particularly of concern, with several limitations encountered as follow:

(i) Financial statements of companies in Thailand are reported annually. Therefore, quarterly positions have to be imputed for compiling quarterly SBS.

(ii) Some companies failed to submit their financial statements in time, hence causing 'missing values' at the time CPFS data were processed for inclusion in SBS compilation.
(iii) Some companies did not report all items in the reporting template for certain years, resulting in omissions scattering across some financial instruments.
(iv) CPFS data are made available to SBS compiler with approximately seven-month lag, thus requiring techniques to produce preliminary estimates for the current year.

Subsequent sections illustrate different methods explored with an aim to overcome these limitations in deriving positions of non-listed companies.

Data preparation: Imputing missing values for selected financial instruments for private non-listed companies (non-listed ONFC)

Step 1: Filtering CPFS data for non-listed ONFC

- The study was conducted on the most recent 3-year CPFS data.
- Companies in the financial sector (ISIC Rev.4 Section K) were excluded.
- State-owned enterprises (SOEs) were excluded.
- Companies listed in the Stock Exchange of Thailand (SET) and Market for Alternative Investment (MAI) were excluded.
- Companies which were suspended or no longer operated were excluded.
- Only those companies submitting financial statements for at least two consecutive years and have total income in the most recent year of more than 100,000 baht were included (i.e., "cutting off tail" with insignificant value of assets/liabilities).

Step 2: Conducting statistical tests to identify optimal method in imputing missing values

Thorough review of CPFS database revealed that most companies consistently reported data on equity and inter-company lending (if any). On the contrary, approximately 10 percent of companies did not report trade credits receivable; and five percent not reporting trade credits payable for at least one of the three years investigated. Therefore, this study emphasized on exploring different options in imputing missing values for these instruments.

Step 2.1: Creating "mock-up test data":

1. Selecting only those companies submitting financial statements for three consecutive years. Actual reported data were used as "benchmark" for this statistical test.
2. Removing some data points for certain years, so as to create mock-up test data:
    - Removing data points of some companies for one of the three years
    - Removing data points of some companies for two consecutive years
    - Removing data points of some companies for two non-consecutive years
    - Removing data points of some companies for the three years

Step 2.2: Testing different methods in imputing missing values:

Four different methods tested include:

Method 1: "Overall Growth Method"

This method involves calculating year-on-year growth (%YoY) of respective financial statement items of all reporting companies combined (i.e., %YoY at aggregate level) and applying this percentage to estimate missing data for non-reporting companies in each of the three years tested. Table 2 illustrates how missing data were imputed under this method (imputed data are in yellow cells).

Table 2: Imputing missing data under Method 1

(assuming 2020 - 2022 %YoY were 15.5%, 12.4%, and 17.9%, respectively)

Company	2019	2020	2021	2022
Company A	1,306	1,306 * (15.5% + 1)	1,500	1,713
Company B	1,306	1,231	1,231 * (12.4% + 1)	1,713
Company C	1,306	1,306 * (15.5% + 1)	(1,306 * (15.5% + 1)) * (12.4% + 1)	1,713
Company D	1,306	1,231	1,231 * (12.4% + 1)	(1,231 * (12.4% + 1)) * (17.9% + 1)

Method 2: "Cluster-wise Growth Method"

This method involves calculating year-on-year growth (%YoY) of respective financial statement items of all reporting companies in each ISIC section (i.e., %YoY for each ISIC 1-digit level) and applying these percentages to estimate missing data for non-reporting companies in each ISIC section for the three years tested. Table 3 illustrates how missing data were imputed under this method (imputed data are in yellow cells).

Table 3: Imputing missing data under Method 2

(assuming 2020 - 2022 %YoY were 5.3%, 4.7%, and 5.6%, respectively)

Company	ISIC	2019	2020	2021	2022
Company A	J	1,306	1,306 * (10.8% + 1)	1,500	1,713
Company B	B	1,306	1,231	1,231 * (4.7% + 1)	1,713
Company C	J	1,306	1,306 * (10.8% + 1)	(1,306 * (10.8% + 1)) * (14.2% + 1)	1,713
Company D	A	1,306	1,231	1,231 * (4.7% + 1)	(1,231 * (4.7% + 1)) * (5.6% + 1)

Method 3: "Hybrid Method"

This method applies Method 2 for 'manufacturing and wholesale-retail trade' ISIC sections, and Method 1 for other ISIC sections.

Method 4: "Time-Based Method"

This method assumes "linear growth" for companies not reporting data for only one year. For companies with missing data for two years or more, Method 2 was applied. Table 4 illustrates how missing data were imputed under this method (imputed data are in yellow cells).

Table 4: Imputing missing data under Method 4

(assuming 2020 - 2022 %YoY were 10.8%, 14.2%, and 15.9%, respectively)

Company	2020	2021	2022
Company A	1,500 + (1,500 – 1,713)	1,500	1,713
Company B	1,231	1,231 + ((1,713 – 1,231)/2)	1,713
Company C	1,231	1,500	1,500 + (1,500 – 1,231)
Company D	1,306 * (10.8% + 1)	(1,306 * (10.8% + 1)) * (14.2% + 1)	1,713
Company E	1,500 * (10.8% + 1)	1,500	1,500 * (15.9% + 1)
Company F	1,231	1,231 * (14.2% + 1)	(1,231 * (14.2% + 1)) * (15.9% + 1)

Step 2.3: Assessing efficacy of each method

The statistics used to assess the efficacy of each of the four methods include:

1. Mean square error (MSE)
2. Root mean square error (RMSE)
3. Mean absolute error (MAE)
4. Mean absolute percentage error (MAPE)

The method which yields the lowest value for all (or most of the) four statistics is deemed to have highest efficacy and would therefore be chosen as preferred method, to be applied in actual compilation of the SBS.

Based on statistical tests conducted on the mock-up data, **"Hybrid Method" yielded the best results for trade credits receivable**. Meanwhile, **Method 1 (Overall Growth Method) returned the lowest statistical error for trade credits payable**. These two methods were then applied to

impute missing data for the entire CPFS dataset for the aforementioned two financial statement items. Completed data were then used for SBS compilation for periods up to 2023.

Compiling Quarterly SBS: Constructing quarterly estimates based on annual data

Not all financial statement items and counterpart sectors have quarterly inputs to support quarterly compilation of the SBS. This is particularly the case for selected ONFC's assets and liabilities, for which the annual CPFS data serve as the only source. Thus, estimates of quarterly positions were derived from annual time series, with the assumption that changes to the annual positions occurred at a constant rate and distributed evenly across the four quarters. Table 7 depicts how quarterly series were derived based on this assumption.

Table 5: Example of how to convert annual data to quarterly data

Q4/2021	Q1/2022	Q2/2022	Q3/2022	Q4/2022
1,000	1,000 + ((1,400 – 1,000) / 4)	Q1/2022 + ((1,400 – 1,000) / 4)	Q2/2022 + ((1,400 – 1,000) / 4)	1,400

Nowcasting the SBS for the current year

As CPFS data are made available to SBS compilers with relatively long time lag, a method to nowcast the SBS for NFC sector needs to be developed to enable a more timely compilation and dissemination of the SBS. The following variables were short-listed as conceptually relevant to the targeted financial statement items. These variables are available with significantly less time lag; and hence could serve as indicative information for designing nowcasting models for the targeted items.

- Relevant variables for trade credits receivable: total sales or income *(e.g. from financial data of listed companies disclosed by the SET, data on sales reported to the Revenue Department's value-added tax (VAT) database)*
- Relevant variables for trade credits payable: production volume, inventory, sales, etc.

CPFS data for 2011 – 2023 were tested for cross-correlation between those short-listed variables and the targeted items (i.e., trade credits receivable/payable). Tests were conducted both on the value of the variables themselves and their year-on-year growth rate. For data on sales, the total sales of all ONFCs as well as sales of those in manufacturing and trading businesses (i.e., ISIC sections C and G, which are deemed to have closer link to trade credits) were tested. The study revealed the following results:

Table 6: Cross-correlation between selected variables and trade credits receivable

For trade credits receivable	Variables' Value	Variables' %YoY
SET	0.66	0.40
VAT	0.90	0.75
VAT C+G	0.88	0.70

Remark : Test period was 2011-2023 (for variables' value), and 2012-2023 (for %YoY)
C denotes manufacturing sector G denotes wholesale and retail sector
Source: SET, The Revenue Department, calculated by the BOT

Table 7: Cross-correlation between selected variables and trade credits payable

For trade credits payable	Variables' Value	Variables' %YoY
SET	0.66	0.39
Production (MPI)	0.87	0.38
VAT_C	0.90	0.70
VAT_C+G	0.88	0.77

Remark : Test period was 2011-2023 (for variables' value), and 2012-2023 (for %YoY)
C denotes manufacturing sector G denotes wholesale and retail sector
MPI denotes manufacturing production index
Source: SET, The Revenue Department, Ministry of Industry, calculated by the BOT

Results in Tables 6 and 7 reflected that "total sales of all business sectors combined" was highly correlated with trade credits receivable; while "sales of manufacturing and wholesale/retail sectors alone" exhibits the highest correlation with trade credits payable. These findings were used as a basis for nowcasting preliminary estimates for ONFC's trade credits receivable/payable. These estimates would be revised thereafter upon availability of the CPFS data of respective year.

Conclusion

In recognition of the need for comprehensive tools to monitor inter-sectoral financial linkages and plausible risk spillovers, the BOT has commenced Thailand's compilation of the SBS some years ago, with continuous effort in exploring and integrating supplementary sources to enhance the coverage and quality of the SBS. In an attempt to narrow the data gaps on ONFC's positions, financial statement data reported annually to the MOC were identified as potential data source. Despite its rich information and considerable granularity, limitations exist in regard to the timeliness and completeness of the data reported.

To overcome these limitations, the author team conducted statistical tests to identify appropriate techniques to (i) impute missing values (for companies not reporting trade credits receivable/ payable for certain years); (ii) derive quarterly estimates of the SBS from the annual financial statement data; and (iii) nowcast quarterly SBS for the current year. This, together with the enhanced coverage of other comprehensive data sources (namely, the MFS, the IIP, and securities database) should ensure considerable improvement to the coverage, quality, as well as timeliness of Thailand's SBS.

BANK OF THAILAND

Uplifting the Usability of Financial Statement Data to Leverage the Quality of the Sectoral Balance Sheet Estimates

Prepared for the workshop on
"New Insights from Financial Statements"
17 October 2024
Madrid, Spain

Natchaya Mahaviriyakul
External Sector Statistics Division
Data Management and Analytics Department
Bank of Thailand

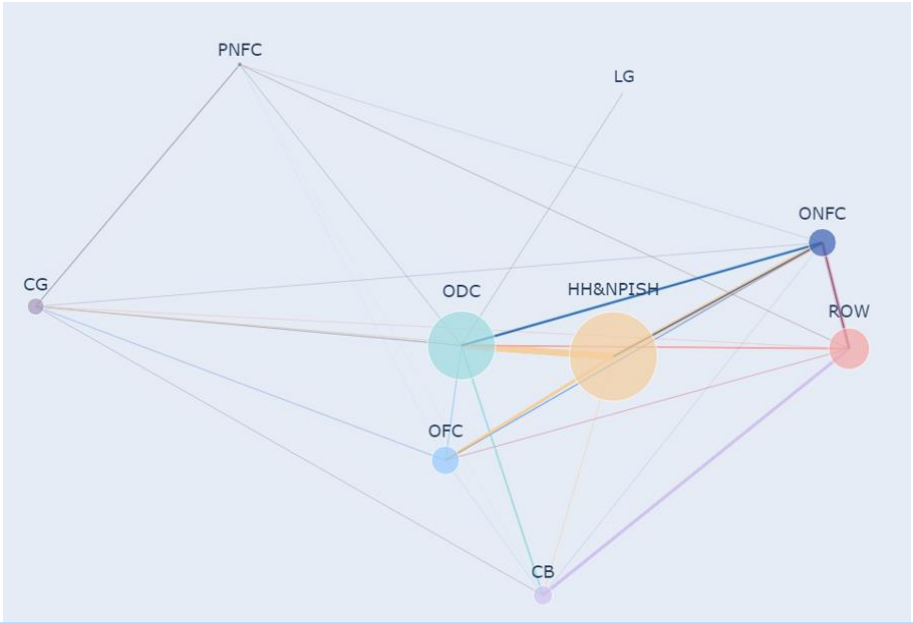
Discussion Topics

- **Importance of Sectoral Balance Sheet (SBS)**
- Components and structure of SBS
- **Challenges and Constraints**
- Addressing Constraints
- **Case study: Other Non-Financial Corporations (ONFC)**
  - Impute Missing data
  - Increasing the Frequency
  - Extrapolating Forward Data
- Conclusion and Way Forward

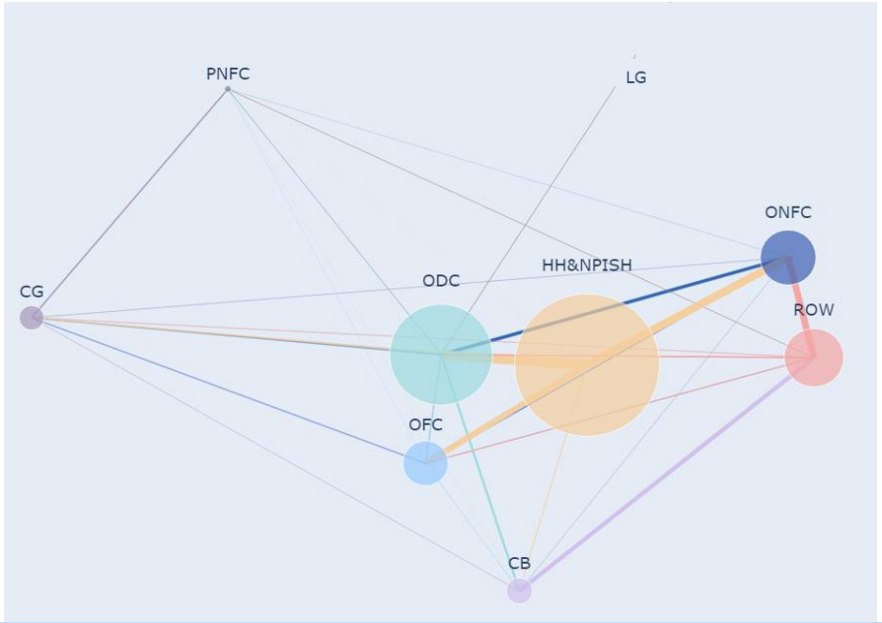
Importance of SBS

Intersectoral Asset and Liability Network

Q4/2013



Q4/2023



Source: BOT

Importance of SBS

Maturity Mismatch

- High Short-term Liabilities
- Illiquid Assets
- Ability to roll-over

Capital Structure Mismatch

- High Liabilities
- Low Equity
- Lead to higher expenses

Currency Mismatch

- High Foreign Currency Liabilities
- Low Foreign Assets
- Vulnerable to changes in FX rate

Solvency Problem

- Unable to repay debt
- Credit risk
- Spill-over effects

Source: Sathit Thaensat and Jitchaya Buranahiran (2014), SDDS Plus Standards: The Next Step for Thailand's Economic and Financial Data Dissemination, FOCUSED AND QUICK (FAQ) Issue 88.

Components of SBS

Economic Sector

- Central Bank **(CB)**
- Other Depository Corporations **(ODCs)**
- Other Financial Corporations **(OFCs)**
- Non-Financial Corporations **(NFCs)**
  - Other Non-Financial Corporations **(ONFC)**
  - Public Non-Financial Corporations **(PNFC)**
- General Government **(GG)**
  - Central Government **(CG)**
  - Local Government **(LG)**
- Households and Nonprofit Institutions Serving Households **(HH & NPISH)**
- Rest of the World **(ROW)**

Financial Instrument

- Monetary gold and Special Drawing Rights (SDRs)
- Currency and deposits
- Debt securities
- Loans
- Equity and investment fund shares
- Insurance, pensions, and standardized guarantees
- Financial derivatives and employee stock options
- Other accounts receivable/payable

Challenges in Data Collection

Coverage assessment of Thailand's SBS

	CB	ODC	OFC	PNFC	ONFC	HH&NPISH	CG	LG	ROW
Monetary gold & SDRs									
Currency & Deposits									
Debt securities									
Loans									
Equity									
Insurance									
Fin derivatives									
Other accounts									

Complete data (direct reports) Complete data (Counterpart data) Incomplete data

Case study : ONFC's equity + trade credits & other accounts receivable/payable

Data source for ONFC

- **Listed companies:** Securities database called "***Financial Market Instrument (FMI) System***"
  Financial statements disclosed by SET
- **Non-listed companies :** Financial statements submitted to Ministry of Commerce's (MOC)
  "***Corporate Profile and Financial Statement (CPFS) database***"

CPFS data constraints

1. Completeness :

- Missing data for some years/some companies for trade credits receivable/payable (TCR, TCP)
- Complete data for equity

2. Frequency :

- Annual

3. Timeliness :

- 7-month lag (i.e., available in July of the following year)

Addressing Constraints

A framework for improving the quality of CPFS data

I. **Impute missing values** for TCR, TCP in the financial statement data which registered companies nationwide reported to the MOC annually

II. **Derive quarterly estimates** of the SBS from annual financial statement data

III. **Produce nowcasting estimates** for SBS of recent quarters for which annual financial statements are not yet reported to the authorities

Data Preparation

I. Use company-level financial statement data for 3 years.

II. Segregating non-listed ONFCs by

- Excluding financial corporations (ISIC = K)
- Excluding State-Owned Enterprises (SOEs), following GFS' list of SOE
- Excluding companies listed in SET & MAI

III. Exclude companies which were no longer operated

And randomly remove data to cover all possible cases in order to create a test dataset for data imputation.

Impute Missing Data

Statistical techniques adopted to impute missing values for TCR/TCP:

1. Overall Growth Method:

- Calculate %YoY of reported data at aggregate level

- Applying this % to estimate missing data for non-reporting companies in each of the 3 years tested

2. Cluster-wise Growth Method:

- Calculate %YoY of reported data at 1-digit ISIC level

- Applying these % to estimate missing data for non-reporting companies in each ISIC section for each of the 3 years tested

3. Hybrid Method:

- Applying Method 2 for 'manufacturing and wholesale-retail trade' (ISIC C & G sections)

- Applying Method 1 for other ISIC sections

4. Time-based Method :

- Applying "linear growth" for companies not reporting data for only 1 year

- Applying Method 2 for companies with missing data for 2 years or more

Method 1 : Overall Growth Method

Example: Assuming 2020 - 2022 %YoY were 15.5%, 12.4%, and 17.9%, respectively

Company	2019	2020	2021	2022
Company A	1,306	1,306 * (15.5% + 1)	1,500	1,713
Company B	1,306	1,231	1,231 * (12.4% + 1)	1,713
Company C	1,306	1,306 * (15.5% + 1)	(1,306 * (15.5% + 1)) * (12.4% + 1)	1,713
Company D	1,306	1,231	1,231 * (12.4% + 1)	(1,231 * (12.4% + 1)) * (17.9% + 1)

Method 2 : Cluster-wise Growth Method

Example: Assuming 2020 - 2022 %YoY for ISIC J were 10.8%, 14.2%, and 15.9%, respectively and %YoY for ISIC A & B were 5.3%, 4.7%, and 5.6%, respectively

Company	ISIC	2019	2020	2021	2022
Company A	J	1,306	1,306 * (10.8% + 1)	1,500	1,713
Company B	B	1,306	1,231	1,231 * (4.7% + 1)	1,713
Company C	J	1,306	1,306 * (10.8% + 1)	(1,306 * (10.8% + 1)) * (14.2% + 1)	1,713
Company D	A	1,306	1,231	1,231 * (4.7% + 1)	(1,231 * (4.7% + 1)) * (5.6% + 1)

Method 3 : Hybrid Method

Example: Assuming 2020 - 2022 %YoY were 15.5%, 12.4%, and 17.9%, respectively and for ISIC C were 10.8%, 14.2%, and 15.9%, respectively

Company	ISIC	2019	2020	2021	2022
Company A	C	1,306	1,306 * (10.8% + 1)	1,500	1,713
Company B	A	1,306	1,231	1,231 * (12.4% + 1)	1,713
Company C	C	1,306	1,306 * (10.8% + 1)	(1,306 * (10.8% + 1)) * (14.2% + 1)	1,713
Company D	A	1,306	1,231	1,231 * (12.4% + 1)	(1,231 * (12.4% + 1)) * (17.9% + 1)

Method 4 : Time-Based Method

Example: Assuming 2020 - 2022 %YoY for ISIC J were 10.8%, 14.2%, and 15.9%, respectively

Company	2020	2021	2022
Company A	1,500 + (1,500 – 1,713)	1,500	1,713
Company B	1,231	1,231 + ((1,713 – 1,231)/2)	1,713
Company C	1,231	1,500	1,500 + (1,500 – 1,231)
Company D (ISIC = J)	1,306 * (10.8% + 1)	(1,306 * (10.8% + 1)) * (14.2% + 1)	1,713
Company E (ISIC = J)	1,500 * (10.8% + 1)	1,500	1,500 * (15.9% + 1)
Company F (ISIC = J)	1,231	1,231 * (14.2% + 1)	(1,231 * (14.2% + 1)) * (15.9% + 1)

Missing value : 1 year

Missing value : 2 years

Impute Missing Data

Statistical Testing :

- **MSE:** Mean Square Error
- **RMSE:** Root Mean Square Error
- **MAE:** Mean Absolute Error
- **MAPE:** Mean Absolute Percentage Error

The method which yields the lowest value for all (or most of the) four statistics is deemed to have highest efficacy and would be chosen as preferred method, to be applied in actual compilation of the SBS.

Results :

Trade Credits Receivable : Method 3

"Hybrid Method"

Trade Credits Payable : Method 1

"Overall Growth Method"

Convert Annual data to Quarterly data

- In compiling SBS, inputs for most sectors & instruments sourcing from MFS, IIP, GFS, FMI system are available quarterly.
- Only the positions of NFC's selected instruments where CPFS is the sole data source are available only on an annual basis.
- Derive quarterly from annual positions, assuming that changes to the annual positions occurred at a constant rate and distributed evenly across the four quarters.

Q4/2021	Q1/2022	Q2/2022	Q3/2022	Q4/2022
1,000	1,000 + ((1,400 – 1,000) / 4)	Q1/2022 + ((1,400 – 1,000) / 4)	Q2/2022 + ((1,400 – 1,000) / 4)	**1,400**

Statistical Tests to Develop Nowcasting Models for SBS

- CPFS data are available with long time lag
- Need to develop a method to nowcast positions for NFC's selected items for more timely compilation/dissemination of SBS

Target : Nowcasting trade credits receivable & payable

Trade Credits Receivable

Variables tested: total sales or income
(from SET, VAT database)

Trade Credits Payable

Variables tested: production volume, inventory, sales
(from SET, MPI, VAT database)

Trade Credits Receivable :

"Total sales (aggregate level)" was highly correlated with trade credits receivable.

Cross-correlation analysis results :

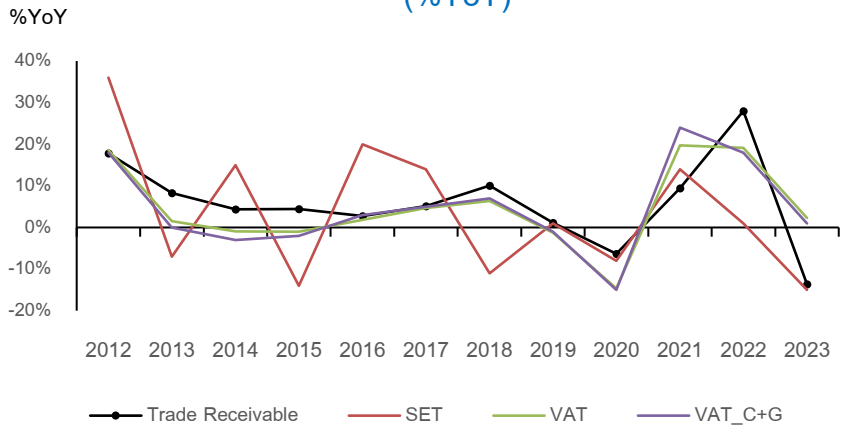
Trade credits receivable

	Variables' Value	Variables' %YoY
SET	0.66	0.40
VAT	0.90	0.75
VAT C+G	0.88	0.70

Remark : Test period was 2011-2023 (for variables' value), and 2012-2023 (for %YoY)
C = manufacturing sector G = wholesale and retail sector

Source: SET, Revenue Department, calculated by the BOT

Trade Credits Receivable vs Determining Variables (%YoY)



Trade Credits Payable :

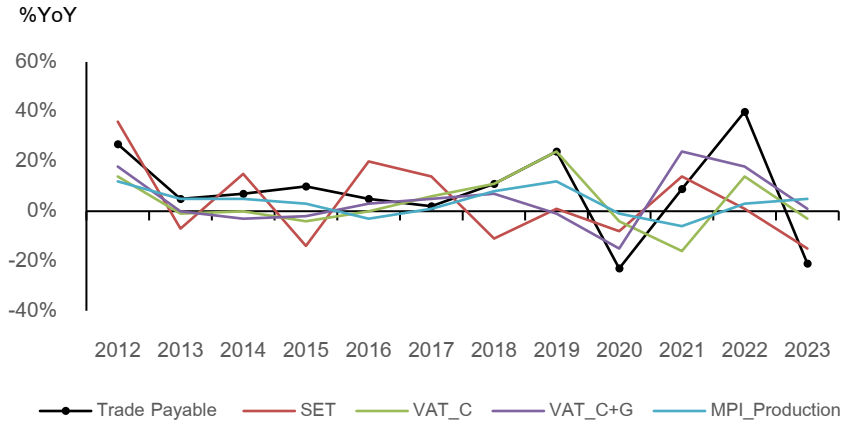
"Sales of manufacturing and wholesale/retail sectors" exhibits highest correlation with trade credits payable.

Cross-correlation analysis results :
Trade credits payable

	Variables' Value	Variables' %YoY
SET	0.66	0.39
Production (MPI)	0.87	0.38
VAT_C	0.90	0.70
VAT_C+G	0.88	0.77

Remark : Test period was 2011-2023 (for variables' value), and 2012-2023 (for %YoY)
C = manufacturing sector G = wholesale and retail sector
MPI = manufacturing production index
Source: SET, Revenue Department, Ministry of Industry, calculated by the BOT

Trade Credits Payable vs Determining Variables (%YoY)



Conclusion and Way Forward

Conclusion : Addressing 3 constraints to improve quality of CPFS and SBS

	Equity	Trade Credits & Other Accounts Receivable/Payable	
		Trade Credits Receivable	Trade Credits Payable
Impute missing values	No missing values	Method 3 : Hybrid Method	Method 1 : Overall Growth Method
Derive quarterly estimates	Assuming that changes to the annual positions occurred at a constant rate and distributed evenly across the four quarters		
Produce nowcasting estimates	CPFS : Equity (carry forward)	Total sales from VAT database (%YoY)	Sales (C+G) from VAT database (%YoY)

Conclusion and Way Forward

Conclusion

Limitations in quality and timeliness of financial statement data can be overcome by means of statistical techniques, thereby leveraging their usability for SBS compilation

Way Forward

- **The BOT plans to disseminate SBS in the near future**
- **Promote the use of SBS for improved policymaking, enhanced risk assessment, greater transparency and strengthened economic resilience**
- **Regularly review statistical methods for nowcasting to further improve the quality of the estimates**

Thank you

Contact Information

External Sector Statistics Division, Bank of Thailand

For more information, please feel free to contact us via the email below.

Angsupalee Wacharakiat	Pornsawan Rugpenthum	Prossaya Jitapunkul	Jittima Dummee	Natchaya Mahaviriyakul
Tel : +66 023567720	Tel : +66 023567340	Tel : +66 022835638	Tel : +66 023567325	Tel : +66 023566425
Email angsupaw@bot.or.th	Email : pornsawr@bot.or.th	Email : prossaj@bot.or.th	Email : jittimad@bot.or.th	Email : natchaym@bot.or.th

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Green financing and the relationship between banks and non-financial corporations through the lens of balance-sheet interaction[1]

Michael Harrison and Mimoza Shabani,
University of East London

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

Green financing and the relationship between banks and non-financial corporations through the lens of balance-sheet interaction

Michael Harrison[1] and Mimoza Shabani[2]

Abstract

Green investment is a crucial element in the transition towards achieving lower emissions on a global scale. Non-financial corporations (NFCs) have a pivotal role in the transition process by investing in green projects together with banks in providing the necessary finance. This paper uses a novel dataset which includes all the green financing transactions between banks and NFCs in all countries in the European Economic Area (EEA), United Kingdom (UK), and Switzerland for the period between 2014-2024. Our analysis focuses on examining the factors that explain the demand side for green financing, which reflects the NFCs, and the supply side, which reflects banks as the providers of green finance. Our empirical results suggest that there are distinct differences in the factors that determine the amount and preference for green finance and relate to the size of NFCs as well as the sectors in which they operate. Similarly, banks' balance sheet variables also explain their supply of green finance and their likelihood to issue loans or bonds.

Keywords: green finance; syndicated loans, non-financial corporation; banks.

JEL classification: Q56; G21; G15.

[1] Senior Lecturer in Economics, University of East London, UK, email: m.harrison@uel.ac.uk
[2] Associate Professor in Economics, University of East London, UK, email: m.shabani@uel.ac.uk

We thank Carmela D'Avino and Maria Tselika for their valuable comments, and conference participants at the Forecasting Financial Markets (FFM) conference (University of Oxford). Special thanks to Siddhesh Metkar and Daniela Tulgara for their research assistance. All errors remain our own.

Contents

1. Introduction 3
2. Literature review 4
3. Data 5
4. Methodology 8
5. Results 10
6. Conclusion and Policy Recommendations 14

References: 16

Appendix 17

Appendix I- Variable Definitions 17

Appendix II- Supplementary findings 18

1. Introduction

Recently, significant attention has focused on the economic impact of climate change with governments, regulators, and international supervisory authorities alike, striving to achieve a greener economy. Banks are key players in the transition, as they play a fundamental role in providing green finance to non-financial corporations (NFC)[3] that need to fund investments (Del Gaudio *et al.*, 2022; Filipava and Murshudli, 2023). This paper explores the factors influencing the interaction between banks and NFC in supporting projects designed to mitigate the global effects of climate change. We obtain bank and firm-level data from Bloomberg, which provides granular financing transaction details, and analyse the factors of the supply and demand for green finance. Using data for 2297 green bonds and 1968 use of proceeds term loans we link banks which have issued loans and have underwritten bonds to/for corresponding NFC. The transaction level data allows for capturing the amount, maturity, and costs of green financing activities. The dataset contains 292 banks and 1910 NFCs located in European Economic Area (EEA) countries, the United Kingdom and Switzerland for 2014-2024.

Our goal is to identify the determinants of banks and NFC's preferences for green finance. To this aim, we employ a binomial logistic regression model that sheds light on and quantifies the factors determining a firm's preference for green financing through bonds or loans. The empirical approach is two-fold: We first estimate the model using NFC's balance sheet variables to determine their bond or loan preference, similar to the methodology employed in Altunbaş *et al* (2010). Secondly, we also estimate banks' likelihood to issue loans or underwrite bonds taking into account other balance sheet activities.

Overall, the results suggest that balance sheet variables are key determinants in explaining the total amount of green finance NFCs obtained. In particular, leverage, liquidity and size appear to be positively associated with the total amount of green finance, encompassing both loans and bonds. However, the findings indicate that profitability is not a significant factor in determining the amount of green finance raised. Distinguishing between green firms, and potentially polluting firms known as brown firms, the findings yield distinct results. More specifically, the size of the firms and the cost of finance do not explain the amount of green finance received by green firms. In contrast, both variables along with other balance sheet variables remain positive and significant for brown NFCs.

Regarding the preference of NFCs for the type of green finance raised, we find that more profitable and larger firms tend to select bond issuance as the source of funding for their green projects rather than loans. Balance sheet variables for green firms have a more pronounced effect on their choice of green finance. Indeed, green firms with greater liquidity ratios are found to favour bonds over loans.

[3] We use the terms NFC and firm interchangeably.

In contrast, brown firms' liquidity does not explain the preference for the type of finance they raise. However, the cost of finance and the firm's size remain key factors in explaining the preference for all sectors in which firms operate.

Banks' balance sheet indicators also play a significant role in explaining the amount of green finance they issue to the NFCs. Most notably liquidity and size of the bank are positive and significant with the total amount of green finance banks provide. Profitability and the loan growth rate on the other hand are negatively associated with green finance. Therefore, when banks achieve higher profitability levels by more traditional lending activities they tend to reduce the amount of green lending to corporations. Regarding the environmental scores recorded for all banks engaged in green lending, the findings point to a positive relationship. Therefore, banks with higher reported environmental scores increase their green lending activities. However, the higher these scores are, the higher the likelihood that a bank would prefer the issuance of bonds rather than loans to support NFCs green investment. Similar results yield when we account for the size of the banks suggesting that larger banks also increase the likelihood of issuing or underwriting a corporate bond as opposed to issuing loans.

The results enable a new understanding of the factors determining the supply and demand dynamics of green financing. The micro nature of the granular data used in this paper provides detailed insight on the determinants of green finance allowing regulators and market participant to optimise their approaches to green finance. This has far-reaching implications for policymakers as it can serve as a guide and assist them in tailoring policy to support green financial markets as a whole. More specifically, we argue that smaller NFCs must have additional support both accessibility and affordability of green finance.

2. Literature review

There is a growing body of literature on the intersection of climate change and banking. Several studies have focused on banks' role in advancing the transition to low-carbon economies after committing to the goals of the 2015 Paris Agreement. Delis et al (2023) examine the loan prices that banks charge fossil fuel firms and find evidence of higher loan rates in the corporate loan market following the Paris Agreement.

Kacperczyk and Peydro (2022) analyse how banks distribute credit among non-financial corporations and their role in reducing emissions by decreasing credit supply to brown firms and promoting green investment. Using data from 2013-2018, the analysis captures banks with significant exposure to the syndicated loan market that have committed to goals in emission reduction levels in light of the 2015 Paris Agreement. The empirical results suggest that committed banks support low-emission firms by allocating more credit to them while reducing credit to high-emission brown firms. Similarly, Reghezza et al (2022) employ loan and firm-level data to assess the impact of the Paris Agreement and find that banks reallocated credit away from polluting firms.

However, others have cast doubt on the credibility of banks in delivering green finance, as their actions often do not fully reflect their reported sustainability

commitments. Nauman and Morris (2021) report that post-2015 global banks have actually expanded their lending to fossil fuel companies, providing a total of $750bn in 2020 alone. There are also concerns about the transparency of banks in disclosing information on their environmental impact (ECB 2023). Giannetti et al (2023) examine the environmental disclosures and the lending decisions for 101 systemically important European banking groups. The findings suggest that banks claiming to emphasise sustainable lending activities are in fact extending more new loans to firms in brown industries, without a corresponding increase in lending to green industries.

Growing concerns have also emerged regarding the 'greenwashing' practices, casting doubts about the reliability of information surrounding green bond issuance, with the latter being an important product in green financing. The strong demand for green bonds among investors is incentivising their issuance even when firms may not have the means to meet their climate pledges fully[4]. Flammer (2021) examines corporate green bond issuance by public and private firms worldwide from 2013 to 2018 and finds that the stock market reacted positively to the announcement of these issuances. Furthermore, the paper tracks the environmental performance of the firms engaged in green bond issuance and observes an increase in their environmental ratings along with a decrease in CO2 emissions.

3. Data

To examine the factors that influence the amount and type of green financing received by firms and issued by banks, we use transactional level data for the issuance of 2297 bonds and 1968 syndicated loans. We do so for all beneficiary NFC located in EEA, Switzerland and the UK[5] for the period between 2014 and quarter two of 2024[6].

Bloomberg captures individual financing transactions which take the form of bond issuances and syndicated loans to companies. Each transaction reflects an investment in reducing the recipient firm's carbon emissions, not necessarily eliminating all emissions, especially in the case of brown industries. In each case, Bloomberg can consistently apply a green flag to financing activity, where set criteria are met. Bloomberg's process for flagging involves verifying the declared use of proceeds. Issuers must demonstrate, usually with the aid of an ESG verification partner, that 100% of the net proceeds are dedicated to projects promoting climate change mitigation, adaptation, or other environmental sustainability purposes (Bloomberg, 2023). We only consider bonds/loans where Bloomberg has verified the use of proceeds so that we can be sure that greenwashed transactions do not feature in the dataset.

---

[4] See 'Fears rise over 'greenwash' bonds', Financial Times, March 21, 2022.

[5] Full list of countries included: Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom, Switzerland, Iceland, Liechtenstein and Norway.

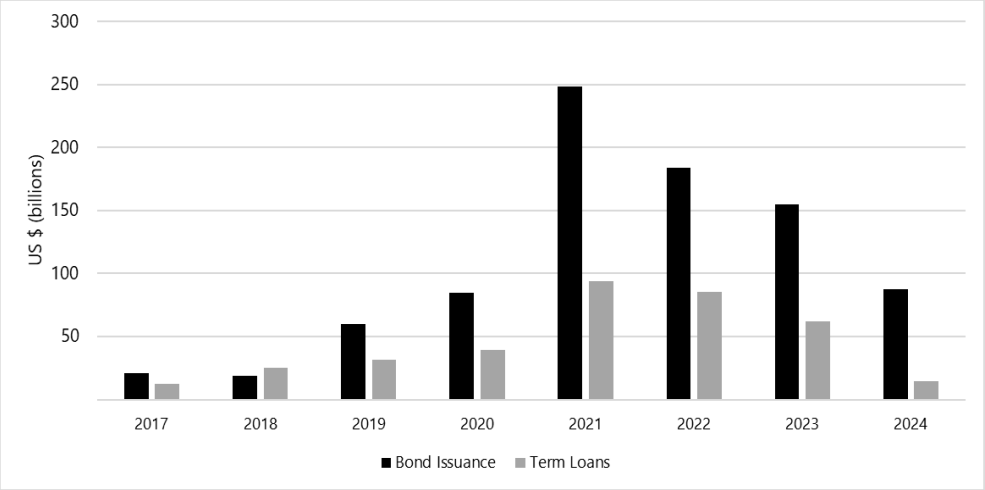
[6] Data from 2014- 2016 is scarce due to inconsistent reporting practices by banks (Bloomberg 2023) and therefore they have been excluded from the final dataset.

We identify a total of 4265 individual transactions for which we also collect balance sheet variables for all NFCs that received green financing. This constitutes all the recorded transactions in the period available from Bloomberg. Further, we identify, in most cases, the bank which has facilitated the transaction. That is, facilitated a bond issuance or served as syndicate lead for a loan issuance. Having identified the banks which are involved in the supply of green finance we also collect individual bank-level data. In this sense we capture data pertaining to both sides of an individual transaction, the demand side using firm-level data and the supply side using bank-level data.

Graph 1 shows the total amount of bond and loan financing obtained by NFCs over the sample period. In 2021, green financing reached a total of $341.65 billion, with bond issuance accounting for $248.09 billion of that amount.

Total bond and loan financing 2017-2024

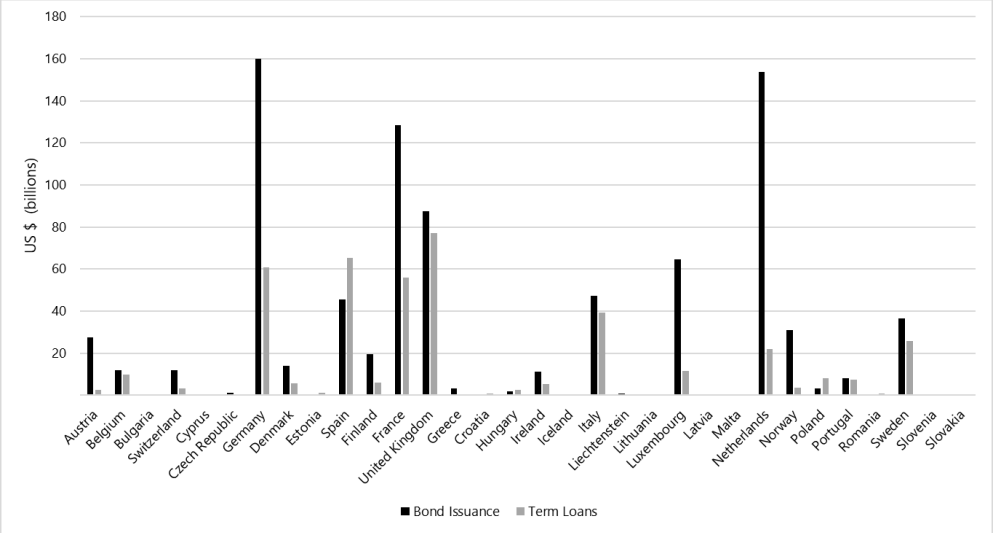
Graph 1



Graph 2 presents the amount and type of green financing based on the domicile of the recipients, highlighting the countries in which the NFCs are located. Firms domiciled in Germany, Netherlands, France, the UK, Spain, and Luxembourg are the largest recipients of green financing. As shown in Graph 2 bond issuance is the dominant channel through which green finance has been obtained.

Total Green Financing sorted by type and domicile of recipient

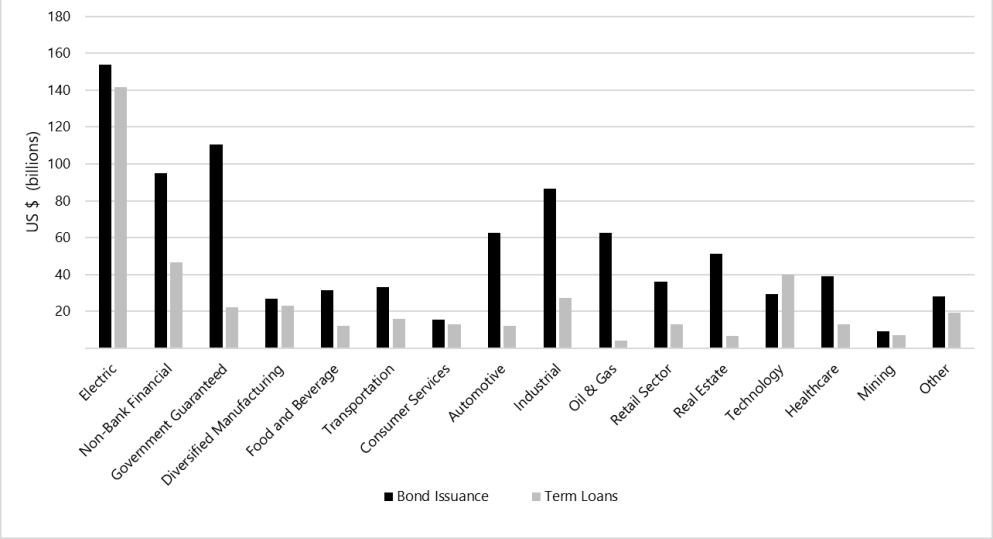
Graph 2



Regarding the sectors in which NFCs operate, as presented in Graph 3, a significant variation in the amount of funding obtained by each sector can be observed. However, the Electric sector emerges as the largest recipient in our dataset, having secured $145 billion in bond issuance and $14 billion in loans.

Total Bond and Loan Financing 2017-2024 (US$) by sector of recipient

Graph 3



On the other side of the transaction, we have identified the banks that have facilitated the issuance of syndicated loans or underwriting corporate bond issuance. Table 1 provides information on the largest top 10 banks that have supported green finance to NFCs. As can be seen, BNP Paribas is a prominent bank providing a total of $53.77 billion in green loans and underwriting green bonds, between 2014-2024.

Ten largest banks by total green financing

Table 1

Bank Name	Bank Country	Syndicated Loans	Bond Issuance	Total Green Financing
BNP Paribas	France	8.22	45.55	53.77
UniCredit (Germany)	Germany	11.99	26.21	38.19
JPMorgan Chase	US	4.49	29.98	34.48
CACIB	France	20.22	12.28	32.50
HSBC	UK	2.36	28.91	31.28
ING	Netherlands	12.20	18.69	30.90
Société Générale	France	7.66	22.63	30.29
Santander	Spain	17.71	10.14	27.85
Mitsubishi UFJ (London)	UK	24.68	0.30	24.99
NATIX	France	13.52	11.04	24.56

Note: Largest banks by total amount of green financed issued for the period between 2014-2024. Mitsubishi UFJ is a subsidiary located in London and does not reflect transactions from the headquarters. Similarly, UniCredit is an Italian-owned bank, but the transactions recorded in our dataset reflect the German-based subsidiary only. All values are in US $ in billion.

4. Methodology

We estimate two models to evaluate the demand side of the green finance market. First, we estimate the following baseline model that examines the factors determining the total amount of green financing received by NFCs:

$$log\ (totalgreenfinancing_{i,j,t,c}) = \alpha + Y_{i,j,t}\beta + \gamma_j + \theta_t + \delta_{c,t} + \epsilon_{i,j,t,c} \quad (1)$$

Where, the dependent variable is the logarithm of the total amount of green finance, including both use of proceeds loans and bonds, obtained by firm *i* in sector *j,* during year *t,* and in country *c*. $Y_{i,j,t}$ is a vector of firm-level balance sheet variables such as leverage, profitability, liquidity, size of the firm, and the cost of finance for firm i and time t with corresponding coefficient vector β. γ_j denotes sector fixed effects for sector *j,* in which firm *i* operates and θ_t denotes time-fixed effects for year *t*. $\delta_{c,t}$ represents country-level macroeconomic variables such as GDP per capita, inflation, and investment for country *c* at time *t*. A complete list of variables and their definitions is provided in Appendix I. [7]

Next, we estimate a binominal logistic regression model aimed at explaining the determinants of the type of green finance an NFC has secured, rather than the

[7] The cost of finance variable consists of the lending rate in the case of loans and the coupon rate where green financing was raised through bond issuance. The coupon rate reflects the cost of finance to the firm at the point of bond issuance, but perhaps not in future periods where interest rates change.

amount[8]. The model considers whether a firm received a bond or a loan, allows us to identify the significant factors influencing NFCs decision-making and has the following form:

$$Indicator_{itb} = \alpha + \beta Y_{i,t,b} + \delta_{c,t} + \gamma_j + \theta_t + \mu_i + \varepsilon_{i,t,b} \quad (2)$$

Here $Indicator_{i,t,b}$ is a binary variable which takes the value of 1 if a loan was issued to firm *i* at time *t*, and by bank *b*, and 0 otherwise to denote the issuance of a bond. $Y_{i,t,b}$ is a vector of bank balance sheet variables. $\delta_{c,t}$ represents country-level macroeconomic variables. θ_t denotes time-fixed effects for year *t*. μ_i represents an indicator variable to denote if the recipient firm is in a green or a brown industry. The Bloomberg data provides information on the industry in which the NFCs operate allowing us to distinguish between green and brown sectors. A key barrier here is the lack of a standardised approach to delineating between green activity and brown activity at the sector level (Shapira *et al.*, 2014). The approach taken in this paper is similar to Ardia *et al.* (2023) and He *et al.* (2024), where the delineation is made based on the sector's tendency to pollute in the course of the production or consumption of the good/service.

We proceed in a similar fashion and estimate the baseline model (1) on the supply side of the green finance market. Specifically, we include bank balance sheet variables and bank-level ESG scores to analyse the factors influencing the amount of green finance provided by a bank in a given year as outlined in the following baseline model:

$$\log (totalgreenfinancing_{i,t,b}) = \alpha + \beta_1 Y_{b,t} + \beta_2 \rho_{b,t} + \lambda_b + \theta_t + \varepsilon_{b,t} \quad (1.1)$$

Where the dependent variable is the natural log of the total amount of green finance a bank b issued to firm i in year t. $Y_{b,t}$ is a vector of bank balance sheet variables such as leverage, profitability, liquidity and loan growth rate. $\rho_{b,t}$ is a vector of bank-level ESG indicators which we consider sequentially. They comprise the Bloomberg environmental score, overall ESG score, ESG disclosure score and climate annual report, which reflects the discussion of climate change risk in the bank's b annual report at time t. λ_b denotes bank-level fixed effects for bank b. θ_t denotes time-fixed effects for year t.

We also estimate a binomial logistic regression model to assess the factors which determine the form of finance a bank is willing or able to offer and the factors which explain the bank's likelihood of issuing loans rather than bonds (and vice versa). The model takes the following form:

$$Indicator_{itb} = \alpha + \beta_1 Y_{b,t} + \beta_2 \rho_{b,t} + \lambda_b + \theta_t + \varepsilon_{b,t} \quad (2.1)$$

[8] The model is estimated using information on transactions which occurred and cannot reflect the demand from NFCs which was not realized.

For which the $Indicator_{i,t,b}$ is a binary variable to denote the issuance of a bond (coded as 0) or a loan made to firm *i*, at time *t*, and by bank *b*. $Y_{b,t}$ is a vector of bank balance sheet variables. $\rho_{b,t}$ is a bank-level ESG indicator at time *t*. λ_b denotes bank-level fixed effects for bank *b*. θ_t denotes time-fixed effects for year *t*.

5. Results

Table 2 presents the results of the baseline model (1), which controls for NFCs' balance sheet variables along with macroeconomic factors. The findings indicate that balance sheet variables play a significant role in determining the amount of green finance which NFCs obtain, as shown in specifications (1) and (2). Higher leverage, greater liquidity, and larger firms are associated with obtaining a higher amount of green finance. Profitability, on the other hand, does not appear to be a significant factor in explaining the amount of green finance, except when accounting for country fixed effect, as shown in specification (4). The results also provide some evidence that macroeconomic variables play a role in determining the amount of green finance raised by NFCs. Interestingly, the results indicate a negative relationship between country-level investment and total green finance, which may suggest that green finance is subject to a crowding-out effect. The results in the specification (3) suggest significant sector-level variations within our dataset. The last two specifications in Table 2 present the results for brown and green firms. For green firms, as shown in specification (5), firm size and the cost of finance do not appear to be significant factors in determining the total amount of green finance firms obtain, although other balance sheet variables remain significant. In contrast, specification (6), reveals that for brown firms, firm size and the cost of finance are the primary determinant of the total amount of green finance.

We extend model (1) to include the maturity of the bonds issued during the period in our dataset. Due to data limitations, the estimation only takes into account the bond issuance for which the results are presented in Table A2, Appendix II. The maturity of the bond appears to have a positive effect on brown firms and does not significantly explain the finance raised from bond issuance for green firms.

Baseline regressions using NFCs balance sheet variables

Table 2

Dependant variable	*Log Total Green Financing*					
Sample	All	All	All	All	Green	Brown
	(1)	(2)	(3)	(4)	(5)	(6)
Leverage	0.003*	0.004**	0.003	-0.003	0.006*	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Profitability	-0.000	-0.000	0.000	0.001***	-0.001	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.002)	(0.001)
Liquidity	0.184***	0.211***	0.155***	0.133***	0.253***	0.077
	(0.048)	(0.047)	(0.042)	(0.040)	(0.087)	(0.057)
Size	0.106***	0.113***	0.143***	0.136***	0.032	0.174***
	(0.020)	(0.019)	(0.018)	(0.017)	(0.029)	(0.026)
Cost of Finance	-0.038**	-0.076***	-0.059***	-0.001	-0.015	-0.045**
	(0.017)	(0.018)	(0.019)	(0.019)	(0.024)	(0.022)
GDP	-0.000***	-0.000**	-0.000***	0.000	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Inflation	0.010*	-0.000	0.002	0.021***	-0.009	0.019**
	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)	(0.007)
Investment	-0.058***	-0.072***	-0.063***	0.055**	-0.099***	-0.016
	(0.012)	(0.012)	(0.012)	(0.025)	(0.021)	(0.015)
Constant	19.216***	19.206***	18.708***	16.799***	20.350***	18.108***
	(0.350)	(0.592)	(0.562)	(0.760)	(0.639)	(0.413)
Observations	1,188	1,188	1,187	1,188	430	757
R-squared	0.102	0.157	0.258	0.376	0.187	0.097
Time FE	No	Yes	Yes	Yes	No	No
Country FE	No	No	No	Yes	No	No
Sector FE	No	No	Yes	No	No	No

Notes: All specifications are estimated using OLS regressions. The dependent variable is the log of the total amount of green finance, including both the use of proceeds loans and bond issuance. Robust standard errors are in parentheses. The statistical significance of results is indicated by ***,**, and * and refer to 1%, 5% and 10% levels, respectively.

Table 3 presents the results of the binomial logistic regression model (2). Specifications (1) and (2) show that profitability influences green finance preferences but not the amounts, as shown in the result of Table 2. Specifically, more profitable firms tend to prefer issuing bonds to finance their green projects. Additionally, larger firms favour bonds over loans, and a higher cost of finance shifts their preference toward bond issuance. In determining the financing choices available to firms, balance sheet variables are slightly more significant for green firms, as shown in specification (3). In particular, firms with greater liquidity tend to prefer raising green finance by issuing bonds. For brown firms, however, only profitability is marginally significant. Regarding macroeconomic conditions, investment stands out as a key difference between the two sectors in that it is a significant factor for brown firms but not green ones.

Binominal Logistic regression

Table 3

Dependant variable	*Green loan =1, bond =0*			
Sample	All	All	Green	Brown
	(1)	(2)	(3)	(4)
Leverage	0.002	0.006	0.007	0.005
	(0.004)	(0.004)	(0.008)	(0.005)
Profitability	-0.005**	-0.003*	-0.006*	-0.002*
	(0.002)	(0.002)	(0.003)	(0.001)
Liquidity	-0.014	0.095	0.466***	-0.178
	(0.089)	(0.095)	(0.151)	(0.126)
Size	-0.241***	-0.258***	-0.207***	-0.280***
	(0.036)	(0.037)	(0.065)	(0.046)
Cost of Finance	-0.399***	-0.589***	-0.660***	-0.545***
	(0.039)	(0.055)	(0.106)	(0.071)
GDP	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Inflation	-0.003	-0.012	-0.034	-0.021
	(0.012)	(0.014)	(0.037)	(0.018)
Investment	-0.123***	-0.128***	-0.054	-0.168***
	(0.026)	(0.029)	(0.040)	(0.041)
Constant	5.088***	4.743***	4.956***	7.679***
	(0.710)	(1.201)	(1.391)	(1.352)
Observations	1,192	1,192	426	761
Time FE	No	Yes	Yes	Yes
ROC	0.750	0.786	0.813	0.797

Notes: All specifications are estimated using a logistic regression model. The dependent variable is a dummy variable taking the value of 1 for use of proceeds loans and 0 otherwise. Robust standard errors in parentheses. The statistical significance of results is indicated by ***,**, and * and refer to 1%, 5% and 10% levels, respectively.

Next, we present the results of the regression model that examines the factors influencing banks' loan and bond issuance to NFCs. Table 4 displays the findings from the baseline regression model (1.1). Similar to the analysis of NFCs discussed above, balance sheet variables appear to play a significant role in determining the total amount of green finance issued by banks during our sample period. Specifically, bank liquidity and size are positively associated with the amount of green finance issues, as indicated in specification (1). However, banks with higher profitability and loan growth rates show a negative relationship with green finance issuance. This could be because banks may reduce green financing to NFCs during periods of strong profitability and rapid loan growth rates. These results remain consistent even after accounting for bank and time-fixed effects, as shown in specifications (2) and (3). The results in Table 4 also show that all environmental performance measures, such as the Bloomberg environmental score, ESG disclosure score and the indicator associated

with the annual reports which discuss climate change, are positive and significant in all specifications.

Baseline regression with bank balance sheet variables

Table 4

Dependant variable	*Log Total Green Financing*					
	(1)	(2)	(3)	(4)	(5)	(6)
Leverage	0.000	-0.000	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Profitability	-0.255***	-0.074***	-0.265***	-0.283***	-0.212***	-0.263***
	(0.024)	(0.017)	(0.018)	(0.024)	(0.019)	(0.021)
Liquidity	0.704**	1.182*	0.551	0.842*	0.821*	0.587*
	(0.343)	(0.688)	(0.337)	(0.433)	(0.427)	(0.321)
Size	0.645***	2.987***	0.582***	0.564***	0.419***	0.570***
	(0.055)	(0.357)	(0.049)	(0.061)	(0.055)	(0.044)
Loan growth rate	-0.040***	0.031**	-0.100***	-0.017	-0.031***	-0.057***
	(0.011)	(0.012)	(0.011)	(0.010)	(0.009)	(0.010)
Environmental score	0.141***	0.311***	0.052***			
	(0.037)	(0.028)	(0.019)			
Overall ESG score				0.373***		
				(0.047)		
ESG disclosure score					0.042***	
					(0.004)	
Climate annual report						0.314***
						(0.072)
Constant	11.733***	-21.736***	12.971***	11.799***	12.997***	13.112***
	(0.675)	(4.885)	(0.639)	(0.710)	(0.623)	(0.574)
Observations	1,573	1,573	1,573	1,565	1,699	1,802
R-squared	0.208	0.826	0.312	0.219	0.221	0.168
Time FE	No	No	Yes	No	No	No
Bank FE	No	Yes	No	No	No	No

Notes: All specifications are estimated using OLS regressions. The dependent variable is the log of the total amount of green finance, including both the use of proceeds loans and bond issuance. Robust standard errors are in parentheses. The statistical significance of results is indicated by ***,**, and * and refer to 1%, 5% and 10% levels, respectively.

Regarding banks' likelihood of the type of green finance issuance, Table 5 shows that most balance variables do not significantly explain their activity. The only notable exception is bank size, which appears to be significant, as shown in specification (1). The results suggest that the larger a bank is the higher the likelihood to issue green bonds over green loans. Additionally, a bank with higher environmental scores also has a tendency towards bond issuance. The results are robust in all three specifications presented in Table 5.

Binominal Logistic regression

Table 5

Dependant variable	*Green loan =1, bond =0*		
	(1)	(2)	(3)
Leverage	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)
Profitability	-0.138	-0.179	0.819
	(0.125)	(0.185)	(0.564)
Liquidity	0.280	0.404	-2.830
	(1.408)	(1.445)	(1.835)
Size	-0.651***	-0.579***	-5.639***
	(0.166)	(0.150)	(1.786)
Loan growth rate	-0.052	-0.074	0.067
	(0.053)	(0.055)	(0.075)
Environmental score	-0.186***	-0.082***	-0.253**
	(0.051)	(0.026)	(0.100)
Constant	9.581***	10.818***	81.112***
	(1.925)	(2.037)	(24.973)
Observations	1,575	1,575	1,302
Time FE	No	Yes	No
Bank FE	No	No	Yes
ROC	0.713	0.734	0.911

Notes: All specifications are estimated using a logistic regression model. The dependent variable is a dummy variable taking the value of 1 for use of proceeds loans and 0 otherwise. Robust standard errors in parentheses. The statistical significance of results is indicated by ***,**, and * and refer to 1%, 5% and 10% levels, respectively.

6. Conclusion and Policy Recommendations

Using transaction-level data on both banks and NFCs, this paper empirically examines the factors influencing the amount of green finance and the choice between use of proceeds loans or bond issuance. The findings highlight that firm balance sheets play a significant role in determining both the quantity and type of green finance obtained. Additionally, we observe distinct differences between brown and green firms in their green finance behaviour. For brown firms, balance sheet variables such as liquidity, profitability and leverage do not significantly impact the total green finance raised. Instead, factors like the cost of finance and firm size are crucial. For brown firms, higher interest or coupon rates reduce the amount of green finance they raise. On the other hand, larger firms tend to secure more green financing.

In contrast, for green firms, the cost of finance is not a key determinant, but liquidity and leverage significantly influence the amount of green finance obtained. Regarding the choice of financing type, profitability drives a preference for bond

issuance over loans. Other balance sheet factors, such as liquidity, also matter more for green firms, with more liquid green firms favouring bonds over loans.

From a bank's perspective, our findings suggest that as they become larger, more profitable or improve their own ESG performance the likelihood of banks in bond issuance increases as opposed to loan issuance. On the other hand, we also find that smaller firms are potentially at a disadvantage in relation to larger firms due to their preference for loans rather than bonds. This potential mismatch between what firms prefer and what banks provide could lead to slowing down the green transition whereby firms are unable to access suitable funds for green projects.

Our findings lead to wide-ranging policy implications. The results presented in this paper reveal that common factors influencing green finance for both green and brown firms include the cost of finance and their size. The implications of the results suggest that the cost of financing can significantly impact a firm's ability to pursue green projects. If green finance is available at lower costs, it could increase the green projects small firms could undertake and hence increase their overall contribution towards the transition to greener economies. Therefore, it is crucial to consider the needs and challenges of smaller firms which are more exposed to the cost of finance given their size predisposes them to the loan market rather than bonds. Furthermore, a lower cost of finance compared to traditional finance could further enhance green projects all-size firms undertake.

References:

Altunbaş, Y., Kara, A. and Marques-Ibanez, D. (2010) 'Large debt financing: syndicated loans versus corporate bonds', *The European Journal of Finance*, 16(5), pp. 437–458.

Ardia, D. *et al.* (2023) 'Climate Change Concerns and the Performance of Green vs. Brown Stocks', *Management Science*, 69(12), pp. 7607–7632.

Bloomberg (2023) *Green, Social, Sustainability & Sustainability linked Instruments Guide*. Bloomberg.

Del Gaudio, B.L. *et al.* (2022) 'Syndicated green lending and lead bank performance', *Journal of International Financial Management & Accounting*, 33(3), pp. 412–427.

Delis, M.D., De Greiff, K. and Ongena, S. (2023) 'Being stranded on the carbon bubble? Climate policy risk and the pricing of bank loans', Swiss Finance Institute Research Paper No.18-10.

Filipava, L. and Murshudli, F. (2023) 'The Development of the Global Green Finance Market: The Role of Banks and Non-banking Institutional Investors', in N. Naifar and A. Elsayed (eds) *Green Finance Instruments, FinTech, and Investment Strategies*. Cham: Springer International Publishing, pp. 27–46.

Giannetti, M., Jasova, M., Loumioti, M. and Mendicino, C. (2023) 'Glossy green" banks: the disconnect between environmental disclosures and lending activities', European Central Bank, Working Paper Series No 2882.

He, W. *et al.* (2024) 'Dynamic spillovers of green, brown, and financial industries under the low-carbon transition: Evidence from China', *Energy Economics*, 139, p. 107901.

Nauman, B., and Morris, S., (2021). Global banks' $750bn in fossil fuels finance conflicts with green pledges, The Financial Times, available on: https://www.ft.com/content/c1e31c6f-6319-4bfc-bde3-3ace80b46a2b

Reghezza, A., Y. Altunbas, D. Marques-Ibanez, C. R. d'Acri, and M. Spaggiari (2022). Do Banks Fuel Climate Change? *Journal of Financial Stability*, Vol.62, 101049.

Shapira, P. *et al.* (2014) 'Probing "green" industry enterprises in the UK: A new identification approach', *Technological Forecasting and Social Change*, 85, pp. 93–104.

Appendix

Appendix I- Variable Definitions

Table A1

Variable	Definition
Balance Sheet Variables	
Log Total Green Financing	The total amount of finance raised by an NFC in an individual financing transaction. *Value is taken in natural logarithms*
Bond/Loan Indicator	An Indicator variable to denote if an individual financing transaction allowed a firm to obtain a bond issuance or obtain a syndicated term loan. 0 denotes a bond.
Leverage	*Debt-to-Capital Ratio*
Profitability	*EBITDA Margin*
Current Ratio	*Current Ratio*
Size	Natural Logarithm of Total Assets
Cost of Finance	The coupon rate of the individual bond transaction (Bloomberg Data) or the lending rate at the country level (World Bank WDI) in year *t* for firm *i* located in country *c*.
Maturity	Number of years for which a bond has been issued (Calculated based upon Bloomberg Data)
Green/Brown Sector Indicator	An indicator variable where 0 denotes an NFC in a Brown industry
Bank Balance Sheet Variables	
Log Total Green Financing	The total amount of green finance raised by given bank *Value is taken in natural logarithms.*
Loan Growth Rate	(Total Loans$_t$ – Total Loans$_{t-1}$) / Total Loans$_{t-1}$
Size	Natural Logarithm of Total Assets
Liquidity	(Cash & Cash Equivalents + Inter-banking Assets + Trading Securities + Assets Available for Sale)/Total Assets
Profitability	Return on Total Assets - Ratio
Leverage	Average Assets/Average – Equity
ESG Variables	
Bank Balance Sheet Variables	
Overall ESG Score	Bloomberg ESG Materiality Scorecard Indications Values between 0-10, 10 is best
Environmental Score	Bloomberg ESG Materiality Scorecard Indications Values between 0-10, 10 is best
Bloomberg ESG Disclosure Score	Values between 0-100, 100 is best
Annual Report Discusses Opportunities and Risks w.r.t Climate Change	Indicator variable, where 1 denotes the annual report discussing Climate Change
Macroeconomic Control Variables	
GDP	Nominal GDP per Capita (World Bank WDI)
Inflation	Annual GDP Deflator (World Bank WDI)
Investment	Gross Fixed Capital Formation (World Bank WDI)

Appendix II- Supplementary findings

Table A2 presents the regression estimates for the baseline model (1) on green bond financing. Due to data limitations, we could only obtain information on the maturity of green finance for NFCs in the context of bond issuances. As a result, model (1) is estimated using data exclusively from bond issuances to account for bond maturity. The results suggest that higher profitability reduces the amount of green bond finance obtained by firms, possibly indicating that firms may prefer to use retained earnings to fund green projects. However, this finding applies only to green firms, as shown in specification (2) of Table A2. In contrast, bond maturity affects only brown firms, with longer bond maturities associated with higher amounts of green bond issuance by firms.

Green bond financing

Table A2

Dependant variable	*Log Green Bond Financing*		
Sample	All	Green	Brown
	(1)	(2)	(3)
Leverage	0.004	0.004	0.003
	(0.003)	(0.004)	(0.004)
Profitability	-0.006***	-0.005***	-0.002
	(0.001)	(0.002)	(0.002)
Liquidity	0.186***	0.163	0.124**
	(0.055)	(0.110)	(0.059)
Size	0.076***	0.057*	0.114***
	(0.025)	(0.033)	(0.035)
Cost of Finance	-0.079***	-0.021	-0.111***
	(0.022)	(0.028)	(0.039)
Maturity	0.001*	0.005	0.001**
	(0.001)	(0.014)	(0.001)
GDP	-0.000***	0.057*	-0.000**
	(0.000)	(0.033)	(0.000)
Inflation	0.002	-0.021	0.007
	(0.006)	(0.028)	(0.007)
Investment	-0.103***	0.005	-0.072***
	(0.015)	(0.014)	(0.018)
Constant	21.252***	21.650***	20.507***
	(0.422)	(0.898)	(0.500)
Observations	762	296	465
Time FE	0.391	0.407	0.377

Notes: The dependent variable is the log of the amount of bond green financing. Robust standard errors are in parentheses. The statistical significance of results is indicated by ***,**, and * and refer to 1% 5% and 10% levels, respectively.

Green financing and the relationship between banks and non-financial corporations through the lens of balance-sheet interaction



Michael Harrison and Mimoza Shabani
University of East London

New Insights from Financial Statements Conference
17th of October 2024
Banco de España

University of East London | 125 Years of Pioneering Futures

Motivation

- Recently, significant attention has focused on the economic impact of climate change with the aim achieve a greener economy
- Banks are key players in the transition, as they play a fundamental role in providing green finance to non-financial corporations (NFC) that need to fund investments (Del Gaudio *et al.*, 2022; Filipava and Murshudli, 2023).
- A growing literature examines banks' engagement in green financing but very little is known about the determinants of green lending behaviour
- What about the role of NFC and what are the (internal) determinants of raising green finance?

Contribution

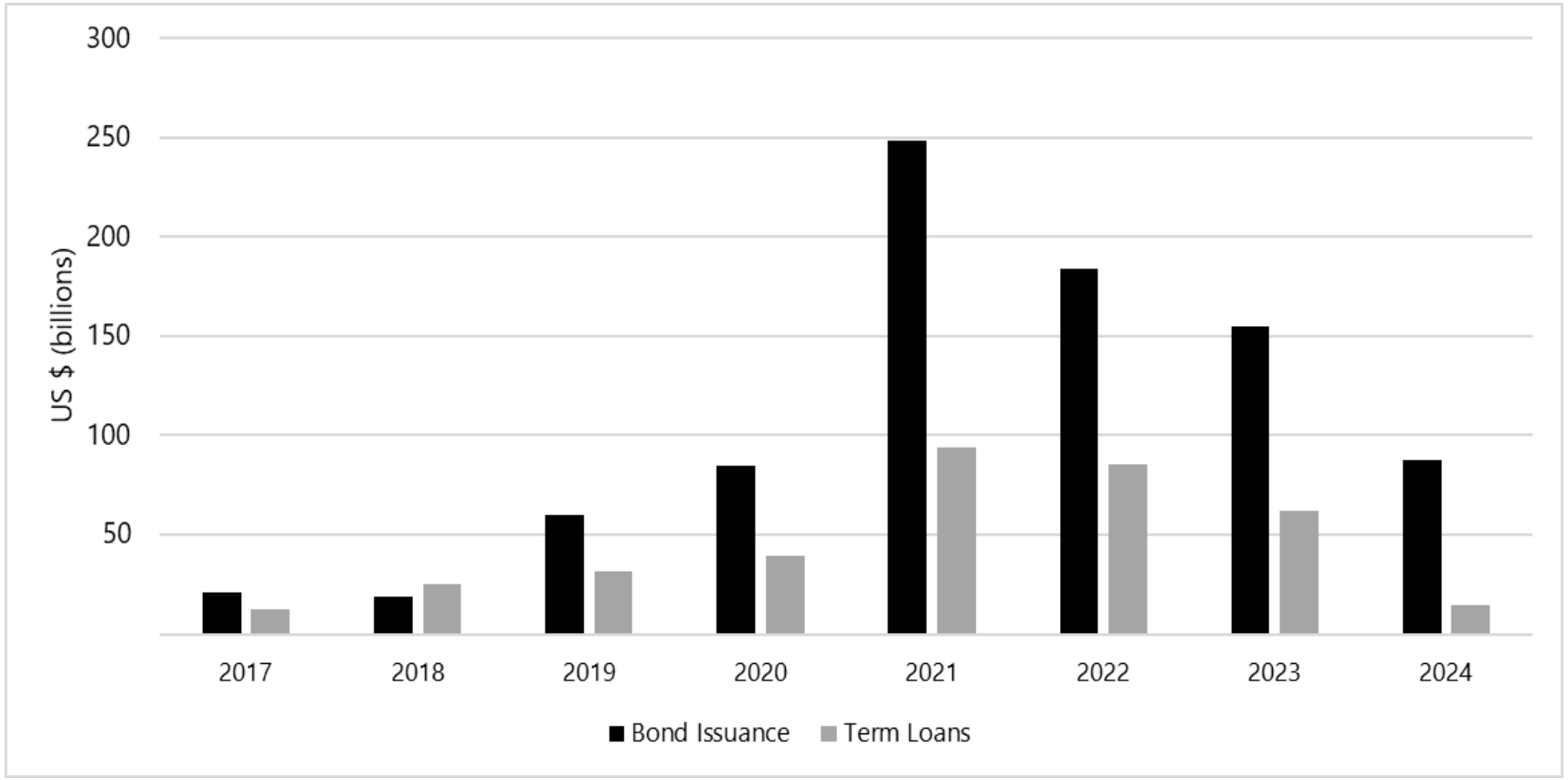
- Our goal is to identify the determinants of NFC's preferences for green finance through green corporate bonds or the use of proceeds loans
- We employ a binomial logistic regression model that will shed light and quantify the factors determining a firm's preference for green financing through bonds or loans
- We also estimate the likelihood of banks issuing loans or/or underwriting bonds taking into account bank-level balance sheet factors
- Using data for 2297 green bonds and 1968 use of proceeds term loans we link banks which have issued loans and have underwritten bonds to/for corresponding NFC.
- The granular data allows us to examine the internal factors on both sides of the market.
- This deeper understanding could guide policymakers to tailor policy to support green finance provision

Dataset and ESG Screening

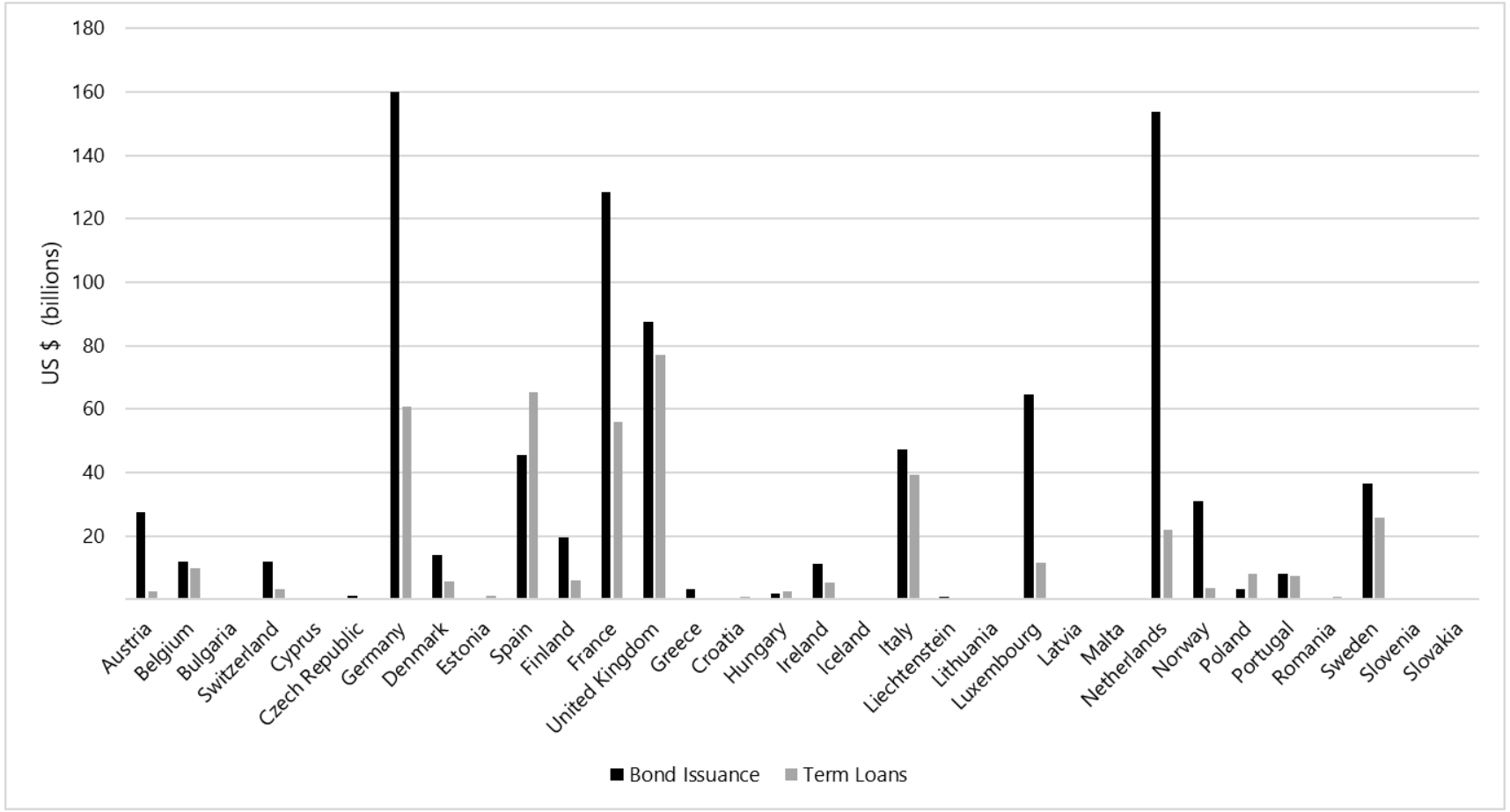
- 4265 individual financing transactions between 2017 and 2024(Q2) identified via Bloomberg LP *SRCH Function*
  - 2297 Bond Issuances
  - 1968 Syndicated Loan Issuances
- Bloomberg screening criteria consistently applied - verified green use of proceeds
  - All transactions considered had an ESG verification partner involved
  - Minimal risk of including greenwashed transactions in our dataset

- Countries Included: EEA, UK and Switzerland
  - Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, United Kingdom, Switzerland, Iceland, Liechtenstein and Norway

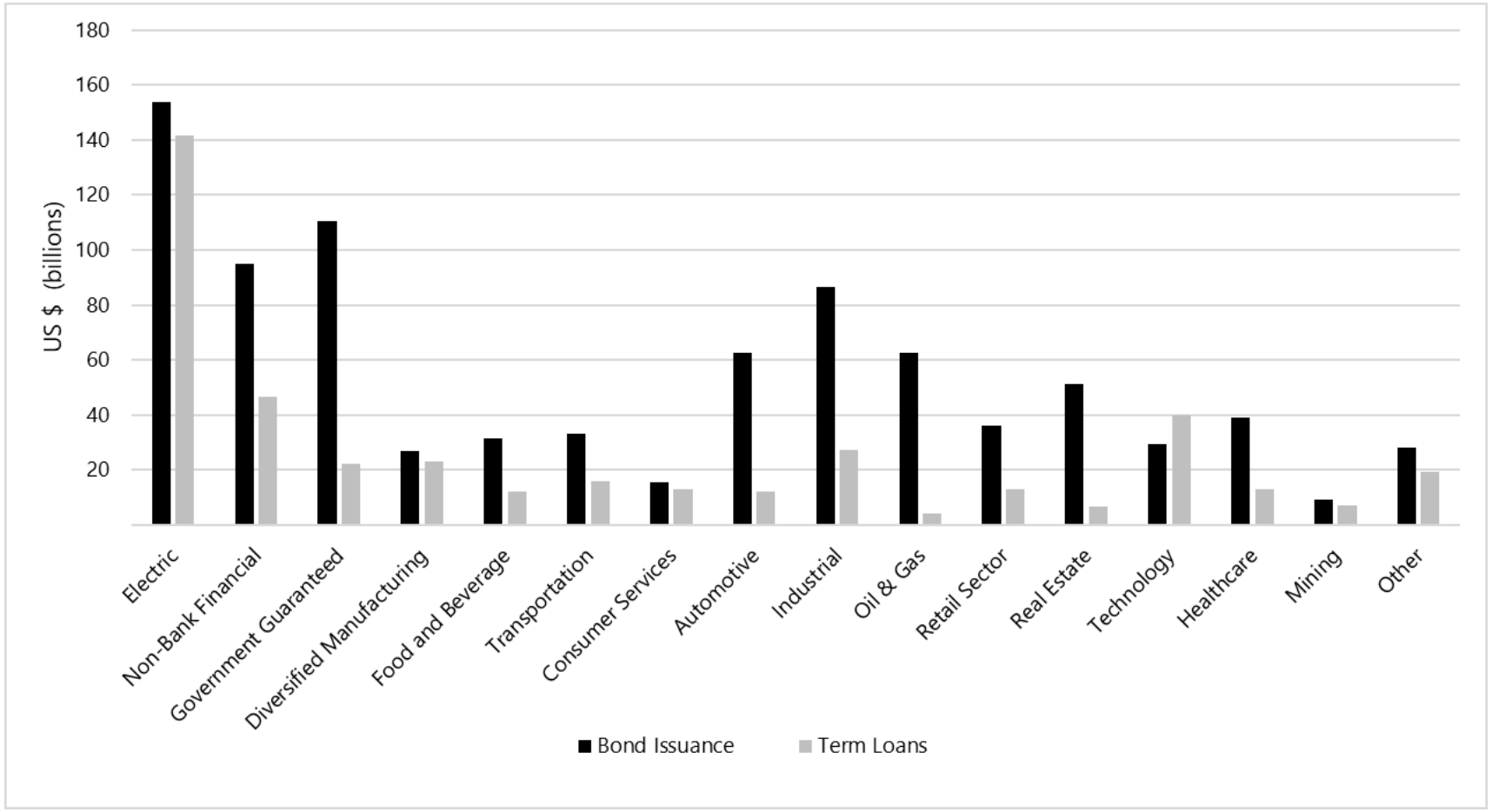
Total bond and loan financing 2017-2024



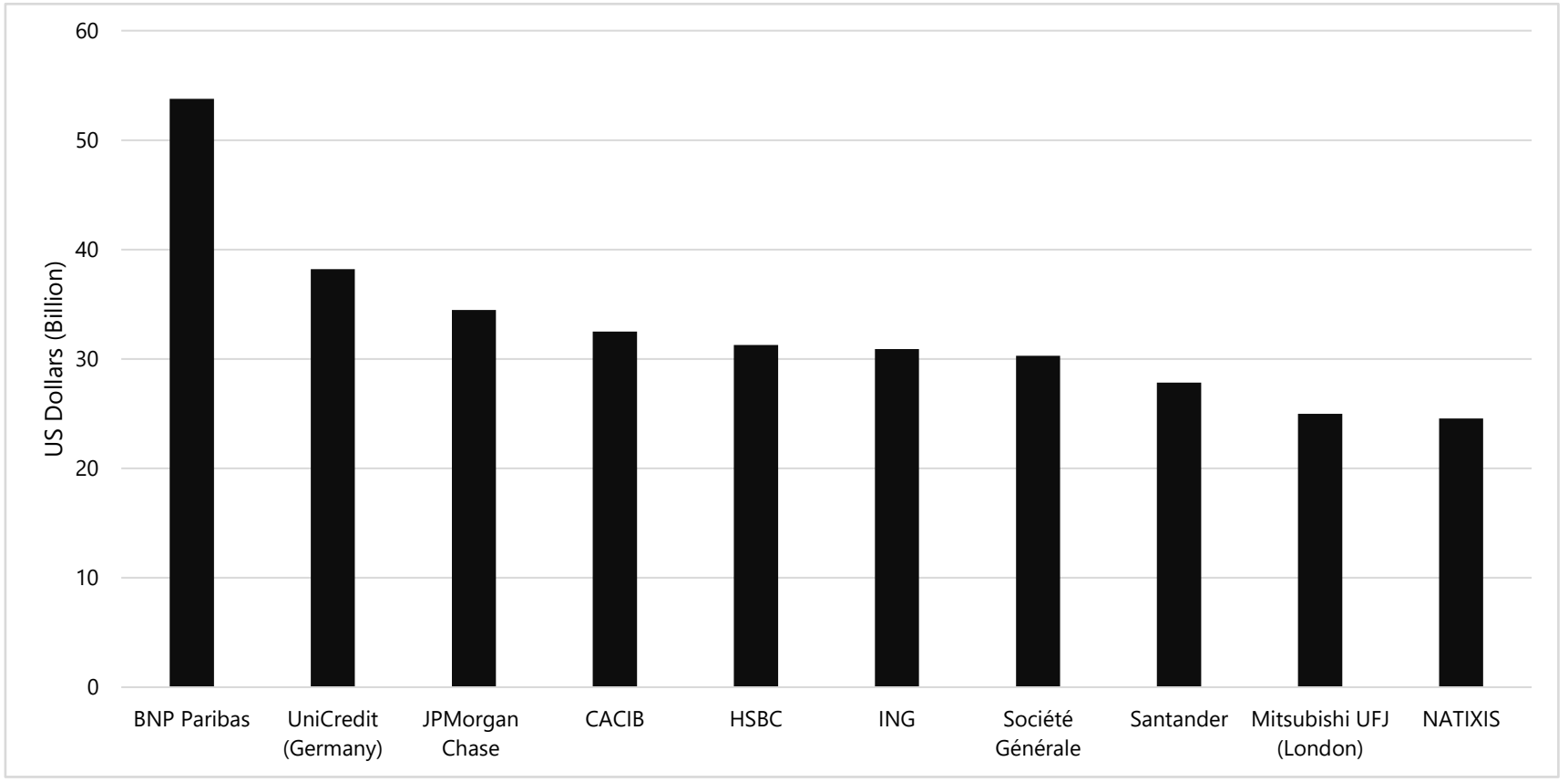
Total Green Financing sorted by type and domicile of recipient



Total Bond and Loan Financing 2017-2024 (US$) by sector of recipient



Ten largest banks by total green financing



Explanatory Variables Collected

Firm Level Balance Sheet Data (From Bloomberg FA Function)	Bank Balance Sheet Variables (From Bloomberg FA Function)
• Financial Leverage – Debt-to-Capital Ratio • Profitability – EBITDA Margin • Liquidity – Current Ratio • Market-to-Book Value • Size – Natural Log of Total Assets • Cost of Finance – Coupon Rate (Bonds) or Lending Rage (Loans) • Industry Classification • Brown/Green Sector Indicator	• Loan Growth Rate – Year on year change on total loans issued • Size – Natural Log of Total Assets • Liquidity – Liquid Assets / Total Assets Profitability – Return on Total Assets Ratio • Leverage - Average Assets/Average Equity
Bank Level ESG Variables (From Bloomberg FA Function)	**Macroeconomic Control Variables** (World Bank WDI Dataset)
• Overall ESG Score - Bloomberg ESG Materiality Scorecard Indicator • Environmental Score - Bloomberg ESG Materiality Scorecard Indicator • ESG Disclosure Score – Bloomberg Transparency Indicator • Climate Annual Report - Indicator variable, where 1 denotes the annual report discussing Climate Change	• Inflation, GDP deflator (annual %) • Nominal GDP per capita • Investment - Gross fixed capital formation

Baseline results - *Firm Level*

- Leverage and liquidity are significant determinants of the amount of green financing but not profitability
- The rate of borrowing and firm size are strongly significant
- Macro variables also play a role
- Higher country investment levels are associated with lower green financing
- Cross-sectoral and country differences in the amount of green financing

Dependant variable	*Log Total Green Financing*					
Sample	All	All	All	All	Green	Brown
	(1)	(2)	(3)	(4)	(5)	(6)
Leverage	0.003*	0.004**	0.003	-0.003	0.006*	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Profitability	-0.000	-0.000	0.000	0.001***	-0.001	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.002)	(0.001)
Liquidity	0.184***	0.211***	0.155***	0.133***	0.253***	0.077
	(0.048)	(0.047)	(0.042)	(0.040)	(0.087)	(0.057)
Size	0.106***	0.113***	0.143***	0.136***	0.032	0.174***
	(0.020)	(0.019)	(0.018)	(0.017)	(0.029)	(0.026)
Cost of Finance	-0.038**	-0.076***	-0.059***	-0.001	-0.015	-0.045**
	(0.017)	(0.018)	(0.019)	(0.019)	(0.024)	(0.022)
GDP	-0.000***	-0.000**	-0.000***	0.000	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Inflation	0.010*	-0.000	0.002	0.021***	-0.009	0.019**
	(0.006)	(0.006)	(0.007)	(0.007)	(0.009)	(0.007)
Investment	-0.058***	-0.072***	-0.063***	0.055**	-0.099***	-0.016
	(0.012)	(0.012)	(0.012)	(0.025)	(0.021)	(0.015)
Constant	19.216***	19.206***	18.708***	16.799***	20.350***	18.108***
	(0.350)	(0.592)	(0.562)	(0.760)	(0.639)	(0.413)
Observations	1,188	1,188	1,187	1,188	430	757
R-squared	0.102	0.157	0.258	0.376	0.187	0.097
Time FE	No	Yes	Yes	Yes	No	No
Country FE	No	No	No	Yes	No	No
Sector FE	No	No	Yes	No	No	No

Logistic model results

- Firms' balance sheet variables do not influence their preference for bonds or loans
- Except for profitability and size
- Some macro variables are also significant determinants in firms' choice for green finance
- Balance sheet variables differ in explaining the preference between green and brown industries

Dependant variable	*Green loan =1, bond =0*			
Sample	All	All	Green	Brown
	(1)	(2)	(3)	(4)
Leverage	0.002	0.006	0.007	0.005
	(0.004)	(0.004)	(0.008)	(0.005)
Profitability	-0.005**	-0.003*	-0.006*	-0.002*
	(0.002)	(0.002)	(0.003)	(0.001)
Liquidity	-0.014	0.095	0.466***	-0.178
	(0.089)	(0.095)	(0.151)	(0.126)
Size	-0.241***	-0.258***	-0.207***	-0.280***
	(0.036)	(0.037)	(0.065)	(0.046)
Cost of Finance	-0.399***	-0.589***	-0.660***	-0.545***
	(0.039)	(0.055)	(0.106)	(0.071)
GDP	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Inflation	-0.003	-0.012	-0.034	-0.021
	(0.012)	(0.014)	(0.037)	(0.018)
Investment	-0.123***	-0.128***	-0.054	-0.168***
	(0.026)	(0.029)	(0.040)	(0.041)
Constant	5.088***	4.743***	4.956***	7.679***
	(0.710)	(1.201)	(1.391)	(1.352)
Observations	1,192	1,192	426	761
Time FE	No	Yes	Yes	Yes
ROC	0.750	0.786	0.813	0.797

Baseline results - *Bank Level*

- Regarding banks, profitability, liquidity and previous lending positively relate to the amount of green financing.
  - However, bank leverage is not a significant factor
- Larger banks lend more
- All ESG variables considered indicate that banks with better environmental performance and transparency engage more in green lending

Dependant variable	*Log Total Green Financing*					
	(1)	(2)	(3)	(4)	(5)	(6)
Leverage	0.000	-0.000	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Profitability	-0.255***	-0.074***	-0.265***	-0.283***	-0.212***	-0.263***
	(0.024)	(0.017)	(0.018)	(0.024)	(0.019)	(0.021)
Liquidity	0.704**	1.182*	0.551	0.842*	0.821*	0.587*
	(0.343)	(0.688)	(0.337)	(0.433)	(0.427)	(0.321)
Size	0.645***	2.987***	0.582***	0.564***	0.419***	0.570***
	(0.055)	(0.357)	(0.049)	(0.061)	(0.055)	(0.044)
Loan growth rate	-0.040***	0.031**	-0.100***	-0.017	-0.031***	-0.057***
	(0.011)	(0.012)	(0.011)	(0.010)	(0.009)	(0.010)
Environmental score	0.141***	0.311***	0.052***			
	(0.037)	(0.028)	(0.019)			
Overall ESG score				0.373***		
				(0.047)		
ESG disclosure score					0.042***	
					(0.004)	
Climate annual report						0.314***
						(0.072)
Constant	11.733***	-21.736***	12.971***	11.799***	12.997***	13.112***
	(0.675)	(4.885)	(0.639)	(0.710)	(0.623)	(0.574)
Observations	1,573	1,573	1,573	1,565	1,699	1,802
R-squared	0.208	0.826	0.312	0.219	0.221	0.168
Time FE	No	No	Yes	No	No	No
Bank FE	No	Yes	No	No	No	No

Bank logistic model

- Balance sheet variables explain the amount of finance issued, but do not explain the likelihood of the type of finance
- Larger banks are more likely to underwrite bonds as opposed to issuing syndicated loans
- Banks with higher ESG scores have a higher likelihood of underwriting green bonds

Dependant variable	*Green loan =1, bond =0*		
	(1)	(2)	(3)
Leverage	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)
Profitability	-0.138	-0.179	0.819
	(0.125)	(0.185)	(0.564)
Liquidity	0.280	0.404	-2.830
	(1.408)	(1.445)	(1.835)
Size	-0.651***	-0.579***	-5.639***
	(0.166)	(0.150)	(1.786)
Loan growth rate	-0.052	-0.074	0.067
	(0.053)	(0.055)	(0.075)
Environmental score	-0.186***	-0.082***	-0.253**
	(0.051)	(0.026)	(0.100)
Constant	9.581***	10.818***	81.112***
	(1.925)	(2.037)	(24.973)
Observations	1,575	1,575	1,302
Time FE	No	Yes	No
Bank FE	No	No	Yes
ROC	0.713	0.734	0.911

Conclusion

- Balance sheet variables are significant in explaining green finance for both firms (receive) and banks (issue)
  - However, for banks the type of finance issued is unrelated to balance sheet factors – only size and ESG factors.
- The determinants of green financing differ among industries

- When considering only green industries' balance sheets and macro variables remain significant but not the same for brown industries
  - The borrowing costs and size of the firm have different effects on brown and green industries

- Common factors for both brown and green industries are the cost of borrowing and the size of the firm, implying needed support for smaller size firms from a policy perspective
- A mismatch between what firms prefer and what banks provide could lead to a slowing down of green transition

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Quality over quantity? Using corporations' climate change related disclosure for risk assessment[1]

Friedrich Bähr, Daniela Gasser and Sabine Wukovits,
Central Bank of the Republic of Austria

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

Quality over Quantity? Using corporation's climate change related disclosure for risk assessment

Daniela Gasser[a], Sabine Wukovits[a], Friedrich Baehr[a]

[a] Oesterreichische Nationalbank[1]

Abstract

Non-financial corporations provide sustainability reports, mainly based on recommendations and market standards to address the increasing demand by financial market participants for climate change risk (CCR) related information. This paper analyses how and to which extent companies disclose CCR information from a credit risk assessment perspective. To study the CCR information in a structured manner, a template was developed and the sustainability reports of 91 Austrian non-financial corporations were analysed. Summing up, the defined analyses aspects risk awareness, actions and target setting are broadly disclosed. However, partial shortcomings in the disclosure quality are identified, also revealing differences between physical risk and transition risk. Notably, companies with higher disclosure quality are more likely to show a downwards trend in their carbon emissions evolution. In sector comparison, the energy sector is a forerunner in CCR disclosure.

Keywords: Climate Change Risk, Transition Risk, Physical Risk

JEL classification: Q54, Q56, Q42

1 Email addresses: Sabine Wukovits (sabine.wukovits@oenb.at), Daniela Gasser (daniela.gasser@oenb.at), Friedrich Baehr (friedrich.baehr@oenb.at)

Any views expressed represent those of the authors only and not necessarily those of Oesterreichische Nationalbank. All remaining errors are those of the authors.

We want to thank Florian Resch and Stefan Löschenbrand for their input and the OeNB ICAS analysts, who have helped to break down the sustainability reports for the data-sample.

Background and Motivation – The need of climate change risks integration into credit risk assessment

The European Commission (EC) published its Action Plan on "Financing Sustainable Growth" in March 2018, following the Paris Agreement on limiting global warming to well below 2°C, preferably below 1.5°C, in the long-term. (European Commission, 2018)

The European Central Bank (ECB) as integral part of the European Union (EU) is contributing to the standardization of regulations as well as to increased clarity concerning the (legal) framework conditions, acknowledging possible effects and dangers for monetary policy resulting from a changing climate. (ECB, 2021) The action plan of the ECB (see (ECB, 2022)) comprises the need to introduce CCR in the in-house credit risk assessment systems (ICAS) of national central banks[2] and banks[3].

The basis for such an assessment is CCR-relevant firm-level information, currently only mandatorily disclosed by companies falling under the Non-financial Reporting Directive (NFRD)[4], which lacks standardisation. However, given the increasing demand for sustainability reporting by financial market participants and supervisors, ever more companies strive to voluntarily provide respective information, mainly based on recommendations and market standards.

Succeeding the NFRD, the Corporate Sustainability Reporting Directive (CSRD) demands a more comprehensive and standardized reporting and specifies sustainability as of financial relevance. (European Parliament; EU Council, 2022) The CSRD is accompanied by a package of European Sustainability Reporting Standards (ESRS) developed by the European Financial Reporting Advisory Group (EFRAG). Thereby, European Sustainability Standards for Climate Change (ESRS E1) are of special interest focusing on reporting aspects for CCR. (EFRAG, 2022)

This paper analyses how and to which extent companies disclose CCR information based on current and prospective disclosure standards to be used for the assessment of CCR in the context of ICASs.

2 Into the so-called Eurosystem Credit Assessment Framework (ECAF) which functions as a quality assurance of mobilised collateral in the Eurosystem credit operations. (Auria, et al., 2021)

3 See Guide on climate-related and environmental risks (europa.eu)

4 The NFRD entered into force in April 2014 and stipulates how financial and listed non-financial corporations with more than 500 employees shall disclose environmental, social and governance topics.

Categorizing a company's sustainability disclosure for climate change risk assessment – methodology and data sample

Description of the template and the analysed criteria

To study the quality of CCR information in a structured manner, a template based on different regulatory and markets standards as well as common reporting practice was developed and the sustainability reports of 91 Austrian non-financial corporations were analysed. Based on this template, the status quo of the companies' CCR reporting is analysed in terms of risk awareness, actions, target setting and emissions. Further it is analysed if there is an interrelation between disclosure quality and emissions' evolution.

Based on the established reporting practice, the template separates between physical and transition risks. According to the adopted single materiality approach for CCR in the context of ICAS only information was collected which potentially has a financial relevance for the company (outside-in perspective). The data collection includes open text questions and categorisations into nominal variables. The assessment of the emissions evolution is based on the historic emissions data reported by the companies and on a calculated emission intensity ratio (scope 1+2 emissions / turnover). Each category is complemented by an evaluation question to verify the quality of the reported information in a harmonized manner. To normalize the evaluation of each category, the outcome is linked to a score value, which is summed up to a final score indicating the quality level of the companies' disclosure. The quality levels are displayed in a range of "very good" to "poor".

Challenges in the filling of the template were mostly related to the open text fields as the required information had to be collected from different parts of the reports, strictly selected regarding the focus of the question, and briefly summarized. The categorizations followed the labelling of the companies, if available, and otherwise was done by the author.

The analysed criteria as well as the evaluation questions for physical and transitions risks are displayed and described in Annex 1: Template description.

Data sample

The sample consists of 91 non-financial companies domiciled in Austria. 39 are listed companies, which mandatorily report under the NFRD, and 52 are groups, which prepare their consolidated statements according to IFRS and voluntarily disclose sustainability information. The analysed reports were published with a financial statement date in a period from 30th June 2022 to 30th September 2023. For the analyses, these firms are classified into the four sectors construction, energy, industry and services, a classification according to the NACE Rev2 4-digit level also used by the ERICA WG[5], belonging to the European Committee of Central Balance Sheet Data Offices (ECCBSO).[6] Distinguishing the different sizes, again based on the classification of the ERICA WG, 17 companies are of small size, 32 are medium and 42 are large.[7]

	Size			
ERICA Sector	Small	Medium	Large	Total
Construction	1	0	3	4
Energy	0	1	8	9
Industry	11	14	20	45
Services	5	17	11	33
Total	17	32	42	91

Table 1: Distribution of companies in the data sample.

5 European Records of IFRS Consolidated Accounts Working Group (ERICA WG)

6 The ECCBSO functions as a consultative body of several national central banks (NCBs) with the goal of supporting central banks' functions by advancing non-financial corporations' analysis. (European Committee of Central Balance Sheet Data Offices (ECCBSO), 2024)

7 The ERICA Working Group defined group sizes based on revenue as follows: small groups < 250 million euros, medium-sized groups 250 million - 1.5 billion euros, and large groups > 1.5 billion euros.

Empirical results

Risk awareness of companies

The level of the risk awareness is measured by the evaluation, whether the companies disclose a comprehensive risk analysis ("Yes"), take a more superficial approach ("Neutral") or do not report on actual or potential climate change risks at all ("No"). In total, out of the 91 assessed companies, a share of around 75% disclose physical and/or transition risks. However, a full risk awareness, meaning a comprehensive and detailed risk description, is only attributed to a smaller fraction of around 43%.

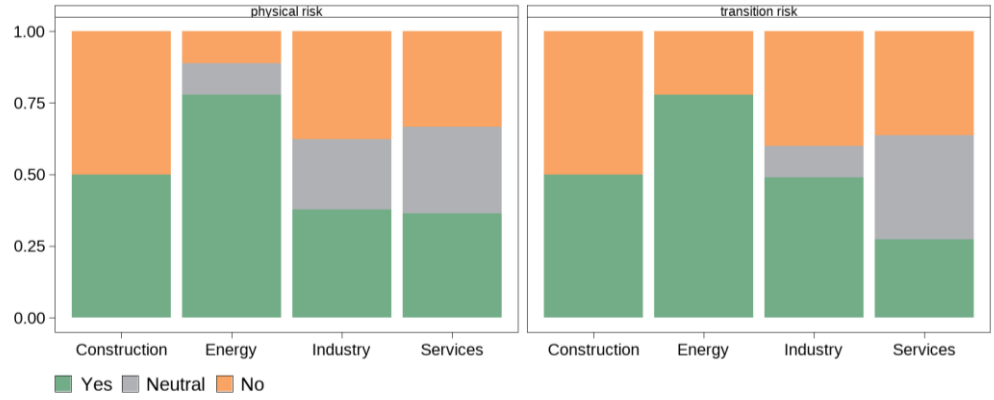


Figure 1 shows the evaluation results separated for physical and transition risks per sectors, revealing that the energy sector takes a leading role, as a comprehensive risk awareness is broadly given in both categories. A less pronounced risk awareness, but at least a certain degree of risk acknowledgment ("Neutral"), is revealed for companies of the industry and services sector. The highest stake of no risk awareness is seen in the construction sector, however, given the small number of companies, these results must be handled with care.

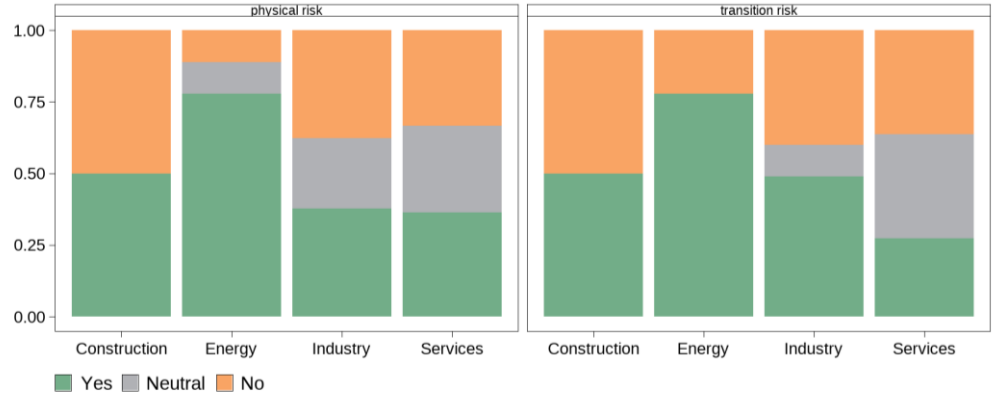


Figure 1: Does the company show a comprehensive risk awareness? Evaluation for physical (plot left) and transition risk (plot right). Note: Share of companies per sector.

The most frequent risk categories in physical risk are water stress (floodings, droughts) and temperature rise (see Figure 2). Companies of the energy sector show the highest frequency in water and temperature, which are related risk categories in

this sector, such as the most common reference is to water scarcity leading to a lower energy production and water scarcity itself is linked to increasing temperatures.

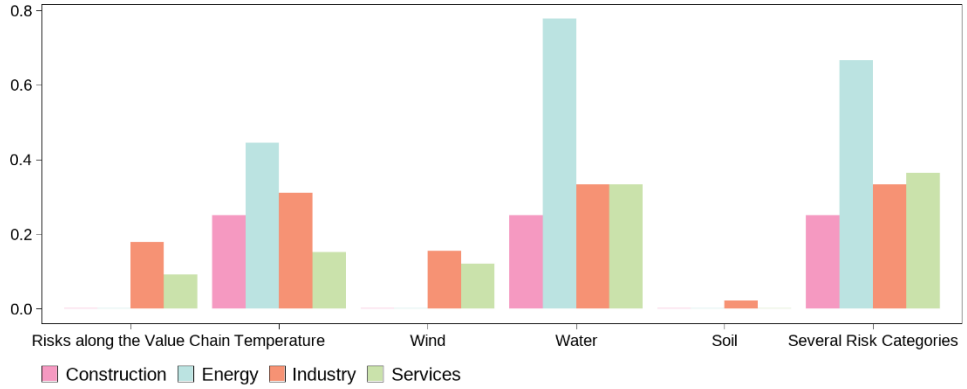


Figure 2: Risk categories of physical risk. Note: Share of companies dealing with physical risk per sector.

Within transition risks, the transmission channel "Legal/Politics" is most frequent, particularly addressed by companies of the construction and energy sector (around 75%, see Figure 3). The risk descriptions mainly refer to regulatory changes connected with the political goals towards CO2 neutrality (energy efficiency / emissions standards) and/or to rise in carbon prices. Changes resulting from the market is particularly relevant for the industry sector. Often the risks are related to rising prices for energy and other input factors of production. Changing customer behaviour leading to a lower demand and to a weaker market position is not only addressed by the industry sector but also by the services sector. Risks accrued from technological developments are particularly pronounced in the energy sector and connected with the risk of stricter energy related regulations, pushing companies to transition from fossil to a renewable energy production. Reputational risks are listed least often across all sectors, in some cases a whole industry is seen confronted with a loss of prestige, e.g. the chemical or textile industry.

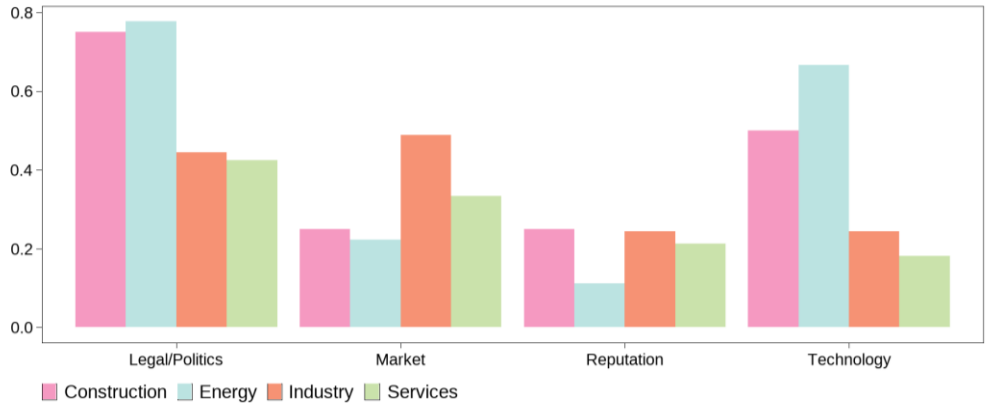


Figure 3: Transmission channels of transition risk. Note: Share of companies dealing with transition risk per sector.

Actions and target setting

The disclosure of actions to address these risks as well as the disclosure of targets along a decarbonisation path are further indicators of the companies' progress in tackling and reporting of CCR.

The assessment of actions separated for physical and transitions risks reveals that the discussion on actions related to physical risks is not as far progressed as for transition risks (see Figure 4).

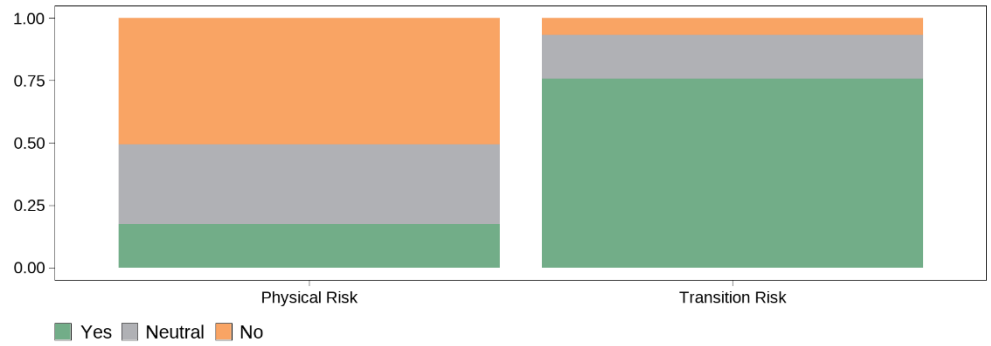


Figure 4: Are the actions in line with the named risks? Note: Share of companies for physical and transition risk.

Only half of the assessed companies reflect on physical risk measures, whereby in the energy sector the share is substantially higher at around two thirds. A fraction of around 30% of the companies show a medium quality regarding their disclosed measures ("Neutral"), and only a minority of 18% is confirmed to disclose comprehensive actions that are in line with the named risks ("Yes"). Physical risk actions are categorized into constructions measures, location measures, process changes and others (see Figure 9 in Annex). Whereas no company plans or takes the action to relocate to an area less endangered by physical risks, construction measures are broadly addressed by all sectors. Process changes are named by a further 20% of the companies.

In contrast, the disclosure of actions to handle transitions risk appears as a common practice, such as more than 75% of all companies publish actions that are in line with the named risks, rising to 100% for companies of the energy sector. However, only a fifth of the actions are precisely defined (i.e. measurable and scheduled). The fraction of measures which are further disclosed together with an implementation plan stands even lower at around 10% of the companies publishing this kind of information.

Published transition actions often comprise a switch to or a higher share of sustainable energy (e.g. purchase of renewable energy resources, purchase/construction of photovoltaic systems) or the increase of energy efficiency. Process changes are particularly in the focus of companies of the energy and construction sector. Actions in "Electrification" or "Fuel switching", mainly referring to the purchase of electric motor vehicles or to use alternative fuels, show a substantial share of mentions across all sectors. Less frequent are measures related to "Material efficiency/consumption reduction" and "Product changes" (see Figure 10 in Annex).

Next to actions, target setting is a crucial disclosure element to tackle the transition to a low-carbon economy (EFRAG, 2022). To assess how far the companies progressed in this reporting field, the disclosure of targets is analysed regarding the timeframe (1-3 years, > 3 years), regarding alignment with the Paris agreement goals and whether the goals are science based (i.e. certified by the Science Based Targets initiative (SBTi)[8]).

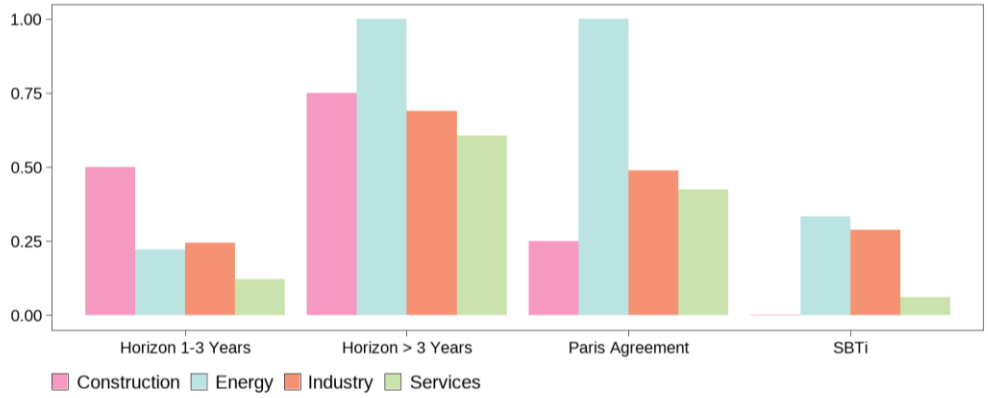


Figure 5: Evaluation of targets. Note: Share of companies per sector.

90% of the analyzed companies disclose at least one target that meets the required criteria, i.e. the target must be measurable and timed, best with reference to a basis year. As shown in Figure 5, goals with a longer horizon than 3 years are clearly more common across all sectors, ranging from 60% in the services sector to 100% in the energy sector. Remarkably, all companies of the energy sector define their goals also in line with the Paris Agreement. Science based targets are disclosed by a third of the energy and industry companies. A full disclosure of target setting, comprising all four defined aspects, is only found in 3 cases. Though incorporated in ESRS E1[9], only 57% of the targets are defined with a clear reference to the emission's scope. Most reported are scope I & 2 references, whereas only a minority of companies also disclose targets for scope 3 emissions.

Summary of the disclosure quality

Figure 6 pictures the disclosure quality (the darker, the better), separating in physical and transition risks, actions as well as transition targets. Generally, the disclosure quality of transition risk and physical risk awareness is balanced. Regarding actions, however, a higher disclosure quality is apparent for transition risk. Acknowledging sector differences, the quality of transition targets broadly meets the disclosure

8 SBTi is a partnership between organizations like CDP, the United Nations Global Compact (UNGC), theWorld Resources Institute (WRI), and the World Wide Fund for Nature (WWF). Its main goal is to help companies to set greenhouse gas (GHG) emission reduction targets to meet the goals of the Paris Agreement. (SBTi, 2024)

9 Once applicable from the financial year 2024 onwards, ESRS E1-4, 35 b specifies that "GHG emission reduction targets shall be disclosed for Scope 1, 2, and 3 GHG emissions, either separately or combined. The undertaking shall specify, in case of combined GHG emission reduction targets, which GHG emission Scopes (1, 2 and/or 3) are covered by the target, the share related to each respective GHG emission Scope and which GHGs are covered.").

quality of actions. It is clearly illustrated, that the energy companies show the highest disclosure quality in all assessed areas. The graph as well visualizes punctual weaker positions, such as the rather low reference to physical actions in the industry sector or to transition targets in the services sector.

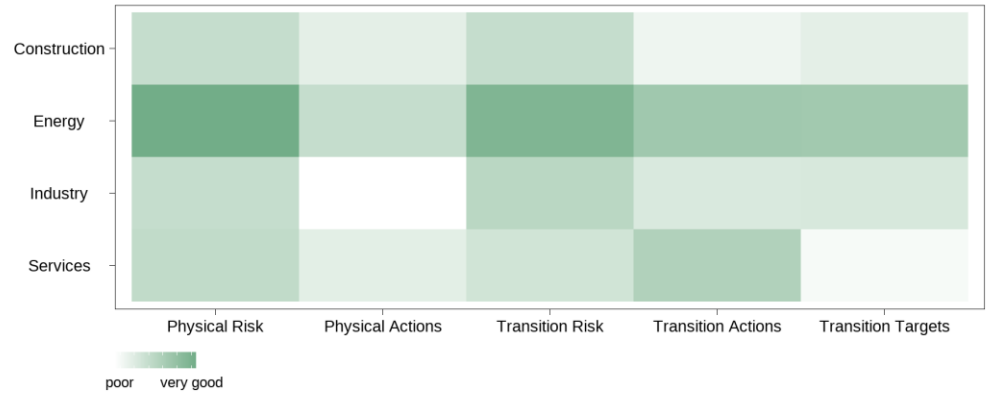


Figure 6: Disclosure quality of physical and transition risk aspects per sector.

Assessment of emissions and relation to disclosure quality

The assessment of the companies' emissions evolution (absolute numbers and emissions intensity[10]) relates to the evaluation question "Does the historic CO2 evolution show a reduction?". In total, a majority of 45% of the companies (45%) show a stable emissions evolution ("neutral"), 38% display an increase and only 17% present an emission reduction. Figure 7 illustrates the results per sector, revealing that companies of the industry and services sector record more often an increase in emissions, whereby the industry sector at the same time also shows the highest share of companies with a downwards trend in emissions. However, it must be noted that a company's business background and the calculation method applied for emissions may influence the observed development of emissions.

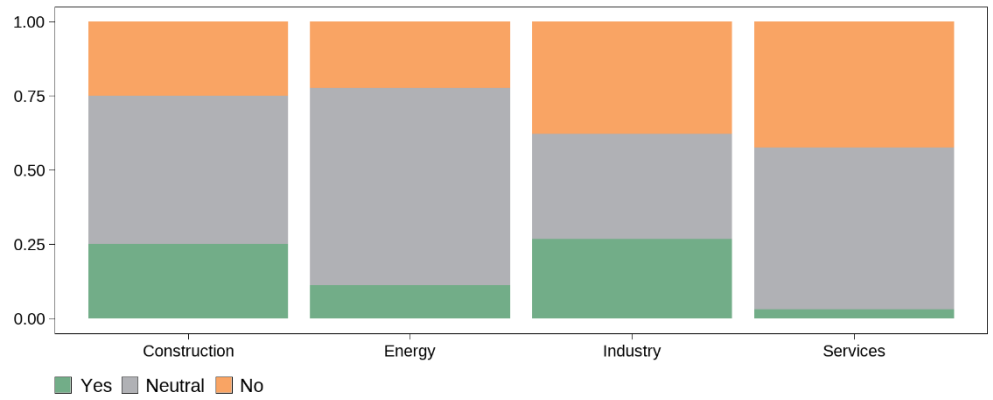


Figure 7: "Does the historic CO2 evolution show a reduction?". Note: Share of companies per sector

10 Emission intensity is measured by scope 1+2 emissions in relation to turnover.

Finally, it is analyzed if companies with a higher disclosure quality are more successful in reducing emissions than other companies. The disclosure quality is determined on the basis of the evaluation questions (see Annex 1: Template description), leading to a final score for each company that is further translated into a quality description graduating from "very good" to "poor".

The relationship between the disclosure quality and the evolution of CO2 emissions is reflected in Figure 8. It is apparent that a high-quality level is often related to a reduction in CO2 emissions, whereas companies with a poor disclosure quality tend to record an increase in emissions. Further, it can be observed that companies with a medium disclosure quality show a "neutral" development of their CO2 emissions. Comparing sectors, the graph depicts the energy sector once more as outperformer.

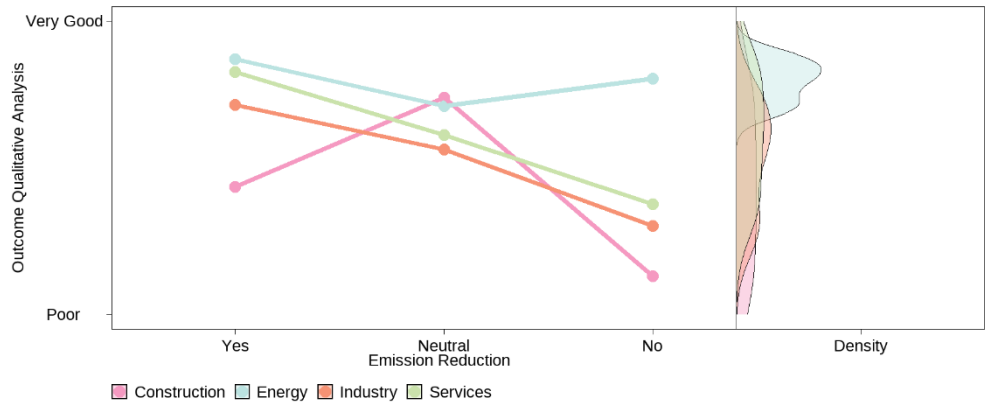


Figure 8: Disclosure quality in relation to emission reduction.

Conclusion

This study analyses the CRR related disclosure practices from a credit risk assessment point of view. Using a sample 91 Austrian non-financial corporations the results are summarized as follows:

- Risk awareness: Generally, a share of 75% of the assessed companies disclose information on physical and/or transition risks. However, a full risk awareness, meaning a comprehensive and detailed risk description, is only attributed to a smaller fraction of around 43%.
- The assessment of actions separated for physical and transitions risks reveals that the discussion on actions related to physical risks is not as far progressed as for transition risks. Whereas only 18% of the assessed companies reflect comprehensively on physical risk measures, the share rises to 75% for transition risk.
- Target setting towards decarbonisation is already well integrated in the companies reporting such as 90% of the analysed companies disclose at least one CO2 reduction target. Long-term targets are much more frequent than short term targets, however science-based targets represent a minority. Notably, only

57% of the targets are defined with a clear reference to the emissions' scope (most common scope 1&2).

- An analyses of the companies' historic emissions evolution illustrates that a majority of 45% companies show a stable emissions evolution, 38% display an increase and only 17% present an emission reduction.
- The assessment showed that there is a positive relationship between the disclosure quality and the evolution of CO2 emissions. A high-quality level is often related to a reduction in CO2 emissions, whereas companies with a poor disclosure quality tend to record an increase in emissions.
- In sector comparison, energy companies are forerunners in CCR disclosure.

Limitations of the study are the small sample size especially in the sector construction and the discretion applied by the authors in categorizing the heterogenous disclosure standards applied by the companies. With the enforcement of the CSRD, which stipulates not only a more comprehensive sustainability reporting but also an assurance by external auditors, it is likely that the disclosure quality on CCR will improve on an accelerating speed.

Bibliography

Auria, L., Bingmer, M., Graciano, C. M., Charavel, C., Gavilá, S., Iannamorelli, A., . . . Sauer, S. (2021). *Overview of central banks' in-house credit assessment systems in the euro area.* Documentos Ocasionales. Retrieved from https://www.bancaditalia.it/pubblicazioni/mercati-infrastrutture-e-sistemi-di-pagamento/questioni-istituzionali/2021-013/N.13-MISP.pdf

ECB. (2021). *ECB presents action plan to include climate change considerations in its monetary policy strategy*. Retrieved from https://www.ecb.europa.eu/press/pr/date/2021/html/ecb.pr210708_1~f104919225.en.html

ECB. (2022). Common minimum standards for incorporating climate change risks into in-house credit assessment systems in the Eurosystem. *ECB Economic Bulletin, Issue 6/2022*. Retrieved from European Central Bank: https://www.ecb.europa.eu/pub/economic-bulletin/focus/2022/html/ecb.ebbox202206_06~d7f88f706f.en.html

EFRAG. (2022). *Draft European Sustainability Reporting Standards.* Retrieved from ESRS E1 Climate hange: https://www.efrag.org/Assets/Download?assetUrl=%2Fsites%2Fwebpublishing%2FSiteAssets%2F08%2520Draft%2520ESRS%2520E1%2520Climate%2520Change%2520November%25202022.pdf

European Commission. (2018). Action Plan: Financing Sustainable Growth. *European Commission*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0097

European Commission. (2021). Commission Delegated Regulation (EU) 2021/2139. *Official Journal of the European Union*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021R2139

European Committee of Central Balance Sheet Data Offices (ECCBSO). (2024). *eccbso.org*. Retrieved from https://www.eccbso.org/

European Parliament; EU Council. (2022). Directive (EU) 2022 /2464. *Official Journal of the European Union*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022L2464

SBTi. (2024). *sciencebasedtargets.org*. Retrieved from About us: https://sciencebasedtargets.org/about-us

Annex

Annex 1: Template description

Physical Risk			
Criteria	**Assessment**	**Category**	**Evaluation**
Risk	Which relevant physical risks are reported by the company? (Open text field)	• Temperature • Wind • Water • Soil • Risks along the value chain • Several risk categories	Does the company show a comprehensive risk awareness? • Yes • Neutral • No
Actions	Which actions are taken to face / reduce the physical risks? (Open text field)	• Construction measures • Location measures • Process-related measures • Other	Are the actions in line with the named risks? • Yes • Neutral • No

Disclosed physical risks and actions are collected in open text and further categorized. The categorization of risks follows the EU regulation 2021/2139, based on the laid-out classification of climate-related hazards into temperature, wind, water, and soil being directly taken from it (European Commission, 2021). It is supplemented by risks along the value chain as a common reporting practice as well as by several risk categories in case of general or overarching risks. The categorization for the key actions a company takes to reduce the physical risks (i.e. construction measures, location measures, process-related measures, other) is defined on basis of commonly reported actions. The evaluation questions point at the risk awareness of the company and how targeted the named measures are. "Yes" stands for a company specific identification of potential or actual physical risk factors respectively for actions that are capable to minimize the named risks, "Neutral" for a partly or more general fulfilment and "No" if there is no information given.

Transition Risk			
Criteria	**Assessment**	**Classification**	**Evaluation**
Risk	Which relevant transition risks are reported by the company? (Open text field)	• Legal/policy • Market • Technology • Reputation	Does the company show a comprehensive risk awareness? Y/N/neutral
Targets	Which CO2 reduction targets does the company define? (Open text field)	N.A.	• Definition of comprehensible short term targets? Y/N • Definition of comprehensible medium / long-term targets? Y/N • Are the targets in line with the Paris Agreement? Y/N • Are the targets sciences based and validated (SBTi)? Y/N
Actions	Which actions are taken to enable the defined targets? (Open text field)	• Energy efficiency • Material efficiency • Fuel switching • Electrification • Renewable energy • Modification product • Modification process	• Are the actions in line with the main CO2 causing activities, risks and targets? Y/N/neutral • Are the actions measurable and

			scheduled? Y/N/neutral • Is a comprehensible implementation plan disclosed? Y/N/neutral
Emissions / Evolution	Historic CO2 emission values (scope 1+2+3)	N.A.	Does the historic CO2 evolution show a reduction? Y/N/neutral

Data collection for transitions risk is broader than for physical risk and comprises next to risks and actions also the aspects targets and emissions evolution. The description of disclosed transition risks is collected in open text, first. The further classification of transition risks into policy/legal, technology, market, and reputation, follows market standards, also adopted by the EFRAG in the defined ESRS E1 climate change (EFRAG, 2022). The related evaluation question regarding risk awareness is answered with "Yes", if the company describes its actual or potential transition risk comprehensively from different perspectives, "Neutral" if risk examination is of superficial nature, and "No" if there is no risk disclosure at all. The targets respectively the CO2 reduction path is collected in an open text field, according to the companies' individual target setting. To normalize this individual information, four target related evaluation questions were defined. The questions focus on the timeline of the disclosed targets (short term, medium to long-term), if they are in line with the Paris Agreement and at best, if the targets are science based. The questions are answered with "Yes" or "No".

The categories for the key measures a company takes to achieve its defined CO2 reduction targets are aligned with the categorization in ESRS E1 climate change (EFRAG, 2022), complemented by the category "Other". The evaluation focuses on the target-orientation of the actions, on the precise definition regarding timeline, and on the availability of a comprehensible implementation plan. The questions are answered with "Yes", "Neutral" or "No".

The assessment of the emissions evolution is based on the historic emissions data reported by the company and supplemented by a calculated emission intensity ratio (scope 1+2 emissions / turnover). The evaluation question "Does the historic CO2 evolution show a reduction?" is answered with "Yes" if there is a clear downwards trend in the emissions' volume, "Neutral" if the reported emissions values are widely unchanged or volatile and "No" if the emissions reveal an increasing tendency. For the correct evaluation of this question also other factors must be considered, such as e.g. a change in the business scope / business units, price related effects on the turnover etc., that might distort the observable trend line.

Annex 2: Detailed analyses results

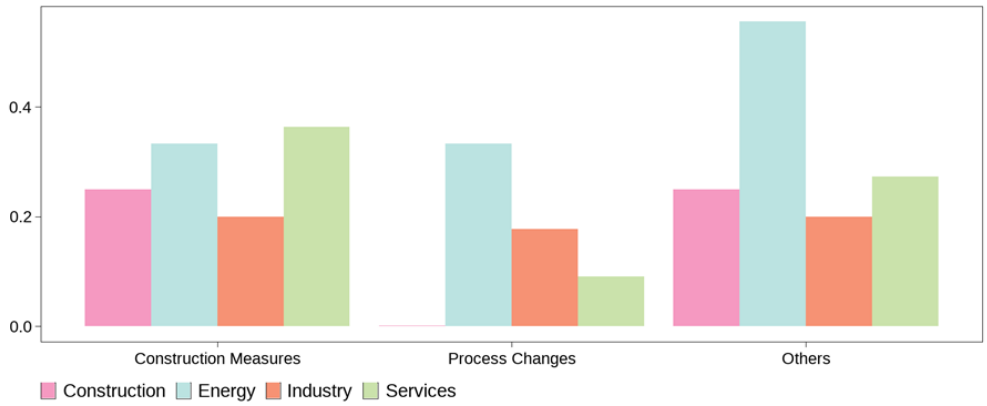


Figure 9: Classification of actions to mitigate physical risk. Note: Share of companies per sector.

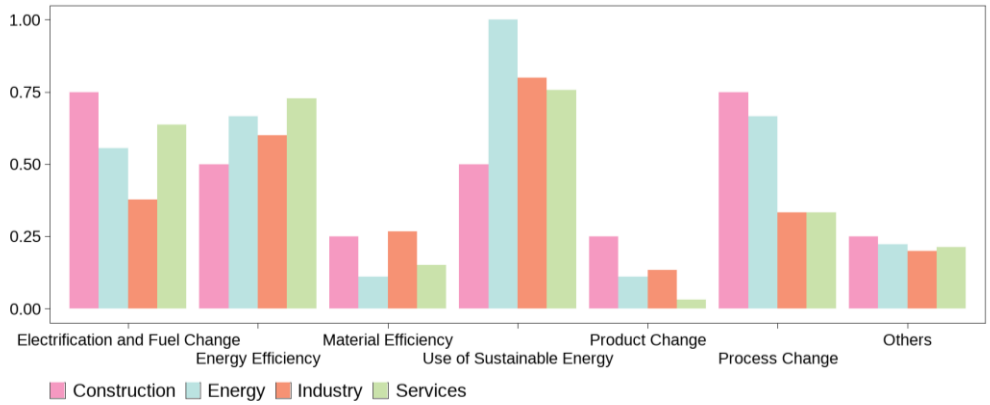


Figure 10: Classification of actions to mitigate transition risk. Note: Share of companies per sector.



Figure 11: Assessment of actions for transition risk. Note: Share of companies per sector.

OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

Quality over quantity? Using corporation‘s climate change related disclosure for risk assessment

Madrid, 17. Oktober 2024

Daniela Gasser
Statistics Department – Supervisory Statistics, Models and Credit Quality Assessment Division
www.oenb.at



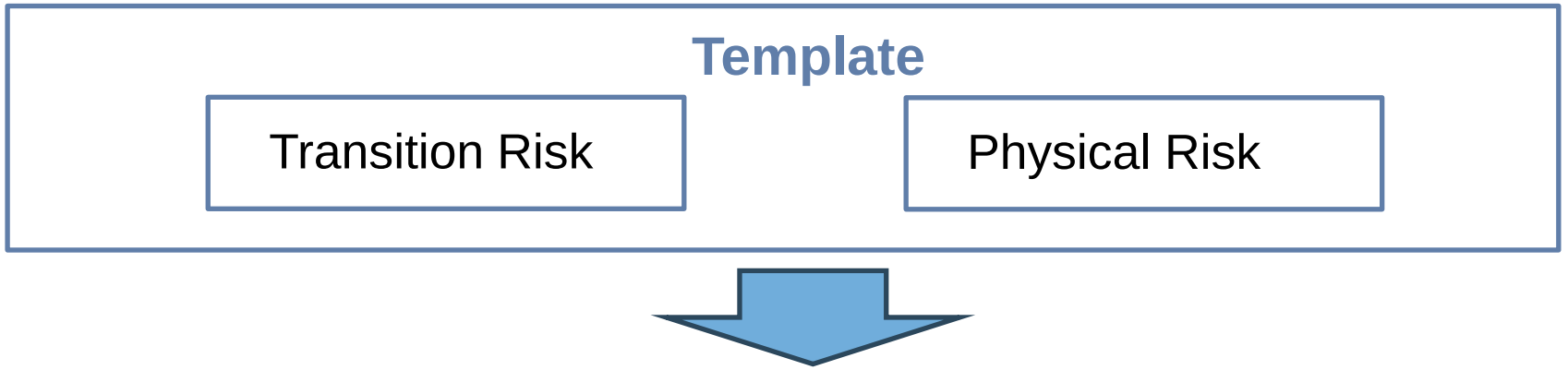
The need of climate change risks integration into credit risk assessment

- Action Plan of European Commission on „Financing Sustainable Growth“
- ECB: Introduction of climate change risk (CCR) in the in-house credit risk assessment systems (ICAS)

  ⇨ Basis: Relevant firm-level information (NFRD / CSRD, voluntarily provided information)

Our paper: Analysis how and to which extent companies disclose CCR information based on current and prospective disclosure standards to be used for the assessment of CCR in the context of ICASs

Methodology - Template and analysed criteria



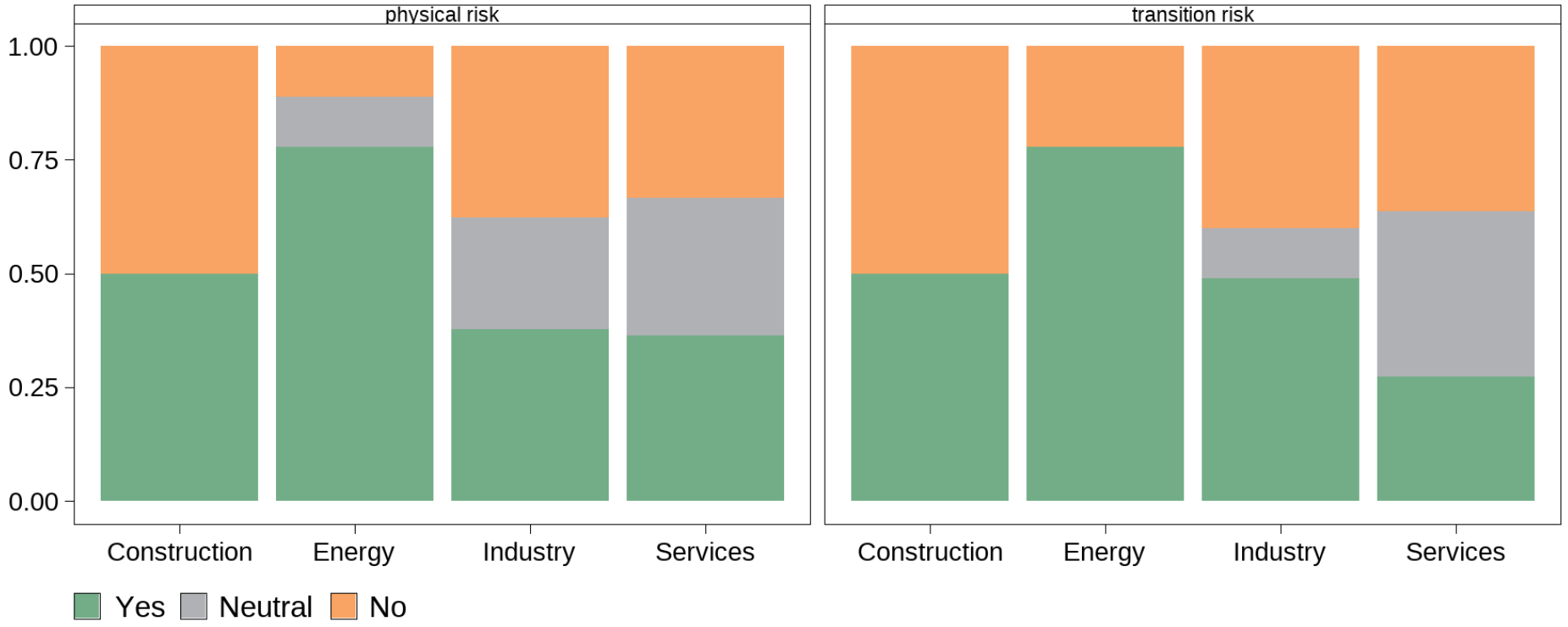
- Sustainability reports of **91 Austrian non-financial corporations**
- **Criteria:** risk awareness, actions, target setting, and emissions,
- **Outside-in** perspective
- **Tools:** Open text questions, categorisation into nominal variables, evaluation questions
- **Emissions:** Historic data reported by companies and calculated emission intensity ratio

The data sample consists of 91 Austrian non-financial companies

ERICA Sector	Size			Total
	Small	Medium	Large	
Construction	1	0	3	4
Energy	0	1	8	9
Industry	11	14	20	45
Services	5	17	11	33
Total	17	32	42	91

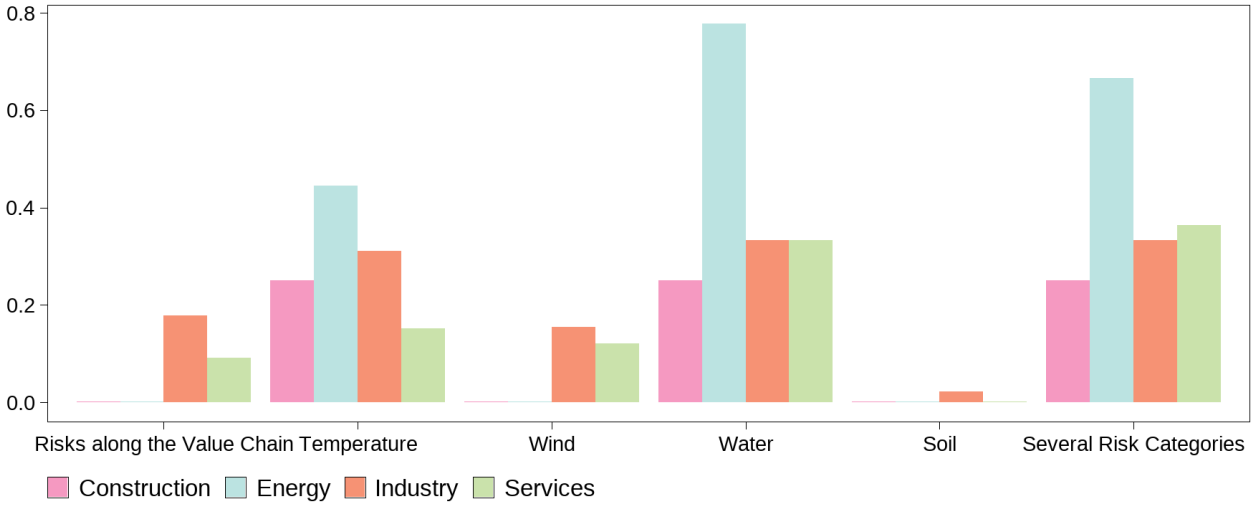
- **39** firms are **listed** companies (NFRD),
- **52** corporations **voluntarily** disclose sustainability information

Empirical results: Risk awareness

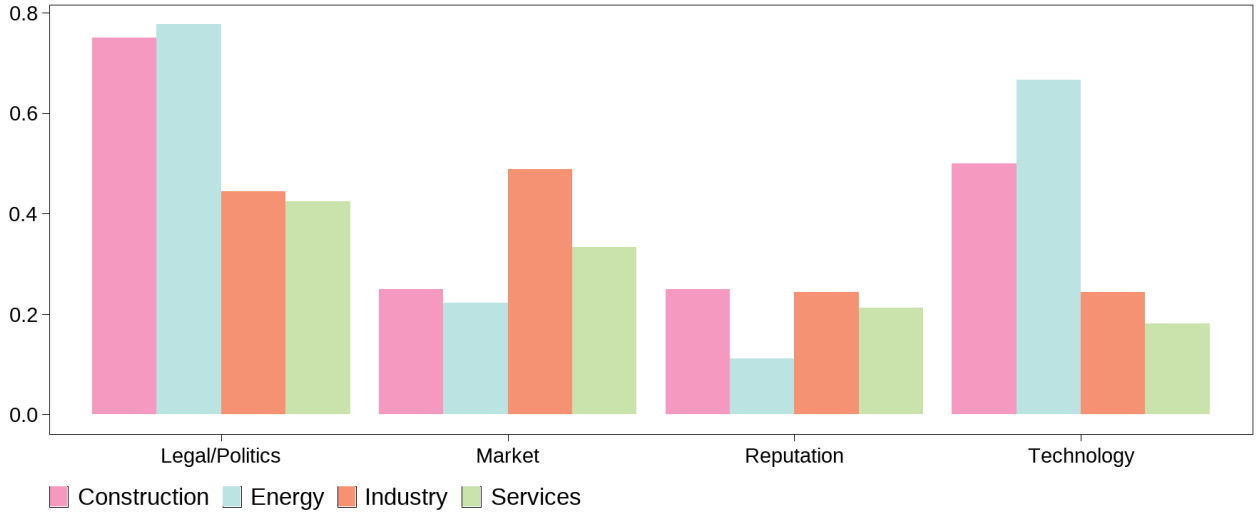


- **75%** of the assessed companies disclose information on physical and/or transition risk
- **Full risk awareness** is only given for **43%**
- Energy sector takes a **leading role**

Empirical results: Most frequent risk categories



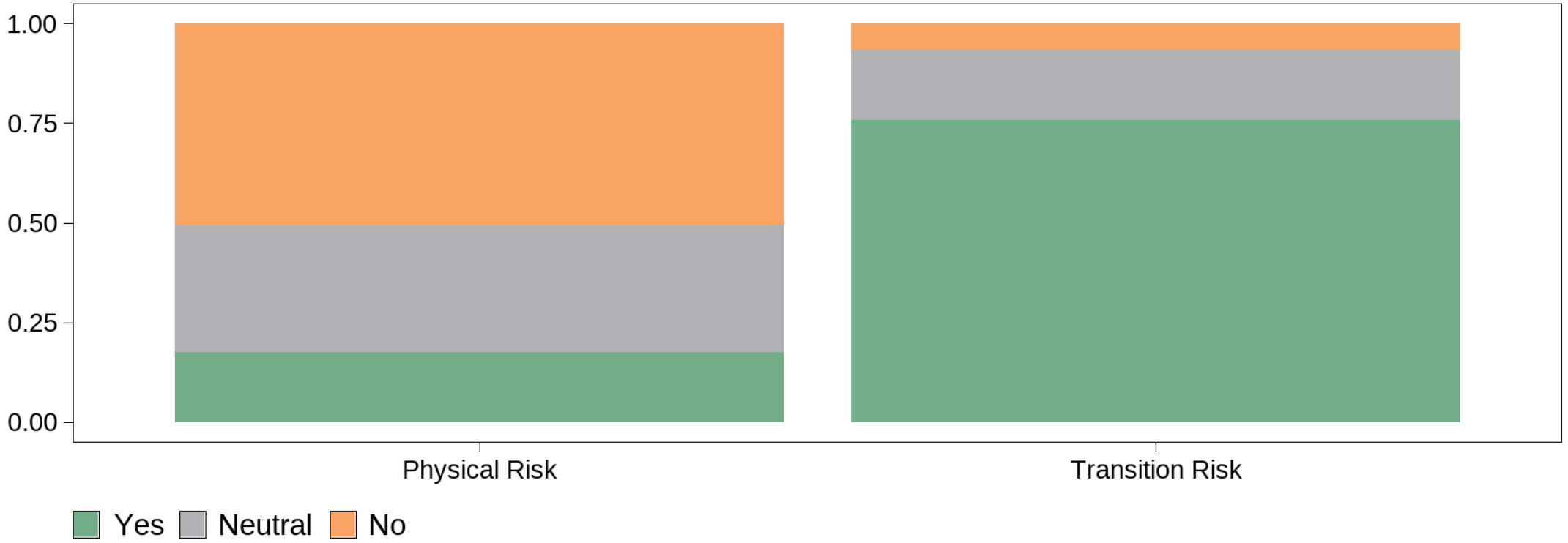
- **Water stress** (e.g. flooding, droughts) is the most frequent risk
- **Temperature** is another crucial physical risk category



- Tranmission channel **„legal/politics** is most frequent
- **Reputational risks** are listed **least** often

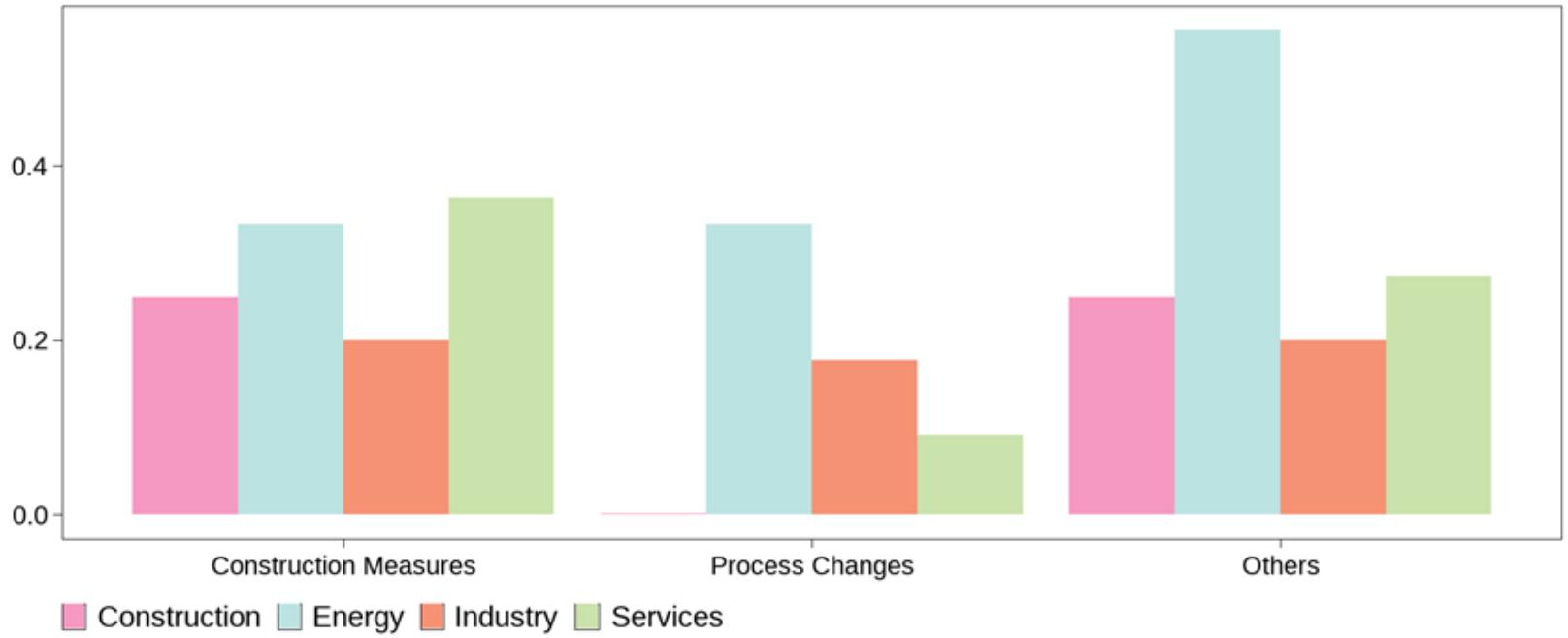
oenb.info@oenb.at

Empirical results: Disclosure of actions



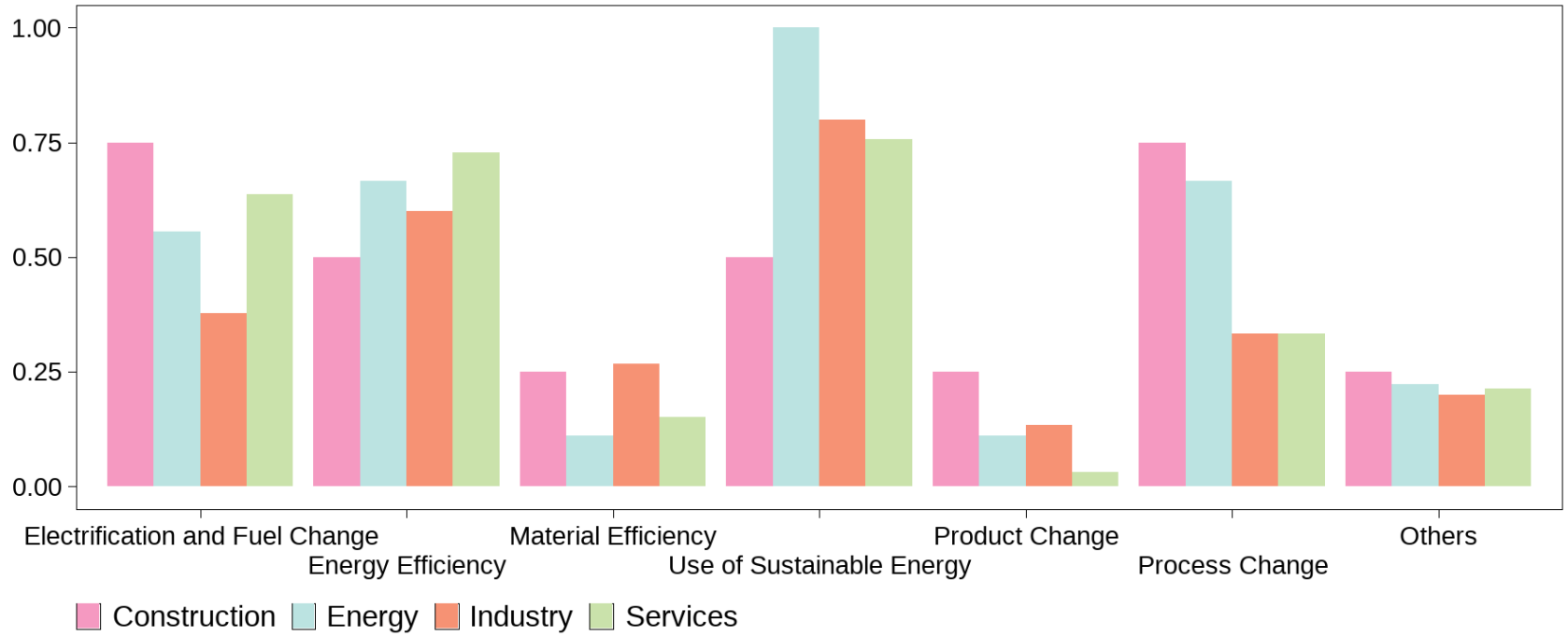
- Actions to mitigate transition risks are common practice
- Actions to mitigate physical risks are not far progressed

Empirical results: Classification of actions to mitigate physical risk



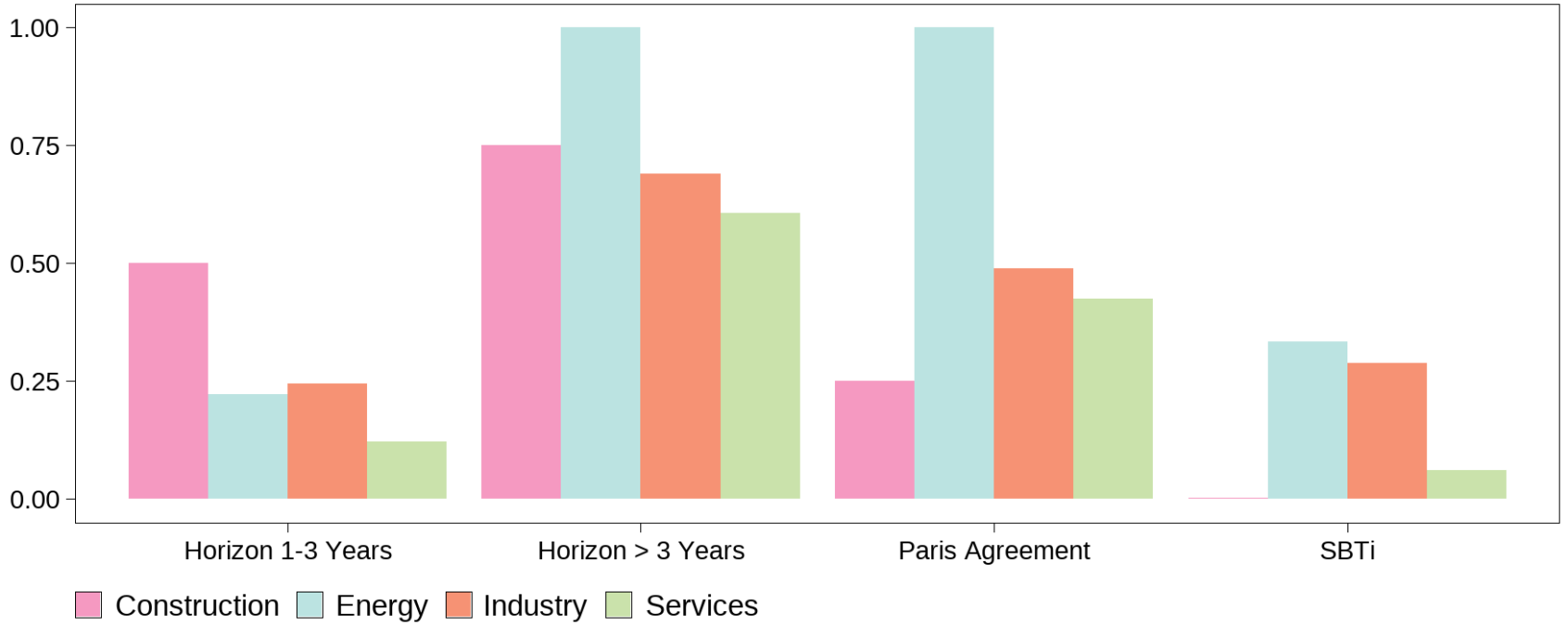
- **The descriptions of construction measures vary in its level of detail.**
- **Companies of the energy sector often focus on contingency plans.**

Empirical results: Classification of actions to mitigate transition risk



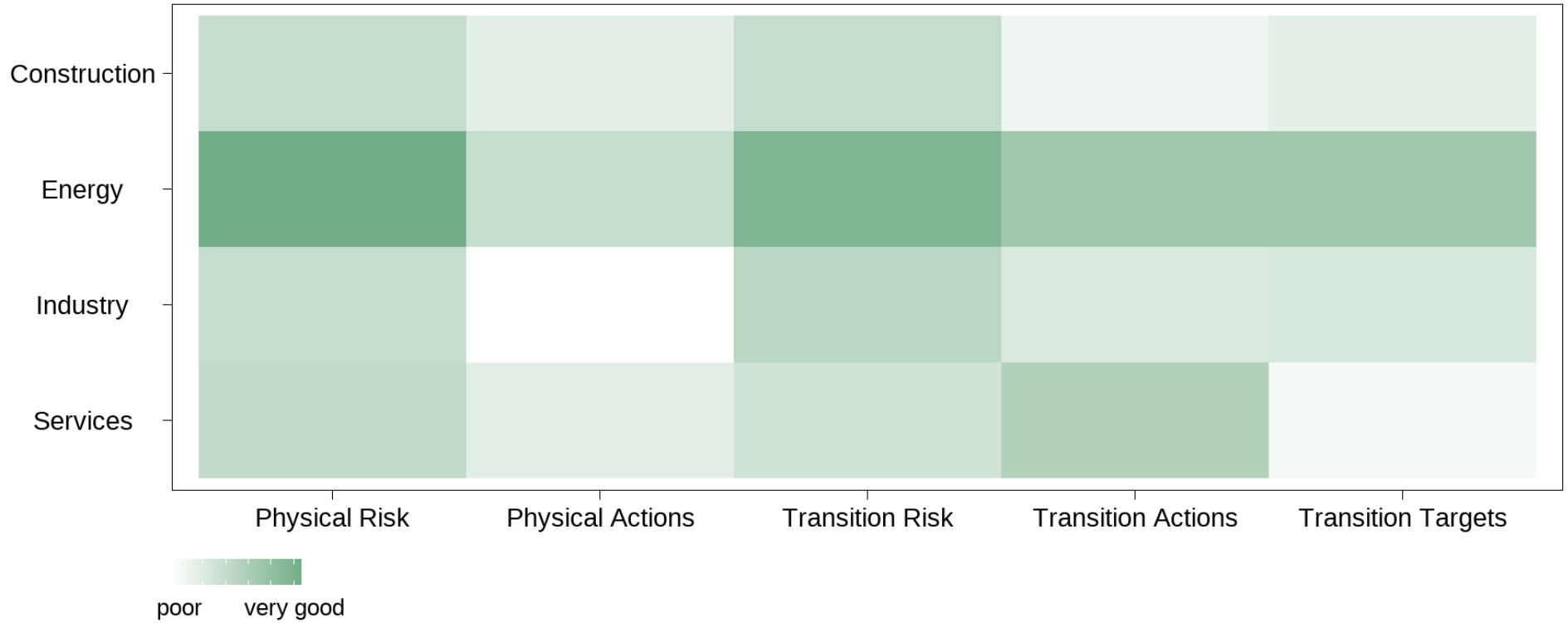
- **Externally driven actions** such as electrification and use of sustainable energy are **preferred**
- **Internal adaptations** such as material efficiency, product or process change are **less common**

Empirical results: Target setting towards CO2 reduction



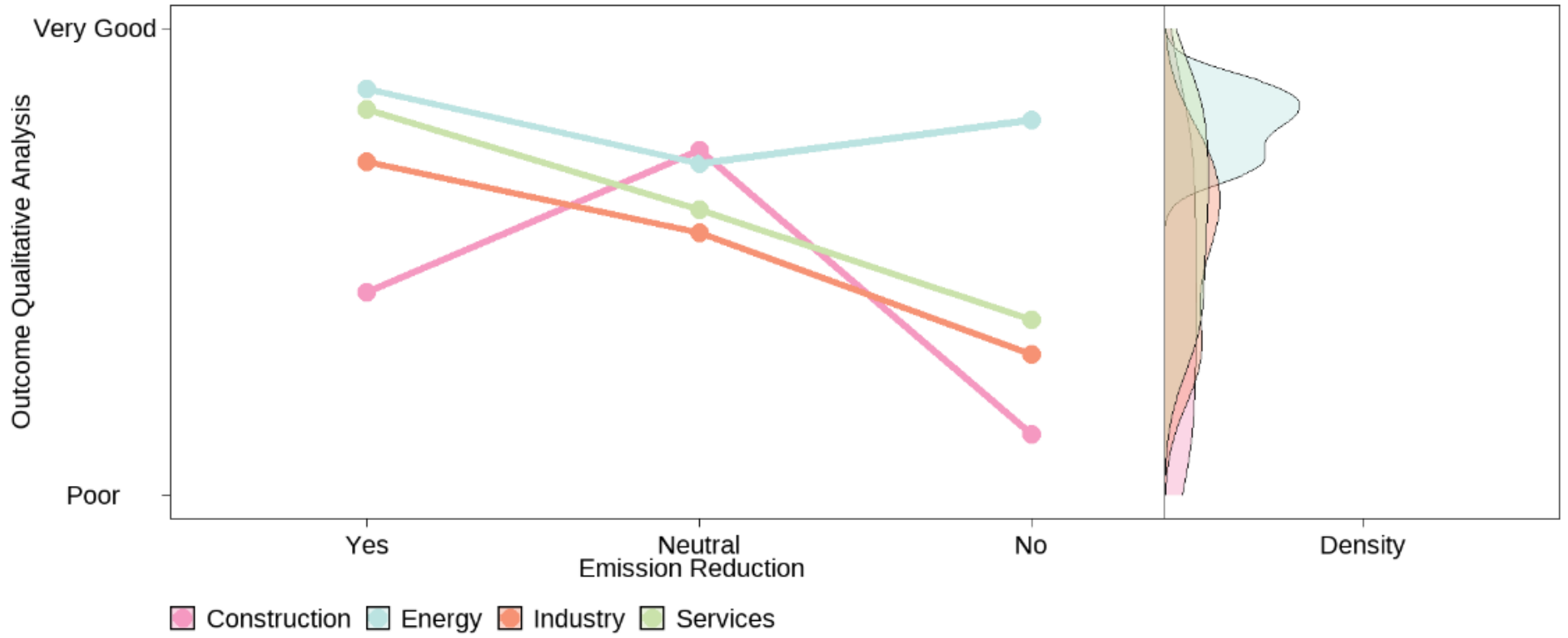
- Goals with a **horizon longer than 3 years** are most common
- All companies of the energy sector disclose goals, which are in line with the Paris agreement
- Science based targets are disclosed by a third of the energy and industry companies

Empirical results: Summary of the disclosure quality



- Risk awareness is balanced for both risk types
- Disclosure quality on actions is higher for transition risk
- Energy sector is forerunner in all categories

Empirical results: Assessment of emissions and relation to disclosure quality



- A **high-quality level of disclosure** is often related to a **reduction in emissions**, whereas
- Companies with a **poor disclosure quality** tend to record an **increase**.

Conclusion

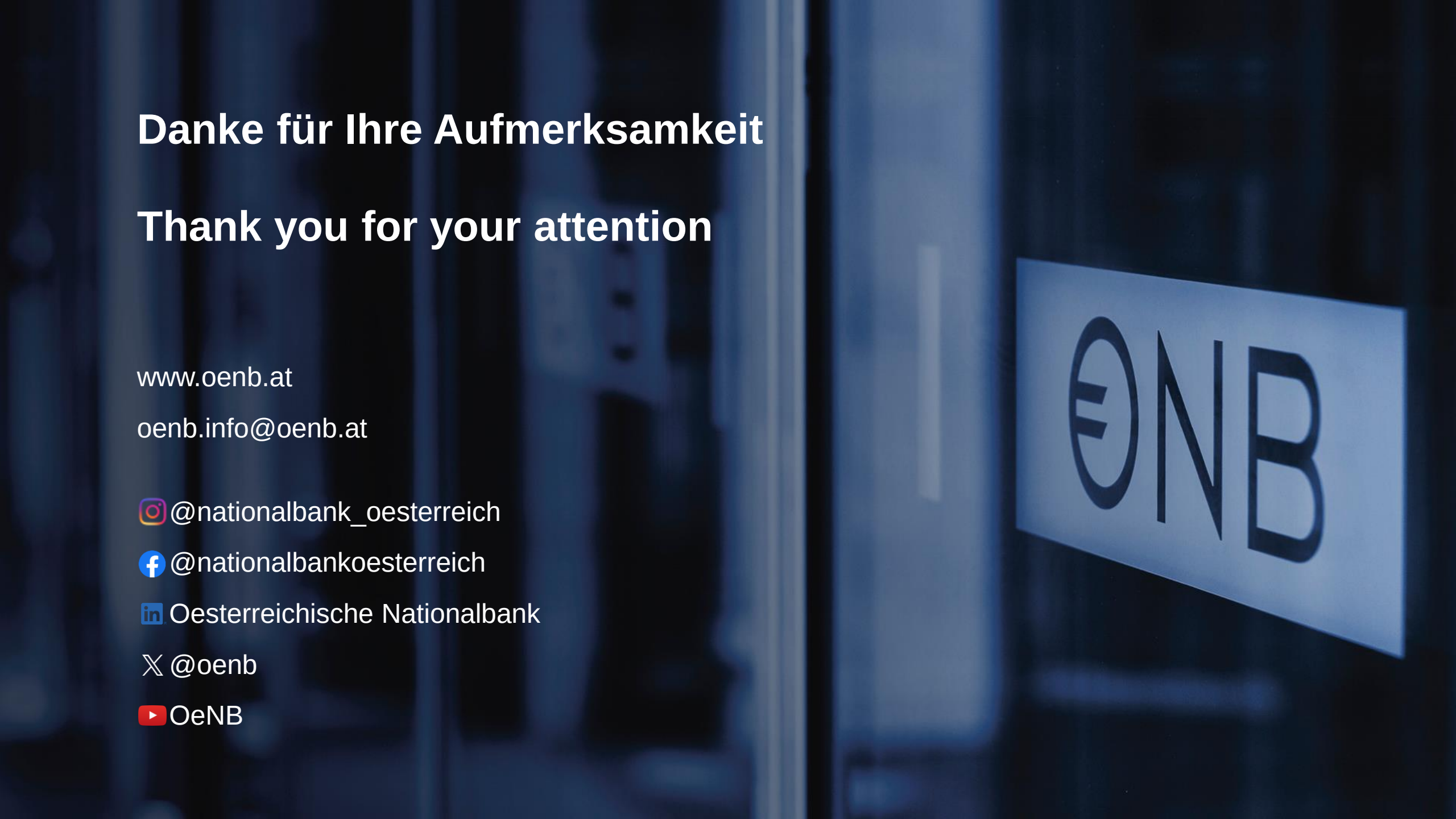
- Companies broadly disclose on risk awareness, actions and targets.
- Differences between physical risk and transition risk are observed.
- Higher disclosure quality is often associated with a downward trend in CO2 emissions.

- Limitations:
  - Small sample size, especially in sector construction
  - Discretion applied by authors in categorizing heterogenous disclosure standards

- Outlook:
  - Standardisation with application of CSRD
  - Assurance by external auditors will be required

Danke für Ihre Aufmerksamkeit

Thank you for your attention



www.oenb.at

oenb.info@oenb.at

@nationalbank_oesterreich

@nationalbankoesterreich

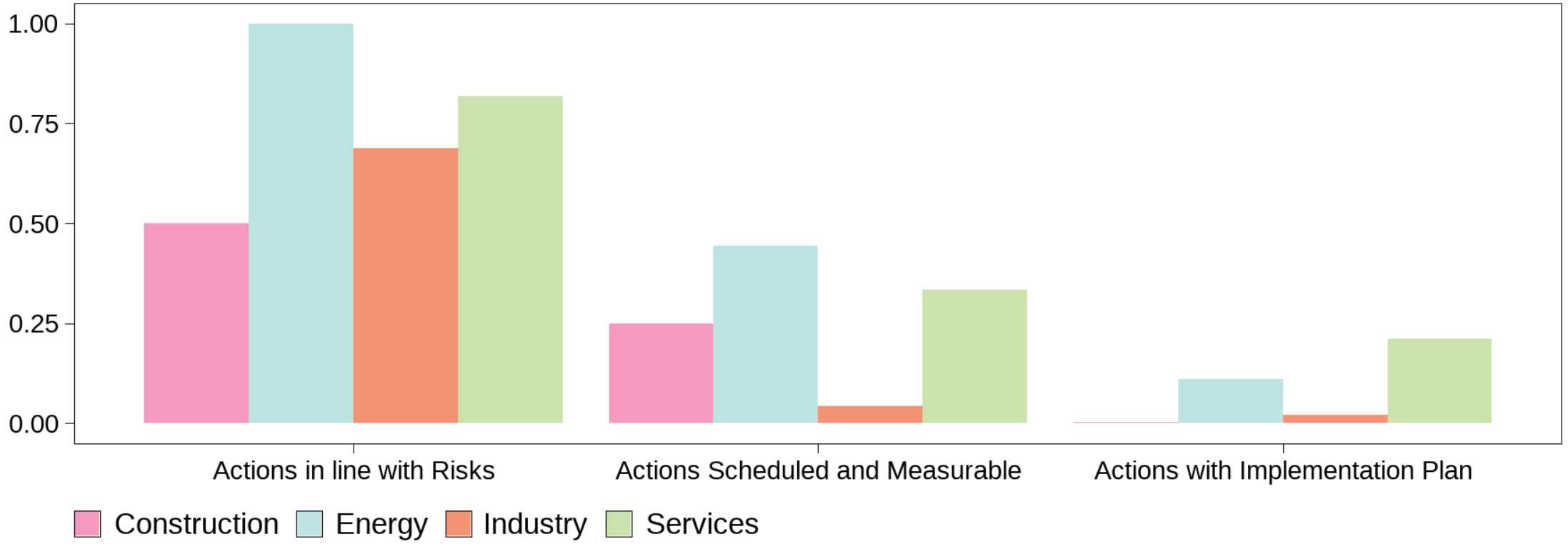
Oesterreichische Nationalbank

@oenb

OeNB

Backup

Empirical results: Assessment of actions for transition risk



IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

It's in the financials, stupid! But is it certain?[1]

Christian Haas, Ulf Moslener and Sebastian Rink,
Frankfurt School of Finance and Management

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

It's in the Financials, Stupid! But is it Certain?

Christian Haas[1], Ulf Moslener[1] and Sebastian Rink[1]

[1]Frankfurt School of Finance & Management

This version: September 2024

Abstract

Sustainability data is increasingly relevant for multinational enterprises (MNEs), financial institutions, and researchers. However, sustainability data remain incomplete, fragmented, or scarce. In our paper, we propose a novel approach to address this challenge using machine learning (ML) to predict sustainability metrics from readily available financial data. This method allows for a more detailed and accurate assessment of sustainability in MNEs and their global value chains. Our approach is tested using a comprehensive dataset of financial and sustainability information at the company level. The results indicate that ML is helpful in predicting key sustainability metrics, such as corporate carbon emissions and water discharge. However, users should reflect on the specific use case when applying ML since model performance can vary sectorally, spatially, and temporally. In addition, we develop a metric to assess the uncertainty of the predictions and find that it can substantially affect the model output. Regulators should build on our findings to encourage the use of ML-generated sustainability data while also requiring more transparency from data providers and model users.

Keywords: corporate carbon emissions, corporate sustainability information, financial data, machine learning

JEL Codes: C53, Q54, Q56

We would like to thank Christina E. Bannier and Maurice Dumrose for thorough feedback as well as participants in the Sustainable Finance Research Platform brown bag lunch for helpful comments. The authors acknowledge funding from the project safe Financial Big Data Cluster (safeFBDC) by the Federal Ministry for Economic Affairs and Climate Action.

I. Introduction

Sustainability commitments such as the Paris Climate Agreement (United Nations, 2015) and the Kunming Biodiversity Declaration (Kunming Declaration, 2021) show the ambition to transform economies worldwide. As a result, multinational enterprises (MNEs), financial institutions, and researchers increasingly require sustainability information. Availability and quality of sustainability data throughout the dimensions environment (E), social (S), and governance (G) are more important than ever. However, these data are incomplete, fragmented, or scarce. In this paper, we present a novel approach to using machine learning to fill these sustainability gaps.

MNEs have different reasons for improving the sustainability of their products and services such as innovation capacity (Acemoglu, 1997; Balasubramanian and Lee, 2008), regulatory and investor pressure (Slager et al., 2023), or financial performance (Kim and Starks, 2016; Adams and Ferreira, 2009). A key component in the evaluation and management of sustainability in MNEs is the availability and quality of sustainability data for MNEs, their subsidiaries, and their global value chains (GVCs) (Marano et al., 2024). As reported and audited data are only gradually and incompletely becoming available, MNEs are forced to rely on other data sources.

Financial institutions and regulators have received increasing attention to sustainability in recent years. In particular, internationally operating banks and investors have committed to steer their portfolios in line with sustainability goals (Bolton et al., 2022) while retail investors display some preference for green financial products (Bauer et al., 2021). An increasing awareness of the financial risk associated with sustainability has led many regulators to strengthen sustainability-related risk management regulation (ECB, 2022). Today, financial markets reflect biodiversity and climate-related risks to some extent (Ilhan et al., 2021, 2023; Giglio et al., 2023; Garel et al., 2023). These developments require increasing amounts of sustainability data from investees and lendees.

Research on sustainability and business practices has grown in recent years (e.g. Marano et al. (2024) and Starks (2023)). Many empirical studies require company-level sustainability data. Sufficient data quality and data availability, as well as understanding modeling assumptions for non-reported data, is required in the process.

Currently, sustainability at the company level is typically measured by ESG ratings. However, these ratings diverge in terms of measuring specific ESG aspects (Berg et al., 2022). Given the difficulty of measuring company sustainability (Edmans, 2023), granular physical indicators would allow different actors to assess corporate sustainability more inde-

pendently, allowing greater diversity of views and, as a result, potentially more efficient functioning of markets, for example, through capital allocation and GVC engagement. These granular physical indicators include corporate carbon emissions, the corporate biodiversity footprint, corporate resource usage (water, primary materials, etc.), and diversity measures such as gender ratios.

However, granular physical sustainability data remain incomplete, fragmented, or scarce. Corporate ESG disclosure around the world is currently evolving but far from being established (Krueger et al., 2024). It is unlikely that full sustainability information on GVCs will be available soon. Corporate carbon reporting seems to be the most evolved, but it remains incomplete (Busch et al., 2022). This can lead to frictions in the efficient use of these data for decision making. Therefore, methods are needed to fill the data gaps.

Data providers and researchers have realized this need and have provided different modeling approaches to estimate ESG data. To date, this work has typically been limited to corporate carbon emissions. Methods include simple estimations, which use a direct proportional relationship between a company's size (e.g., revenue, number of employees) and the corporate carbon footprint, and regression setups, where emissions are regressed against a variety of operational and financial predictors to encapsulate a company's business model, scale, and technological practices (e.g., Goldhammer et al. (2017) and Griffin et al. (2017)). However, these approaches suffer from simplicity and thus potential biases for out-of-sample predictions, as well as a lack of a generally accepted theory on how accounting data drive corporate emissions.

Machine learning can help overcome these shortcomings, even beyond corporate carbon footprints. Companies are complex systems (Loughran and McDonald, 2023). They have different ages, sizes, product lines, geographic presences, financing structures, and company cultures. These aspects may all play a role for companies' sustainability metrics, and it is very likely that relationships between and within company-level metrics are nonlinear. Machine learning with its ability to search large non-parametric algorithmic spaces (Jordan and Mitchell, 2015) seems well positioned for this environment.

Previous work employing machine learning to predict sustainability data already shows promising results, but is limited in metrical and methodological scope. Nguyen et al. (2021) and Nguyen et al. (2022) estimate corporate Scope 1-3 emissions using a set of regression-based supervised learning algorithms and achieve up to 30% improvement compared to parametric approaches. Other estimates of environmental data at the company level remain rare. Tian (2023) estimates the water efficiency of the companies. The literature is largely silent on the estimation of social aspects at the company level. Governance

aspects are discussed, especially diversity aspects, with Ranta and Ylinen (2023) using text-based machine learning to generate diversity indicators from social media posts and Khan et al. (2023) predicting board diversity from company characteristics. This diversity of approaches to applying machine learning requires the user to tailor new data sets, code environments, and machine learning algorithms for each sustainability metric. This is time-consuming, costly, and challenging in environments where a variety of sustainability data points are necessary, such as in disclosures by MNEs' global value chains, in risk management by banks, or research involving various company-level sustainability aspects.

In addition, even if such estimations are available, regulators normally do not accept their use by financial institutions or MNEs. The main reason is that only point estimates are provided and no additional information is available as to how reliable that estimation is at the firm or portfolio level. Point estimates as they are generated by machine learning models represent the conditional expectation of the predicted variable. Information about confidence intervals or coverage intervals is not provided. This is also a challenge when making causal inferences from these data in research.

We propose an approach that solves these challenges. Essentially, this is achieved through three characteristics or our approach: First, we restrict the independent variables (features) to a large, multidimensional but readily available set of company-level financial data supplemented by fundamental data such as industry, country, and company age. Second, the code architecture systematically integrates the search for the best-performing algorithms in combination with the appropriate data preprocessing before those combinations (pipelines) are optimized and then fed into a final meta-model training. Third, our approach incorporates considerations about prediction uncertainty. Based on recent developments in the machine learning literature (e.g. Tibshirani et al. (2019) and Barber et al. (2023)), we add two prediction intervals to our framework. The first refers to the probability that *on average* in all predictions the true value lies within the interval (*marginal coverage*). The second refers to the probability that the true value lies within the interval *for a specific set* of input values (*conditional coverage*).

We apply our setting to consider three questions:

Q1: To what extent can we derive corporate sustainability data from corporate financial data using machine learning?

Q2: How does the prediction performance change for different dimensions, such as time, region or sector?

Q3: What can uncertainty measure reveal about the quality of the point estimates within

our approach?

In order to demonstrate that a lot of information on the sustainability metrics is actually captured "in the financials" (Q1), we test our approach on the sustainability metrics Scope 1 and Scope 2 emissions, air pollution (NOx emissions), water discharge, and female board share. Despite the versatility of the approach (i.e. just one dataset for all indicators), the predictions seem reasonable for these different metrics. The performance of the framework varies between sustainability indicators and seems to work better for non-truncated data. In doing so, we contribute to the literature by (i) confirming that financial data contain sustainability relevant information and (ii) demonstrating that our approach outperforms existing machine learning-based approaches.

We show that the performance of the models also varies between the time, region, and sector dimensions (Q2). For the earlier years in the dataset (2005 until about 2015), the predictions tend to be better than in the following years. When looking at different world regions, we also observe variations with the predictions for Europe being better than those in other parts of the world. With respect to the different sectors, the approach performs differently and is dependent on the sustainability indicator predicted. In agriculture, for example, the prediction of water discharge is comparatively good, while that of air pollution is relatively bad. This is vice versa for transportation. These findings indicate that the literature in the field should not only rely on the global means of model performance metrics but should make it conditional on the respective use case.

The issue of prediction uncertainty (Q3) is to date not considered in the existing literature related to the imputation of missing sustainability data. We fill this gap by adapting recent developments in the technical literature on machine learning. We find that the *conditional coverage* provides relatively precise information about the confidence intervals of the predictions. As such, this metric should be reported. The intervals are dependent on the coverage needed, that is, the desired confidence a user would like to have her predictions. Using a 68% coverage versus a 95% coverage as examples, we find that a higher risk tolerance at 68% would lead to very small intervals, whereas a low risk tolerance would result in large intervals. This leads us again to deduce that the consideration of use case-specific parameters is important when using machine learning in general and specifying confidence intervals in particular.

We conclude by making the case for the use of machine learning to fill data gaps, as it is a cost-efficient way to generate company-specific sustainability data while maintaining high standards for data integrity. Our suggested approach should help regulators to accept the use of estimations as they are more qualified. Additionally, our approach will help MNEs

and financial institutions meet the disclosure requirements on their GVCs and portfolios, as well as enable them to run more granular analyses including the consideration of risk tolerances in sustainability data predictions.

The remainder of the paper is structured as follows. The next section outlines the methodology, including the introduction of the uncertainty measure. Section III presents the results along the three questions. Section IV concludes.

II. Methodology

Predicting cardinal sustainability data is a supervised ML regression task. Here, we suggest a versatile approach to using regression-based machine learning to predict a variety of company-level sustainability data only from financial data. This approach could help make machine learning a more widely used tool in sustainability and business research (Bosma and van Witteloostuijn, 2024) as well as regulation and industry. In addition, we propose a method to estimate the prediction uncertainty next to the traditional model performance criteria based on point estimates.

A. Model Space

To ensure the versatility of our approach, we define a large model space that we search using a Bayesian approach[1] to maximize our chance of capturing the "best" model for a given sustainability metric. We maximize the coverage of the model space by first training and optimizing a large set of model pipelines (running up to 9,600 trials per sustainability metric[2]), then optimizing the best-performing model configurations, and finally using them in meta-model training, see Figure 1. We evaluate the performance throughout the process using the Mean Squared Error (MSE) and double 10-fold cross-validation.

Base Model Training - Pipeline Selection: The first step of base model training consists of the search for suitable model configurations (pipeline elements per learner) for the sustainability metric at hand by exploring a variety of data preprocessing techniques along with different regression learners as input to hyperparameter optimization. In this step, we run up to 25 optimization trials per model configuration.

Data preprocessing is an essential aspect of the modeling process as it can significantly influence the performance of predictive models. In the literature on sustainability data prediction, this step is usually treated separately from the actual model training. Here, we use data preprocessing as a hyperparameter for optimization in itself. In doing so, we expand the considerations of the no free lunch (NFL) theorem (Wolpert and Macready, 1997) to preprocessing in our setup. The options for data preprocessing in our setup span missing indicator flags, imputation methods (mean, median, iterative), outlier removal strategies (none, winsorization), scaling techniques (standard, robust), transformation approaches (none, quantile) and feature selection methods (none, Lasso). We introduce

[1] Unlike grid or random search, Bayesian optimization utilizes past evaluation results to choose the next set of hyperparameters, efficiently narrowing down to the best possible model settings (Snoek et al., 2012).

[2] We have devised an early stopping mechanism to boost computational efficiency. For more information, see Appendix B.

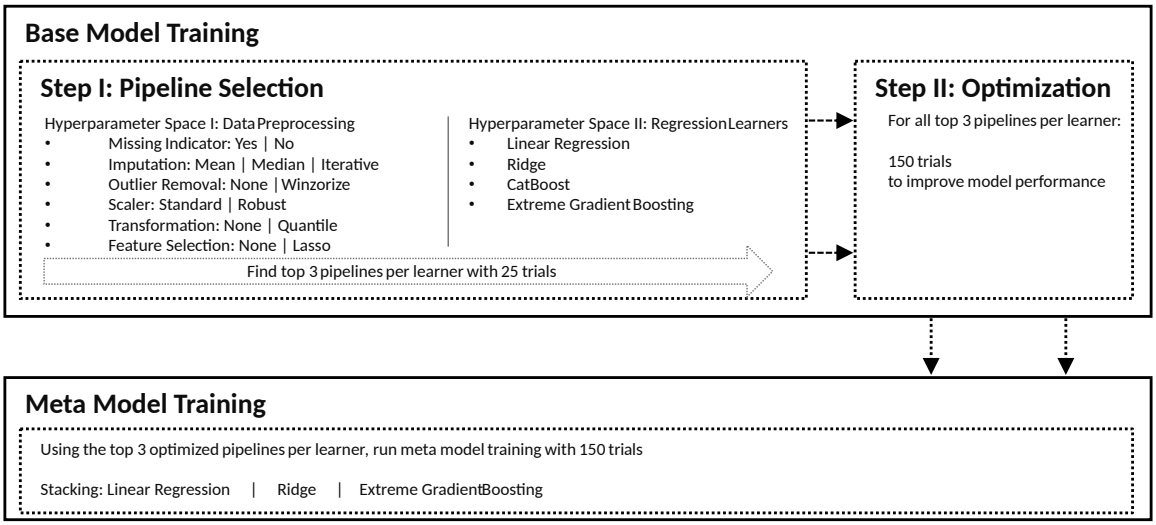


Figure 1. Machine Learning Approach

The figure illustrates the process of Base Model Training, including pipeline selection and optimization using Bayesian hyperparameter tuning, followed by Meta Model Training with a stacked regression approach. Evaluation uses Mean Squared Error (MSE) and double 10-fold cross-validation. Base model training involves selecting top pipelines per learner, followed by further optimization. Meta model training uses the top pipelines in different stacking configurations.

the missing indicator flag (a binary feature) to allow the model to learn from the reporting behavior of a company. The imputation methods are relevant for handling missing data, a common issue in financial datasets, and can affect the model's bias and variance. Imputation strategies such as mean and median are simple and widely used, while iterative methods can provide a more sophisticated approach that accounts for correlations between features (van Buuren, 2007). Outlier removal and scaling enhance the robustness and stability of the models, particularly in financial applications where outliers can represent noise (Aggarwal, 2013). Feature engineering and selection further refine the model by introducing new features that could capture non-linear relationships or selecting the most relevant features to avoid overfitting (Iguyon and Elisseeff, 2003).

In the space of regression learners, the approach contemplates linear regression, ridge regression, extreme gradient boosting (XGBoost) and CatBoost. Linear regression and ridge are fundamental techniques with ridge introducing regularization to manage multicollinearity and overfitting (Hoerl and Kennard, 1970). They are relevant in this study for comparing the training output with methods that are used regularly in econometrics. However, it is unlikely that these learners form the basis of the "best" performing models in the realm of machine learning. On the other hand, Gradient Boosting algorithms, including XGBoost (Chen and Guestrin, 2016) and CatBoost (Prokhorenkova et al., 2018), are powerful ensemble methods that have shown high performance on a wide range of prediction tasks, particularly in the presence of non-linear and complex relationships. One

of the two algorithms is expected to perform best in the context of sustainability data prediction.

Compared to Support Vector Machines (SVMs), K-Nearest Neighbors (KNN), neural networks, and simpler tree methods, XGBoost and CatBoost bring a blend of depth and breadth to the modeling process. SVMs, while effective for small to medium-sized datasets, can be outperformed by tree-ensemble methods in handling large and complex data sets (Fernández-Delgado et al., 2014). KNN suffers from the curse of dimensionality and is inherently slower in making predictions due to its instance-based nature (Beyer et al., 1998). Neural networks, although powerful for large-scale and complex non-linear relationships, require extensive tuning and larger datasets to generalize effectively without overfitting (LeCun et al., 2015). This could reduce the versatility of our setup. Simpler tree methods such as CART or C4.5 can provide interpretable models but usually lack the predictive power of boosted ensembles, which aggregate multiple trees to reduce variance and bias (Breiman, 1996). Therefore, for our regression tasks using financial data, where the data can be noisy and feature relationships complex, XGBoost and CatBoost are likely to be good fits.

Base Model Training - Optimization: The second step involves optimizing the top three model configurations per learner and target variable, selected according to their initial performance in Step I. Each model configuration undergoes 150 Bayesian hyperparameter optimization trials. With this step, we substantially expand the optimization efforts to ensure that the model performance is close to "best".

Meta Model Training: Finally, meta-model training is applied using the three best-performing models per learner. The approach employs stacking, which combines the predictions of multiple models by training a meta-learner, often leading to performance improvements (Wolpert, 1992).

B. Model Performance Metrics

We evaluate the performance of the top three models from both base and meta-training using performance assessment metrics as summarized in Table I. We evaluate the point estimates using these metrics.

Global assessments encompass a set of quantitative measures designed to evaluate the predictive performance of global models. The metrics in this category include the accuracy distribution, which examines the variability in prediction accuracy through the distribution of relative errors between predicted and actual values. The mean absolute error (MAE) and the mean squared error (MSE) quantify the average prediction error and

Metric	Measure/Value	Description
Global Assessment		
Accuracy distribution	Distribution	Refers to the distribution of relative errors between predictions and actual values.
Mean Absolute Error(MAE)	Scale	The average of the absolute errors between the predicted and actual values, representing average prediction error.
Mean Squared Error(MSE)	Numeric value	The average of the squared errors, emphasizing larger errors more than MAE. Also used in the model training as loss function.
R^2 (R-squared)	Numeric value	Measures the proportion of the variance in the dependent variable that is predictable from the independent variables.
Local Assessment		
Quintile allocation	Percentages	Percentage of correctly allocated predictions to actual quintiles (or allocation of n quintiles deviations).
Context	MSE	Evaluation of the model for sectorally, temporally, or spatially local performance.

This table reports a summary of the performance metrics used to assess the point estimates of trained models. Global and local in this context do refer to full or partial results of the model, not to geographic coverage.

Table I. Performance Assessment Metrics for Machine Learning Models

the average of squared errors, respectively, with MSE placing greater emphasis on larger errors. We use MSE as the metric in our loss function and thus as our main measure of model performance. The R-squared (R^2) measures the proportion of variance in the dependent variable that can be predicted from the independent variables, offering insight into the explanatory power of the model.

Local assessments show how the models perform in their operational contexts. This category includes quintile allocation, which evaluates the model's ability to accurately rank predictions within specific quintile brackets. This could be useful when seeking best-in-class investment strategies (Edmans et al., 2022) or when reducing exposure to highly emitting companies in portfolios (Rink et al., 2024). The context examines the performance of the model in sector, space, and time. This should ensure transparency about performance differences prevalent in other sustainability data sets (Dobrick et al., 2023).

C. *Prediction Uncertainty*

Estimating sustainability data from corporate financial data is a task that is inherently characterized by significant uncertainty. Point estimates generated by classical statistical or machine learning models represent the conditional expectation of the target variable - in our case the conditional mean - but do not provide information about the uncertainty or variability in the possible range of prediction values. However, quantifying this uncertainty is crucial to make reliable statements about the accuracy of the estimation and associated risks, to possibly exclude certain ranges from use (*measure not estimate*), or to opt for more conservative or optimistic values instead of the point estimates of the conditional expectation.

A key challenge in accurately and consistently assessing the uncertainty of and between different models is the construction of valid prediction intervals. Our selection of approaches and methods for generating prediction intervals is, therefore, guided by the goal of achieving (sufficiently) good marginal coverage and conditional coverage.

Marginal coverage refers to the probability that, on average across all predictions, the actual target value lies within the prediction interval. This property ensures that the prediction intervals are correct on average across the entire distribution of input data. It is particularly useful for making consistent statements about uncertainty across different models and is relevant in scenarios where models are applied for multiple predictions. Marginal coverage does not guarantee accurate coverage for every individual prediction, but ensures that the average coverage meets the desired level.

Conditional coverage refers to the probability that the true target value lies within the predicted interval for a specific set of input values. This property is stronger than marginal coverage and is crucial when individual-level uncertainty quantification is needed, such as in company-specific predictions. Achieving conditional coverage requires that the prediction intervals are accurately calibrated for each possible input, capturing the uncertainty for that particular instance. However, this is generally more challenging to achieve, especially in the presence of complex data structures or heteroskedasticity.

In our data-driven approach, where financial data are used as predictors of sustainability-related information, additional challenges arise due to the (completely) unknown relationship between these variables (absence of theory). This uncertainty with respect to the underlying distribution suggests a preference for methods that do not rely on strong assumptions about the distribution.

In the literature on uncertainty quantification, there are various approaches to quantify-

ing the uncertainties of model prediction (Soize, 2017; Abdar et al., 2021). A common method involves using scalar uncertainty measures, such as the standard deviation (error), which provides a general measure of the spread of predictions. However, these methods offer only a global perspective on the underlying uncertainty and typically fail to adequately account for the variability of uncertainty across different areas of the distribution. Moreover, theoretical guarantees regarding marginal coverage and conditional coverage are only available under strong assumptions about the distribution and are typically not empirically validated (Palmer et al., 2022).

Quantile regression (Koenker and Bassett, 1978) offers an alternative approach that allows the calculation of prediction intervals using the estimation of quantiles of the target variable. This method is particularly useful when asymmetric uncertainties in predictions cannot be ruled out. A central advantage of quantile regression is that it asymptotically ensures both marginal coverage and conditional coverage for sufficiently large (approaching infinite) sample sizes (Chernozhukov et al., 2009; Romano et al., 2019).

Recently, conformalized prediction has been developed (Vovk et al., 2005; Lei and Wasserman, 2014; Lei et al., 2018). The underlying methods provide theoretical guarantees for marginal coverage in finite samples without making strong assumptions about the distribution.[3]

We employ a conformalized version of the quantile regression (Romano et al., 2019). This approach retains the favorable properties of quantile regression concerning the adaptivity of the prediction intervals and (asymptotically) conditional coverage, while also providing theoretical guarantees for marginal coverage in finite samples under the assumption of i.i.d. data. As with our systematic approach to identifying the optimal model for predicting the conditional mean, this approach to uncertainty quantification aims to be sufficiently close to optimal to allow meaningful conclusions.

Specifically, our approach involves generating symmetric prediction intervals with a target coverage rate of $\tau \in \{68\%, 95\%\}$ to assess prediction uncertainty. Given the size of the datasets, we employ a split method in which the test data set is divided into a calibration set and a (uncertainty) test set. The training of the models using an adapted loss function is performed on the training set. The calibration set is then used to compute conformal scores, which are subsequently employed to construct prediction intervals that achieve the desired coverage levels. The test set is used to evaluate the coverage and analyze prediction uncertainty. This approach is intended to balance the trade-off between

[3]Our approach builds on work that assumes exchangeability, which is satisfied under the assumption of i.i.d. data. For conformalized prediction approaches beyond exchangeability, see Tibshirani et al. (2019) and Barber et al. (2023).

statistical efficiency and computational efficiency. We implement this approach in four steps:

1. *Quantile Regression:* We train the best base models for each learner as well as the meta models on the training data set using the loss function $L_\alpha(\hat{y}_\alpha, y) = (y - \hat{y}_\alpha)\,\alpha \mathbb{1}\{y > \hat{y}_\alpha\} + (\hat{y}_\alpha - y)\,(1-\alpha)\,\mathbb{1}\{y \leq \hat{y}_\alpha\}$, with $\alpha \in \{(1-\tau)/2, (1+\tau)/2\}$. y is the target variable and $\hat{y}$ the (quantile) prediction of the target variable.

2. *Quantile Prediction:* We then use the trained models to predict the conditional quantiles $\hat{y}_\alpha(x)$ for each observation in the calibration data set. x is a vector of the predictor variables for each observation.

3. *Conformal Scores:* Based on these quantile predictions, we determine conformal scores, $c(x, y) = \max\{\hat{y}_{(1-\tau)/2}(x) - y, y - \hat{y}_{(1+\tau)/2}(x)\}$ for each observation in the calibration data set.

4. *Rectifying Quantiles:* Defining $\hat{r} = \text{Quantile}\left(\frac{\lceil (n_{cal}+1)\tau \rceil}{n_{cal}}, \{c_1, ..., c_{n_{cal}}\}\right)$ we can derive conditional prediction intervals

$$I(x) = \left[\hat{y}_{\tau/2}(x) - \hat{r}, \hat{y}_{1-\tau/2}(x) + \hat{r}\right]$$

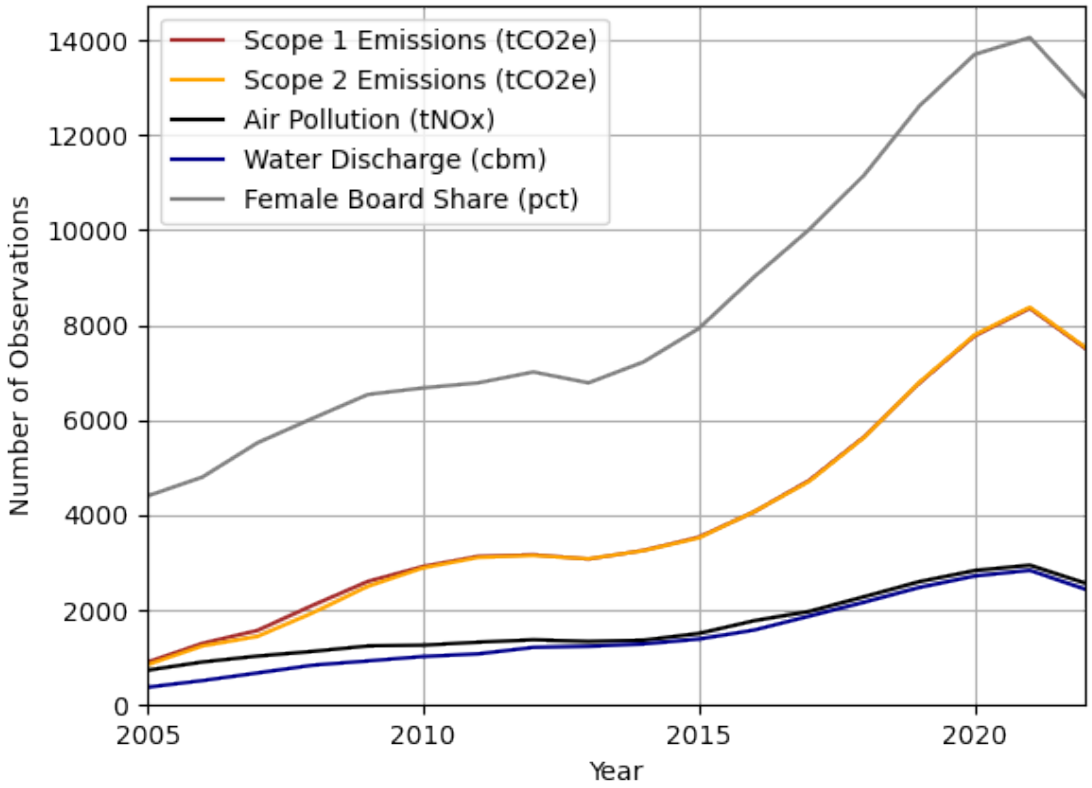
for each observation.

D. Data

Our data set comprises a comprehensive collection of company-year observations for listed equities obtained from the London Stock Exchange Group (LSEG) Data and Analytics database. This data set is global in scope, encompassing data from 95 countries, which provides a diverse and extensive foundation for modeling and analysis. To maintain the integrity and reliability of the results, only reported sustainability data are included in this study, thereby avoiding the potential biases introduced by the LSEG Data and Analytics' estimation models or unaudited data. This conservative approach ensures that the analysis is based solely on real and verifiable metrics.

As shown in Figure 2, the data set includes key sustainability indicators such as Scope 1 and Scope 2 emissions, air pollution, water discharge, and female board share (our "target variables"). Data availability has recently increased substantially, enabling ML applications in the field.

Table II provides a detailed summary of the data, highlighting the breadth and depth of



The figure illustrates the availability of reported sustainability data by listed companies over time.

Figure 2. Reported Sustainability Data over Time

the data set. The data set time period runs from 2005 to 2022.

- Scope 1 and Scope 2 Emissions: The dataset contains nearly 50,000 observations each for these emissions categories, covering 83 sectors across 83 countries, with data available for more than 8,000 companies. We select these indicators to benchmark against other studies and due to the relevance of climate change to business.
- Air Pollution and Water Discharge: These variables have fewer observations, reflecting the more limited availability of environmental data. However, they still provide significant coverage, with data on more than 3,000 companies from more than 60 countries. These indicators are included to reflect emerging topics in business and finance research such as biodiversity, blue economy, and livable cities.
- Female Board Share: This social governance indicator is well-represented, with over 100,000 observations across 95 countries and 86 sectors, providing the richest data set in our analysis. It should demonstrate that our approach is applicable beyond

environmental sustainability data.

The data set includes 212-256 predictor variables, depending on the specific sustainability metric, with data completeness ranging from 57% to 65%. This breadth of variables offers a comprehensive view of company characteristics and operational contexts.

Dataset	Scope 1 Emissions	Scope 2 Emissions	Air Pollution	Water Discharge	Female Board Share
General Information					
Number of observations	47685	47320	20980	18426	108834
Number of sectors	83	83	76	74	86
Number of countries	83	83	63	63	95
Number of companies	8391	8335	3389	3098	14406
Start year	2005	2005	2005	2005	2005
End year	2022	2022	2022	2022	2022
Number of predictor variables	240	240	213	211	255
Data completeness (in %)	63.86	63.77	65.72	65.62	57.92
Target Variable Information					
Mean	3786965	1051686	19914	187082423	16.00
Standard deviation	26440465	48614503	179304	1318742226	14.00
Minimum	0.00	0.00	0.00	0.00	0.00
Maximum	4421000000	7386660000	14042000	26877900000	100.00
Target Variable Information					
Log (1+value) Mean	10.75	11.07	6.31	14.98	2.17
Log (1+value) Std	3.55	2.72	3.28	3.60	1.40
Log (1+value) Min	0.00	0.00	0.00	0.00	0.00
Log (1+value) Max	22.21	22.72	16.46	24.01	4.62

This table presents summary statistics for the dataset, including Scope 1 and Scope 2 emissions, air pollution, water discharge, and female board share. The table details general information such as the number of observations, sectors, countries, companies, the time period (2005-2022), and data completeness rates. It also includes target variable information, such as means, standard deviations, and the range (minimum to maximum) of the absolute and log-transformed values.

Table II. Summary Statistics

The features selected for this study focus on fundamental and financial data, excluding direct sustainability-related metrics (e.g., energy use) to challenge the model's ability to infer sustainability data solely from financial and fundamental data. The feature set includes the following. A complete list of the variables including summary statistics is presented in Appendix E.

- Financial data (e.g., income statement, balance sheet, cash flow metrics), capturing idiosyncratic company characteristics such as size, age, innovative capacity, and asset intensity.
- Fundamental data
  - Industry indicators to account for general emission intensity trends within specific sectors.
  - Spatial variables that reflect policy and socio-economic operating conditions.

  - Company age that reflects the general development stage of a company.

- Temporal effects captured by the inclusion of year variables.

Data preprocessing steps included retrieving data via the LSEG Data and Analytics API, filtering out non-listed equity observations, and excluding entries without date variables or those with quarterly reporting. The target variables were logarithmically transformed to improve predictive performance and for highly correlated features (correlation >= 99%) only one feature is retained to mitigate multicollinearity. Furthermore, features with very low data availability (missingness >= 99%) were excluded, as imputation was not feasible in these cases.

This rigorous data selection and processing framework prior to the pipeline preprocessing steps ensures that the resulting model is robust and reliable, providing information on how financial metrics can be used to predict sustainability data.

III. Results

A. It's in the Financials, Stupid!

Our analysis reveals that sustainability data can be predicted from corporate financial data with a high degree of accuracy. Complex and non-parametric machine learning models significantly outperform linear models in predicting sustainability data from financial data, see Table III. This indicates that the underlying structure of sustainability data is quite complex and cannot be adequately captured by simple linear relationships.

Specifically, our best-performing models achieve mean squared errors (MSE) in the range of 0.6-1.9 and mean absolute errors (MAE) between 0.3 and 0.7. These results are substantially better compared to a benchmark study on corporate carbon emissions by Nguyen et al. (2022), which reported MAE values around 1.1 and 0.8 for Scope 1 and Scope 2 emissions where we achieve 0.6 and 0.5 respectively. Moreover, our models demonstrate high explanatory power. The best performing models consistently explain more than 90% of the variance in the sustainability data sample. This high level of performance underscores the potential of advanced machine learning techniques in enhancing the predictive accuracy of sustainability metrics based on financial data. Hence, it's in the financials, stupid!

The performance of the prediction varies between the different sustainability metrics. The performance of the female board share is weaker in relative terms compared to environmental variables. This indicates that performance cannot always be improved in

Target Variable	Metric	Base Learner				Meta Learner		
		Linear Regression	Ridge	CatBoost	XGBoost	Linear Regression	Ridge	XGBoost
Scope 1 Emissions (tCO2e)	MAE	1.261	1.240	0.648	0.662	0.610	0.609	0.589
	MSE	3.401	3.252	1.579	1.617	1.622	1.622	1.589
	R2	0.713	0.726	0.867	0.864	0.863	0.863	0.866
Scope 2 Emissions (tCO2e)	MAE	1.252	1.118	0.513	0.627	0.556	0.556	0.527
	MSE	3.400	2.959	1.163	1.269	1.275	1.275	1.204
	R2	0.540	0.599	0.842	0.828	0.827	0.827	0.837
Air Pollution (tNOx)	MAE	1.501	1.444	0.695	0.905	0.672	0.672	0.650
	MSE	4.598	4.320	1.923	2.262	1.907	1.907	1.909
	R2	0.564	0.590	0.818	0.786	0.819	0.819	0.819
Water Discharge (cbm)	MAE	1.782	1.782	0.278	0.623	0.275	0.275	0.273
	MSE	7.528	7.522	1.055	1.526	1.063	1.063	1.053
	R2	0.443	0.443	0.922	0.887	0.921	0.921	0.922
Female Board Share (pct)	MAE	0.876	0.876	0.473	0.501	0.448	0.448	0.428
	MSE	1.253	1.249	0.598	0.638	0.595	0.595	0.592
	R2	0.347	0.349	0.688	0.667	0.690	0.690	0.691

This table reports the performance metrics of the machine learning models Linear Regression, Ridge, CatBoost, and XGBoost, used to predict sustainability data. The models are evaluated based on their Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R2) values in five target variables Scope 1 Emissions, Scope 2 Emissions, Air Pollution, Water Discharge, and Female Board Share. Target variable is in log + 1 format.

Table III. Global Model Performance

predicting sustainability data from financial data by increasing the sample size. It also highlights the difficulty that some models seem to have in dealing with truncated data (by definition, the female board share is scaled between 0% and 100%).

The relative errors in the predictions remain non-negligible. To show the applicability of the predicted data to different use cases, we dissect the results at the local level in the next step.

B. Model Performance Varies

In subsequent analyses, we show the results for the best-performing model for each of the sustainability metrics. These are base XGBoost models for Scope 1 and 2 emissions, a meta Ridge model for air pollution, and meta XGBoost models for water discharge and female board share.

Quintile Analysis

First, we evaluate the performance of our models using a quintile-based approach. Specifically, we divide the data set into quintiles with roughly the same number of observations based on the size of the target variable[4]. This allows us to assess how well our models perform across different ranges of sustainability metrics.

The quintile-based approach shows that the models generally perform the worst in the

[4]Note that female board share has nearly 30% of observations equaling zero. Therefore, in this case, we manually adjusted the binning. As a result, more observations are in Quintile 1 in this case.

smallest quintile, which includes the lowest values of the target variables; see Table IV. This pattern is consistent in the five target variables analyzed. A key factor contributing to this performance disparity is the inherent difficulty in predicting values close to zero, which is only necessary in the smallest quintile.

Target Variable	Metric	Q1	Q2	Q3	Q4	Q5
Scope 1 Emissions (tCO2e)	MAE	1.112	0.571	0.476	0.466	0.618
	MSE	3.922	0.903	0.718	0.757	1.593
Scope 2 Emissions (tCO2e)	MAE	1.004	0.377	0.404	0.335	0.446
	MSE	3.517	0.413	0.500	0.405	0.982
Air Pollution (tNOx)	MAE	1.137	0.464	0.451	0.698	0.612
	MSE	4.074	0.780	0.803	2.015	1.861
Water Discharge (cbm)	MAE	0.532	0.222	0.205	0.185	0.221
	MSE	3.146	0.467	0.482	0.362	0.804
Female Board Share (pct)	MAE	0.692	0.301	0.315	0.344	0.401
	MSE	1.376	0.254	0.293	0.331	0.448

This table presents the performance metrics of the best-performing machine learning model across different quintiles of the target variables. The dataset is divided into quintiles based on the size of each sustainability metric Scope 1 Emissions, Scope 2 Emissions, Air Pollution, Water Discharge, and Female Board Share. Performance is measured using Mean Absolute Error (MAE) and Mean Squared Error (MSE) for each quintile, labeled Q1 through Q5, where Quintile 1 (Q1) represents the smallest values of the target variable, while Quintile 5 (Q5) represents the largest values. All input values are in log + 1 format.

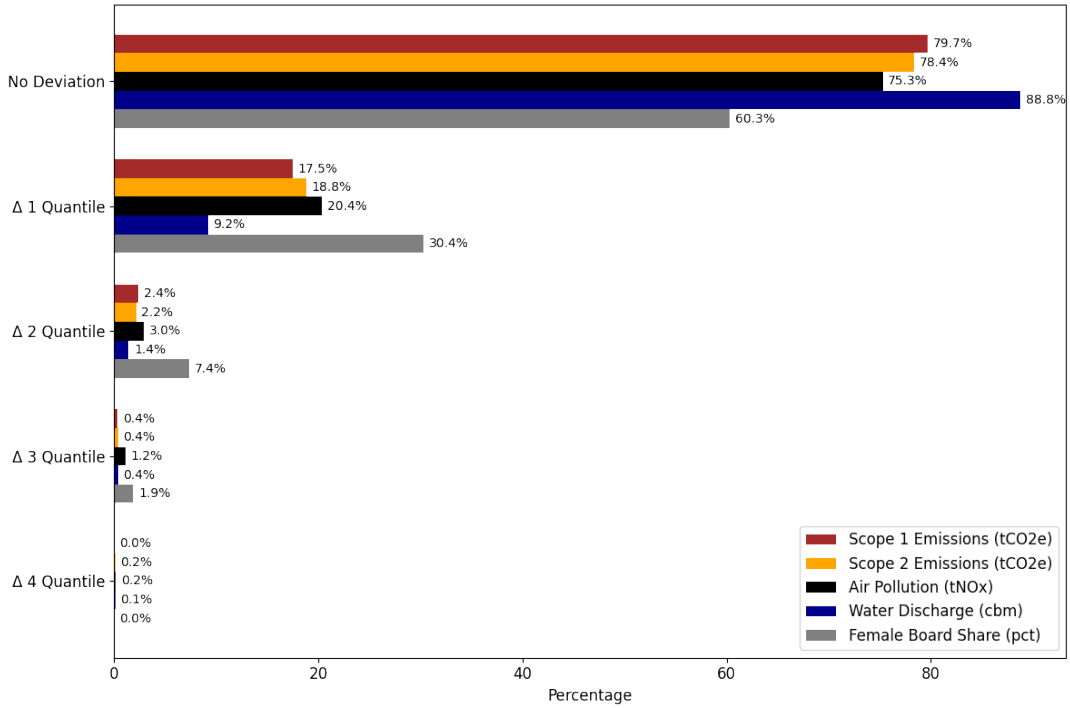
Table IV. Performance of Best Models by Quintiles

The performance variation between the four remaining quintiles is relatively modest, with some target variables showing a slight decrease in performance in the fifth (largest) quintile. We do not observe any particular quintile consistently performing the best or second worst across all metrics, and the deviations in performance are generally small. This suggests that the models maintain a stable performance level in most of the predicted data distributions.

Finally, we categorically examine the degree of deviation between the predicted and actual quintiles. We find that our predictions fall into the right quintile 60.3% (female board share) to 88.8% (Scope 1 emissions) of the time, see Figure 3. In most cases of deviation, the predicted quintile deviates by only one quintile from the actual value. This underscores the overall robustness and performance of our models for categorizing companies based on our continuous predictions.

In practical terms, for an MNE, these findings indicate that the models are reliable for identifying high-impact areas within GVCs. This can be particularly useful for companies seeking to prioritize their strategic focus and allocate resources effectively to areas with the greatest impact on sustainability.

Next, we analyze the performance of the model in the temporal, spatial, and sectoral



The pie charts display the proportion of observations of absolute deviating quintiles for the key variables: Scope 1 and Scope 2 emissions, air pollution, water discharge, and female board share.

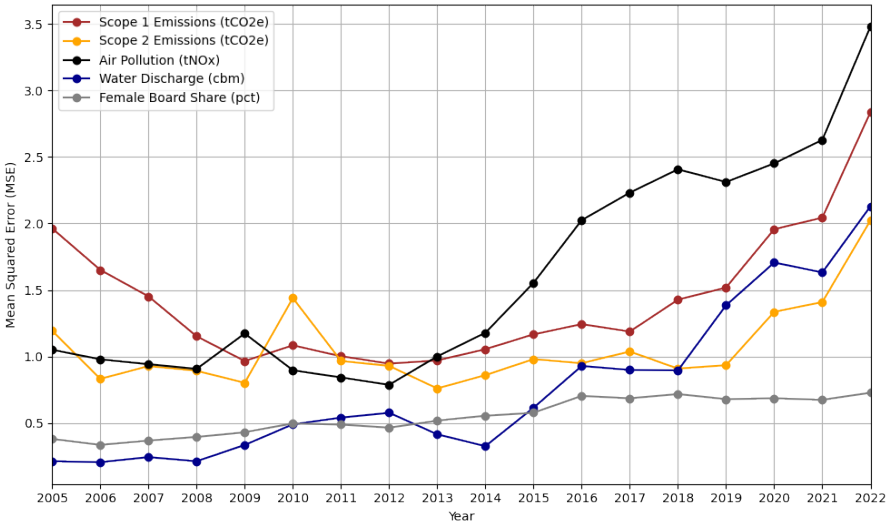
Figure 3. Number of Deviating Quintiles

dimensions to assess how our models perform under different conditions.

Temporal Dimension

Our analysis indicates that the models perform best during the period up to 2015, with a slight decrease in performance in subsequent years; see Figure 4. This decline is relatively minor for all target variables except air pollution. We attribute it to the increased variability in the data set over time, that is, more companies reporting their sustainability data. Although more data should generally help in training machine learning models, higher heterogeneity among reporting companies (business models, sizes, technology mix, etc.) make it more difficult for the algorithm to predict with the same performance level, especially if the heterogeneity grows faster than the data availability.

When applying machine learning models to real-world scenarios, it is essential to consider these temporal variations. Users should ensure that the specific use case is well understood so that the models can be appropriately adapted. Particularly when focusing on current data, it might be necessary to apply variations or modifications to the models trained on older data. One potential approach is to incorporate a discount factor in the loss function that adjusts for temporal discrepancies.



This figure illustrates the model performance over time (2005–2022) for five different variables: Scope 1 emissions, Scope 2 emissions, air pollution, water discharge, and female board share. The ordinate represents the MSE, showing how prediction performance has evolved for each variable across the years.

Figure 4. Temporal Model Performance

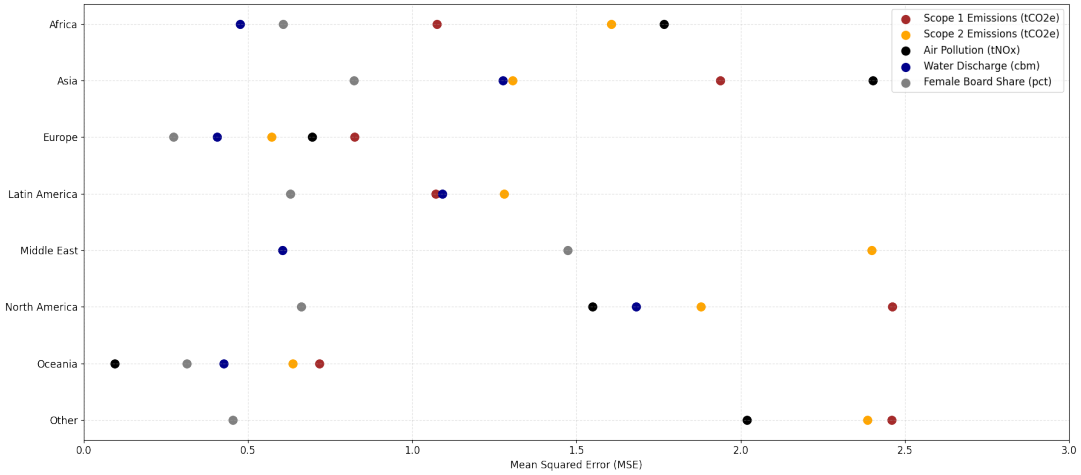
Spatial Dimension

In the spatial analysis, we observe some variation in model performance between different geographic regions; see Figure 5. In particular, our models tend to perform better in Europe (and Oceania) than in other parts of the world. This discrepancy could be partially attributed to stricter and more comprehensive reporting regimes in Europe, resulting in a larger and more reliable dataset (Krueger et al., 2024), that is, reporting is more homogeneous due to mandatory guidelines that improve model performance in these regions (or reduce heterogeneity). However, despite these differences, the models still perform well in all regions. For MNEs looking to optimize global supply chains, our approaches are applicable, although some caution is advised.

Sectoral Dimension

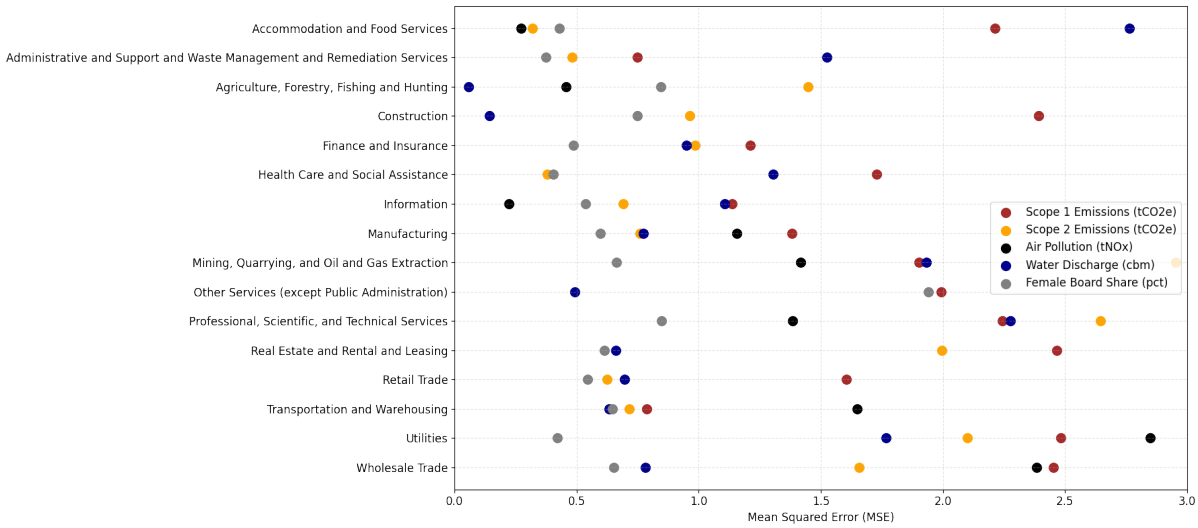
Regarding the sectoral dimension, we find minor variations in model performance across different industries, with some outliers; see Figure 6. For most industries, the models perform reliably, suggesting that they are well suited for applications that involve analyzing portfolios spanning multiple sectors and industries, such as those of banks or asset managers. However, if a particular sector is of special importance to an economic actor, it may be beneficial to use a weighted loss function tailored to sectoral importance.

In summary, while there is some variation in the model performance across the temporal, spatial, and sectoral dimensions, the models generally perform well in all three areas.



This figure shows the model performance across different global regions, as measured by the MSE for the five target variables: Scope 1 emissions, Scope 2 emissions, air pollution, water discharge, and female board share. Each region is represented along the ordinate, while the MSE values are plotted as dots.

Figure 5. Spatial Model Performance



This figure displays the MSE of model predictions across industries for the five target variables: Scope 1 emissions, Scope 2 emissions, air pollution, water discharge, and female board share. Each dot represents the MSE for a particular variable within a specific industry.

Figure 6. Sectoral Model Performance

However, the variation underscores the importance of understanding the specific use case and tailoring the models accordingly to ensure optimal performance for the task at hand.

C. Beware The Prediction Uncertainty

The analysis of prediction uncertainty across the target variables reveals heterogeneous patterns. For smaller coverage requirements, the uncertainty is generally negligible,

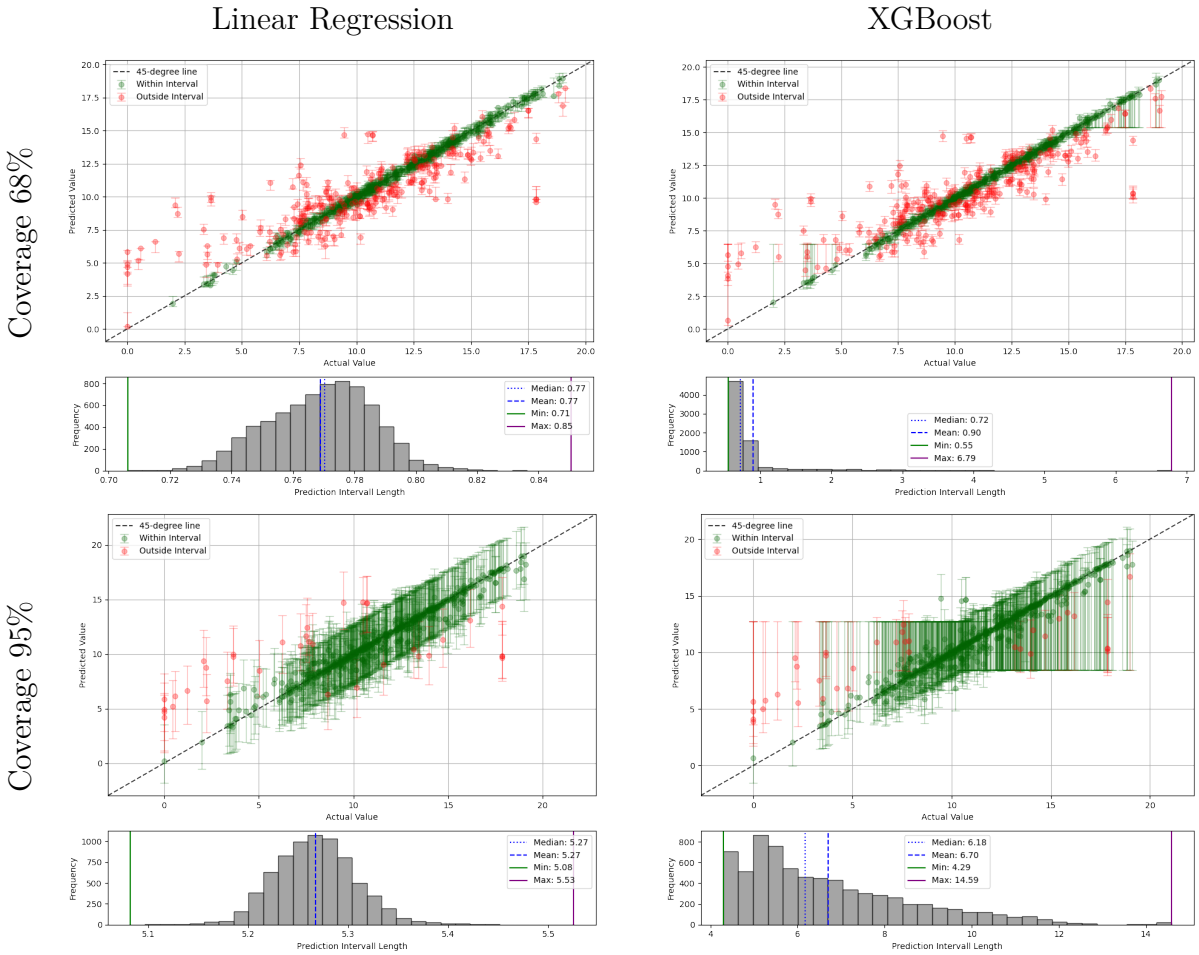
whereas for larger coverage requirements, it becomes significant. Our conformalized quantile regression approach demonstrates empirically valid coverage (with an absolute difference between empirical and target coverage $\leq 0.68\%$ for all variables and coverage rates). The prediction uncertainty remains largely symmetrical, balancing under- or overestimation of prediction uncertainty compared to other standard uncertainty measures. In this section, we show the results for Scope 1 Emissions; see Appendix D for the results for the other target variables.

Figure 7 illustrates the prediction uncertainty in four settings, including conformalized quantile regression using linear regression and XGBoost and with coverage rates of 68% and 95%. The figure demonstrates that, while prediction uncertainty always exists, it is relatively low for the 68% coverage rate. This suggests that users who are not highly risk-averse can use the point estimates without encountering (on average) significant uncertainty. However, at the 95% coverage rate, uncertainty increases substantially, indicating a wider range of possible outcomes, which may be of concern to users who require greater confidence in their predictions.

In addition, the choice of the learner to estimate the uncertainty affects the results. Linear regression models tend to produce on average narrower prediction intervals with relatively constant interval lengths across the distribution of the target variable. In contrast, XGBoost exhibits a more adaptive behavior, with wider prediction intervals towards the extremes of the target variable values. For lower actual values, the uncertainty intervals become larger, particularly skewing toward values above the actuals. Similarly, for the largest values, the intervals widen, though skewed downward, toward lower-than-actual values.

Figure 8 further examines the distribution of the prediction intervals, confirming their general symmetries. In this figure, we use predictions of the conditional mean from the (best performing) XGBoost meta-leaner. Linear regression models show tighter distributions of interval lengths, while XGBoost produces more widely spread intervals.

This figure also presents alternative uncertainty measures for comparison. In addition to the conformalized quantile-based measure, a more naive approach is shown using simple one- and two-standard deviations (σ) from the conditional mean, which are typically used for coverage levels 68% and 95%. This naive approach results in significantly larger intervals compared to the conformalized quantile-based method in most settings. In addition, an alternative method has been employed that uses (standard) quantile regression (ν). This method underestimates the uncertainty compared to our conformalized approach most of the time. This leads us to conclude that the conformalized quantile-based



This figure displays the prediction uncertainty for Scope 1 Emissions. The settings are for the learners linear regression and XGBoost and the target coverage rate of 68% and 95%. For better readability, the actual vs. predicted plots display only 2% of the total observations, randomly selected.

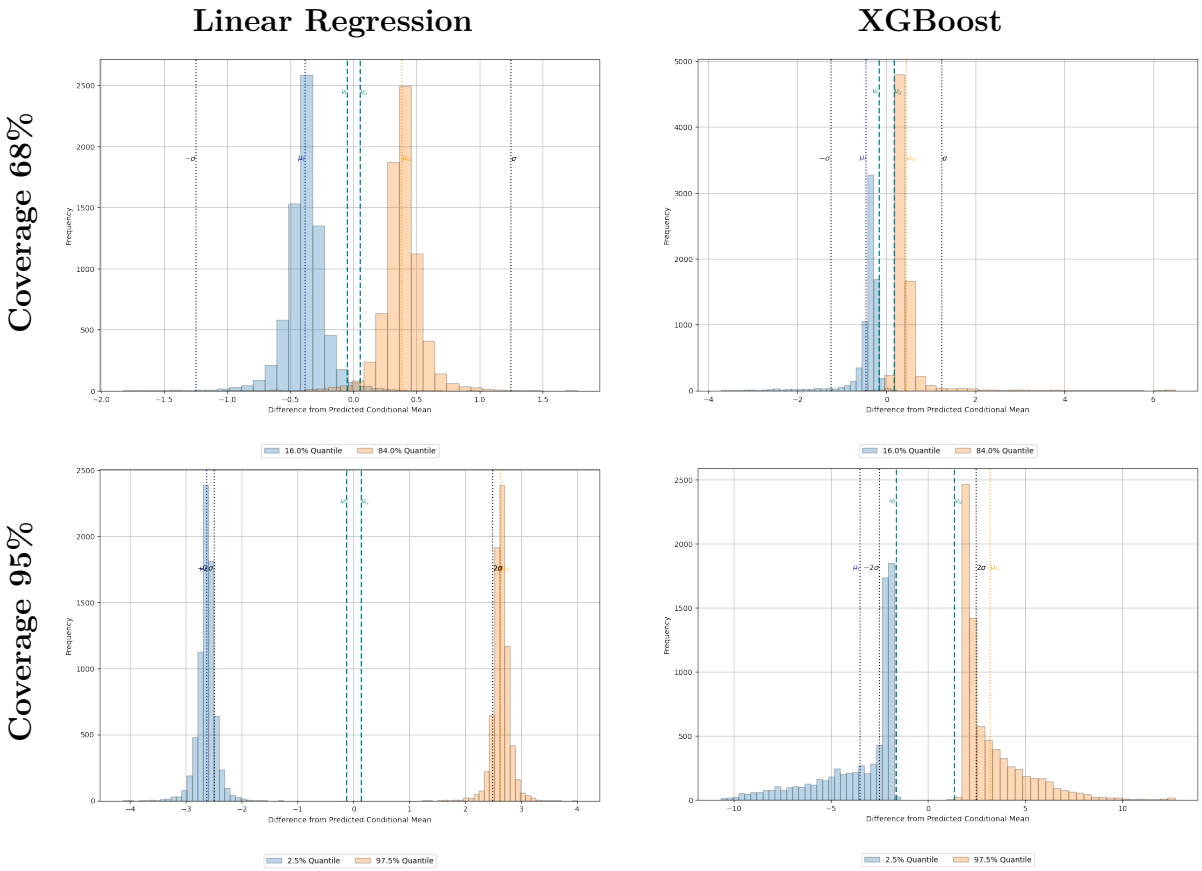
Figure 7. Prediction Uncertainty in Different Settings for Scope 1 Emissions

method is well suited to test prediction uncertainty.

The results of the uncertainty measures for other target variables than Scope 1 emissions show consistent results for environmental variables; however, for the variable representing the female board share (see Appendix D), which is truncated between zero and one, the uncertainty distributions differ, reflecting the specific characteristics of the data.

IV. Conclusion

This study demonstrates that ML can serve as an effective tool for generating company-level sustainability data using only financial data. The findings of this research have several important implications for policymakers, MNEs, and the financial industry.



This figure displays the deviation from the predicted conditional mean for Scope 1 Emissions in different settings. The settings are for uncertainty estimates based on linear regression and XGBoost and for targeted coverage rates of 68% and 95%. In all settings, the conditional mean is predicted by the best-performing XGB meta-model. The black dotted lines represent the prediction intervals using one and two standard deviations (σ), respectively. The green dashed lines represent the mean lower (upper) quantile prediction from standard quantile regression, ν_l (ν_u). Finally, the blue and orange bars show the distribution of quantile predictions for the lower and upper quantile from conformalized quantile regression. The mean of lower (upper) predictions is represented by the blue and orange dotted lines, μ_l (μ_u).

Figure 8. Deviation from Conditional Mean for Scope 1 Emissions

First, policymakers should reconsider the prevailing trend of relying solely on sustainability-related raw data within industry and the financial systems. Although the use of raw data has its merits, particularly in ensuring accuracy, there are instances where the transaction costs associated with obtaining such data are prohibitively high. In these cases, the use of machine learning-generated data offers a viable alternative to using generalized metrics, such as industry averages. This approach can improve the granularity and relevance of sustainability data for decision making.

Second, the study highlights the potential variations and uncertainties inherent in any prediction model, particularly those related to sustainability data. To address this, there is a need for increased transparency requirements for both data providers and companies

when using machine learning or other modeling techniques to generate sustainability data. This transparency is crucial to ensure that the uncertainties associated with these models are well understood and that any biases in the underlying data are clearly communicated. Policymakers could enhance transparency standards, similar to the current efforts for ESG ratings in the European Union (General Secretariat of the Council, 2024), and extend these standards throughout the industry to ensure consistency and reliability.

Third, our research emphasizes the importance of considering the specific use case when applying machine learning models. The performance of these models can vary significantly depending on temporal, spatial, or sectoral factors, which must be taken into account to ensure the accuracy and relevance of the generated data. Adjustments to the loss function can improve the suitability of the model for specific applications, but these considerations should be made transparent to users, illustrating the methodology behind the data generation process.

It is important to acknowledge the limitations of our research. In particular, questions remain regarding the generalizability of our models. The models developed in this study are trained primarily on data from MNEs, which may not be directly applicable to government organizations, non-profit entities, or small and medium-sized enterprises (SMEs). These other parts of the economy may require different data sets, and it remains uncertain whether sufficient data is available to support the development of ML models in these contexts. Future research should explore these gaps, potentially identifying alternative approaches to address the limitations in data availability for these use cases.

In conclusion, this study illustrates that machine learning has a significant role to play in making sustainability data more accessible and cost-effective. By improving the availability and quality of sustainability data through ML, companies and policymakers can make better informed decisions, ultimately advancing sustainability initiatives throughout the global economy.

References

Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., Makarenkov, V., Nahavandi, S., 2021. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. doi:10.1016/j.inffus.2021.05.008.

Acemoglu, D., 1997. Training and Innovation in an Imperfect Labour Market. The Review of Economic Studies 64, 445. doi:10.2307/2971723.

Adams, R.B., Ferreira, D., 2009. Women in the boardroom and their impact on governance and performance. Journal of Financial Economics 94. doi:10.1016/j.jfineco.2008.10.007.

Aggarwal, C.C., 2013. Outlier analysis. volume 9781461463962. doi:10.1007/978-1-4614-6396-2.

Balasubramanian, N., Lee, J., 2008. Firm age and innovation. Industrial and Corporate Change 17. doi:10.1093/icc/dtn028.

Barber, R.F., Candès, E.J., Ramdas, A., Tibshirani, R.J., 2023. Conformal prediction beyond exchangeability. The Annals of Statistics 51. doi:10.1214/23-AOS2276.

Bauer, R., Ruof, T., Smeets, P., 2021. Get Real! Individuals Prefer More Sustainable Investments. Review of Financial Studies 34. doi:10.1093/rfs/hhab037.

Berg, F., Kölbel, J.F., Rigobon, R., 2022. Aggregate Confusion: The Divergence of ESG Ratings. Review of Finance 26. doi:10.1093/rof/rfac033.

Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U., 1998. When is "nearest neighbor" meaningful?, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). doi:10.1007/3-540-49257-7{_}15.

Bolton, P., Kacperczyk, M., Samama, F., 2022. Net-Zero Carbon Portfolio Alignment. Financial Analysts Journal 78. doi:10.1080/0015198X.2022.2033105.

Bosma, B., van Witteloostuijn, A., 2024. Machine learning in international business. Journal of International Business Studies 55, 676–702. doi:10.1057/s41267-024-00687-6.

Breiman, L., 1996. Bagging predictors. Machine Learning 24. doi:10.1007/bf00058655.

Busch, T., Johnson, M., Pioch, T., 2022. Corporate carbon performance data: Quo vadis? Journal of Industrial Ecology 26. doi:10.1111/jiec.13008.

van Buuren, S., 2007. Multiple imputation of discrete and continuous data by fully conditional specification. Statistical Methods in Medical Research 16. doi:10.1177/0962280206074463.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. doi:10.1145/2939672.2939785.

Chernozhukov, V., Hansen, C., Jansson, M., 2009. Finite sample inference for quantile regression models. Journal of Econometrics 152. doi:10.1016/j.jeconom.2009.01.004.

Dobrick, J., Klein, C., Zwergel, B., 2023. Size bias in refinitiv ESG data. Finance Research Letters 55, 104014. doi:10.1016/j.frl.2023.104014.

ECB, 2022. 2022 climate risk stress test. Technical Report. European Central Bank. Frankfurt am Main. URL: https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.climate_stress_test_report.20220708~2e3cc0999f.en.pdf, doi:10.2866/97350.

Edmans, A., 2023. Applying Economics—Not Gut Feel—to ESG. Financial Analysts Journal doi:10.1080/0015198X.2023.2242758.

Edmans, A., Levit, D., Schneemeier, J., 2022. Socially Responsible Divestment. SSRN Electronic Journal doi:10.2139/ssrn.4093518.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? Journal of Machine Learning Research 15.

Garel, A., Romec, A., Sautner, Z., Wagner, A.F., 2023. Do Investors Care About Biodiversity? SSRN Electronic Journal doi:10.2139/ssrn.4398110.

General Secretariat of the Council, 2024. Proposal for a Regulation of the European Regulation and the Council on the transparency and integrity of Environmental, Social and Governance (ESG) rating activities, and amending Regulation (EU) 2019 2088.

Giglio, S., Kuchler, T., Stroebel, J., Zeng, X., 2023. Biodiversity Risk. SSRN Electronic Journal doi:10.2139/ssrn.4420552.

Goldhammer, B., Busse, C., Busch, T., 2017. Estimating Corporate Carbon Footprints with Externally Available Data. Journal of Industrial Ecology 21. doi:10.1111/jiec.12522.

Griffin, P.A., Lont, D.H., Sun, E.Y., 2017. The Relevance to Investors of Greenhouse

Gas Emission Disclosures. Contemporary Accounting Research 34. doi:10.1111/1911-3846.12298.

Hoerl, A.E., Kennard, R.W., 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12. doi:10.1080/00401706.1970.10488634.

Iguyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection.

Ilhan, E., Krueger, P., Sautner, Z., Starks, L.T., 2023. Climate Risk Disclosure and Institutional Investors. Review of Financial Studies 36. doi:10.1093/rfs/hhad002.

Ilhan, E., Sautner, Z., Vilkov, G., 2021. Carbon Tail Risk. Review of Financial Studies 34. doi:10.1093/rfs/hhaa071.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. doi:10.1126/science.aaa8415.

Khan, H.u.R., Bin Khidmat, W., Hammouda, A., Muhammad, T., 2023. Machine learning in the boardroom: Gender diversity prediction using boosting and undersampling methods. Research in International Business and Finance 66, 102053. doi:10.1016/j.ribaf.2023.102053.

Kim, D., Starks, L.T., 2016. Gender Diversity on Corporate Boards: Do Women Contribute Unique Skills? American Economic Review 106, 267–271. doi:10.1257/aer.p20161032.

Koenker, R., Bassett, G., 1978. Regression Quantiles. Econometrica 46, 33. doi:10.2307/1913643.

Krueger, P., Sautner, Z., Tang, D.Y., Zhong, R., 2024. The Effects of Mandatory ESG Disclosure Around the World. Journal of Accounting Research doi:10.1111/1475-679X.12548.

Kunming Declaration, 2021. Declaration from the High-Level Segment of the UN Biodiversity Conference 2020 under the Theme: Ecological Civilization: Building a Shared Future for All Life on Earth. Technical Report. URL: https://www.cbd.int/doc/c/c2db/972a/fb32e0a277bf1ccfff742be5/cop-15-05-add1-en.pdf.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. doi:10.1038/nature14539.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R.J., Wasserman, L., 2018. Distribution-Free Predictive Inference for Regression. Journal of the American Statistical Association 113. doi:10.1080/01621459.2017.1307116.

Lei, J., Wasserman, L., 2014. Distribution-free prediction bands for non-parametric regression. Journal of the Royal Statistical Society. Series B: Statistical Methodology 76. doi:10.1111/rssb.12021.

Loughran, T., McDonald, B., 2023. Measuring Firm Complexity. Journal of Financial and Quantitative Analysis , 1–28doi:10.1017/S0022109023000716.

Marano, V., Wilhelm, M., Kostova, T., Doh, J., Beugelsdijk, S., 2024. Multinational firms and sustainability in global supply chains: scope and boundaries of responsibility. Journal of International Business Studies 55, 413–428. doi:10.1057/s41267-024-00706-6.

Nguyen, Q., Diaz-Rainey, I., Kitto, A., Mcneil, B., Pittman, N.A., Zhang, R., 2022. Scope 3 Emissions: Data Quality and Machine Learning Prediction Accuracy. URL: https://deliverypdf.ssrn.com/delivery.php?ID=735105085006000008086072113005079106120009055009062036075099117085093102127079083094C pdf&INDEX=TRUE.

Nguyen, Q., Diaz-Rainey, I., Kuruppuarachchi, D., 2021. Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. Energy Economics 95. doi:10.1016/j.eneco.2021.105129.

Palmer, G., Du, S., Politowicz, A., Emory, J.P., Yang, X., Gautam, A., Gupta, G., Li, Z., Jacobs, R., Morgan, D., 2022. Calibration after bootstrap for accurate uncertainty quantification in regression models. npj Computational Materials 8. doi:10.1038/s41524-022-00794-8.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. Catboost: Unbiased boosting with categorical features, in: Advances in Neural Information Processing Systems.

Ranta, M., Ylinen, M., 2023. Board gender diversity and workplace diversity: a machine learning approach. Corporate Governance (Bingley) 23. doi:10.1108/CG-01-2022-0048.

Rink, S., Matheis, Y., Dumrose, M., 2024. Private Capital under the Paris Agreement: Institutional Investors and Company Decarbonization. Unpublished Work .

Romano, Y., Patterson, E., Candès, E.J., 2019. Conformalized quantile regression, in: Advances in Neural Information Processing Systems.

Slager, R., Chuah, K., Gond, J.P., Furnari, S., Homanen, M., 2023. Tailor-to-Target: Configuring Collaborative Shareholder Engagements on Climate Change. Management Science doi:10.1287/mnsc.2023.4806.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms, in: Advances in Neural Information Processing Systems.

Soize, C., 2017. Uncertainty Quantification. volume 47. Springer International Publishing, Cham. doi:10.1007/978-3-319-54339-0.

Starks, L.T., 2023. Presidential Address: Sustainable Finance and ESG Issues—Value versus Values. Journal of Finance 78, 1837–1872. doi:10.1111/jofi.13255.

Tian, M., 2023. Impact of Climate Water Risk on Corporate Operational and Capital Markets Performance: A Machine Learning Approach. Ph.D. thesis. University of Michigan. URL: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/192421/mytian_1.pdf?sequence=1&isAllowed=y.

Tibshirani, R.J., Barber, R.F., Candès, E.J., Ramdas, A., 2019. Conformal prediction under covariate shift, in: Advances in Neural Information Processing Systems.

United Nations, 2015. Paris Agreement - UNFCCC. URL: https://unfccc.int/process-and-meetings/the-paris-agreement.

Vovk, V., Gammerman, A., Shafer, G., 2005. Algorithmic learning in a random world. doi:10.1007/b106715.

Wolpert, D., Macready, W., 1997. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation 1, 67–82. doi:10.1109/4235.585893.

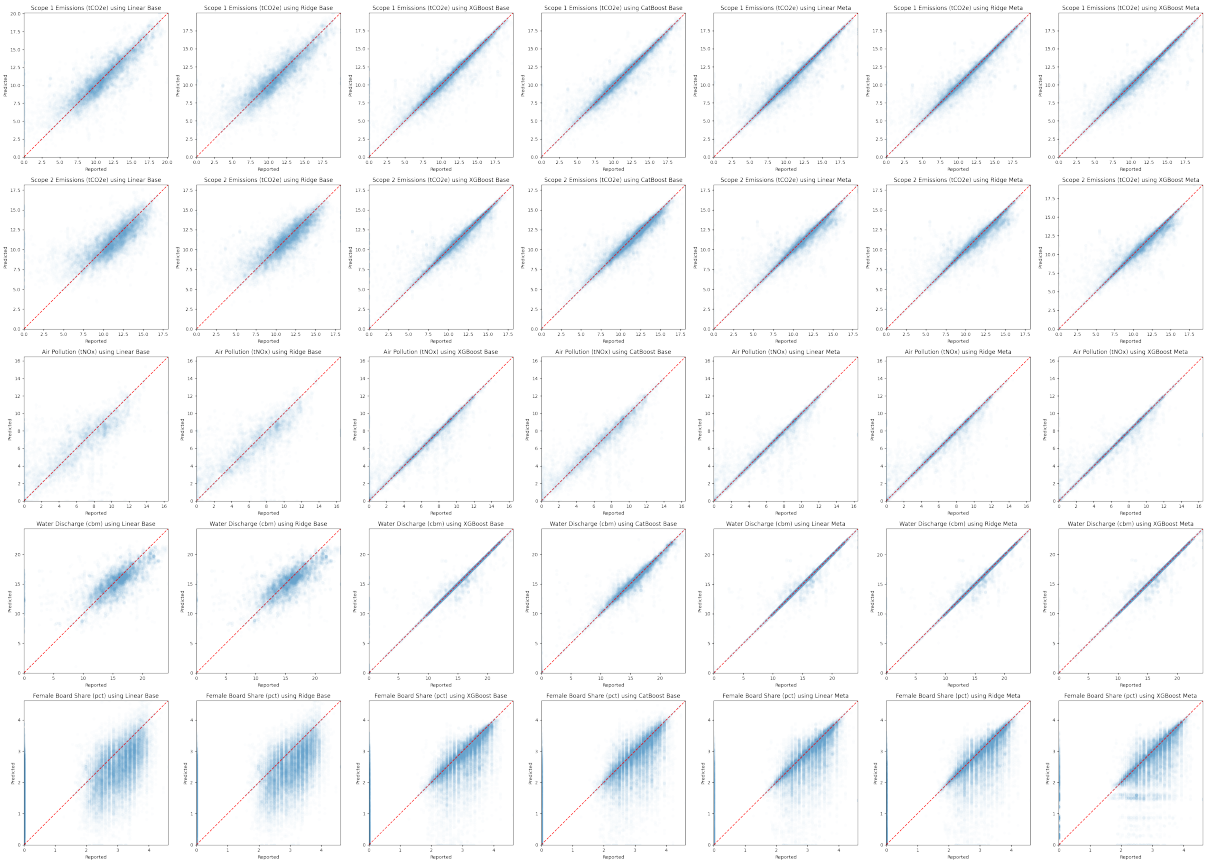
Wolpert, D.H., 1992. Stacked generalization. Neural Networks 5. doi:10.1016/S0893-6080(05)80023-1.

A. Additional Results

In this annex, we provide supplementary graphical representations of our model performance results.

A. Model Performance: Reported vs. Predicted Values

Figure 9 presents a comparison between reported and predicted values within our test data set for both baseline and meta-models in the five sustainability metrics. This comparison reaffirms that more complex models, such as XGBoost and CatBoost, as well as the meta-models, consistently outperform linear models like OLS and Ridge by a significant margin.

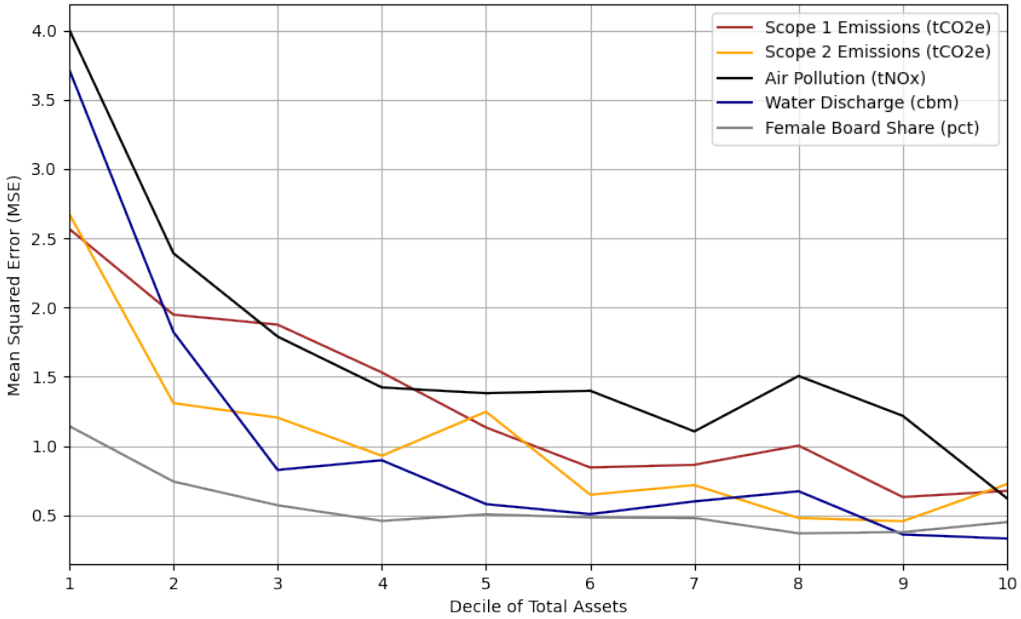


This figure displays the plots of the reported vs predicted variables for the five target variables vertically (Scope 1 emissions, Scope 2 emissions, air pollution, water discharge, and female board share) and the best performing model per learner horizontally (linear regression, ridge, CATBoost, XGBoost, meta-linear, meta-ridge, meta-XGBoost).

Figure 9. Plots reported vs predicted for best model per target variable and learner

B. Model Performance by Company Size

Figure 10 illustrates the variation in model performance relative to company size, represented by deciles of company size and the corresponding mean squared error (MSE) per decile. The results indicate that model performance is notably poorer in the lowest decile, which includes the smallest companies. Despite this, the performance remains superior compared to previous studies even in this decile. Following the initial decile, there is a sharp improvement in model performance, which then levels off with only minor variations around deciles 7 to 9.

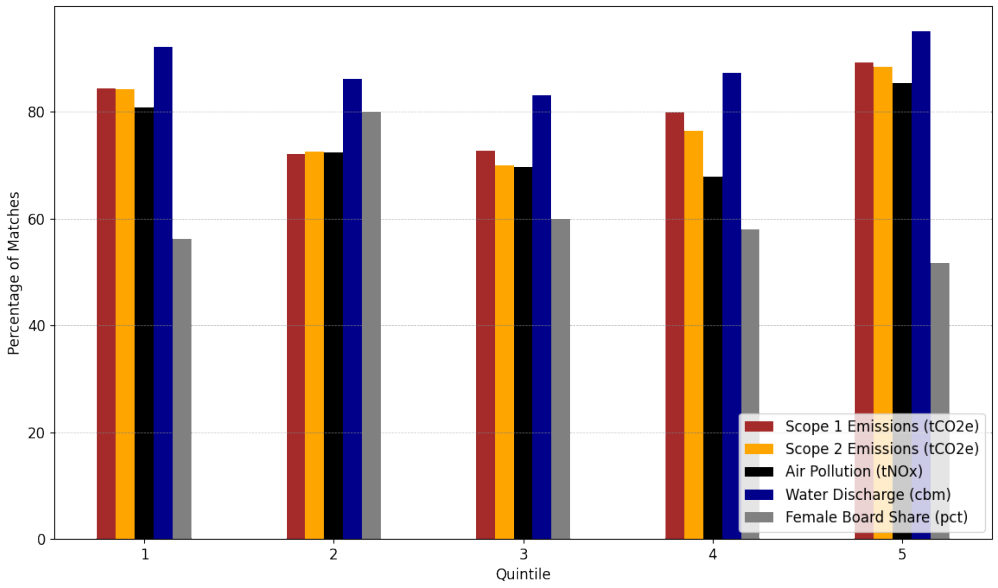


This figure displays the MSE of model predictions for company-size deciles for the the five target variables: Scope 1 emissions, Scope 2 emissions, air pollution, water discharge, and female board share.

Figure 10. Model Performance by Company Size

C. Quintile Analysis of Sustainability Metrics

Figure 11 presents the quintile results, also discussed in the main body of the paper, showing the percentage of matches for each sustainability metric within their respective quintile brackets. The primary finding is that the variation in model fit across quintiles is generally minimal for most sustainability metrics, with the exception of the female board share metric, where substantial variation is observed among the quintiles. This suggests that the predictive accuracy for the female board share is more sensitive to the distribution of data between different quintiles compared to other sustainability metrics.

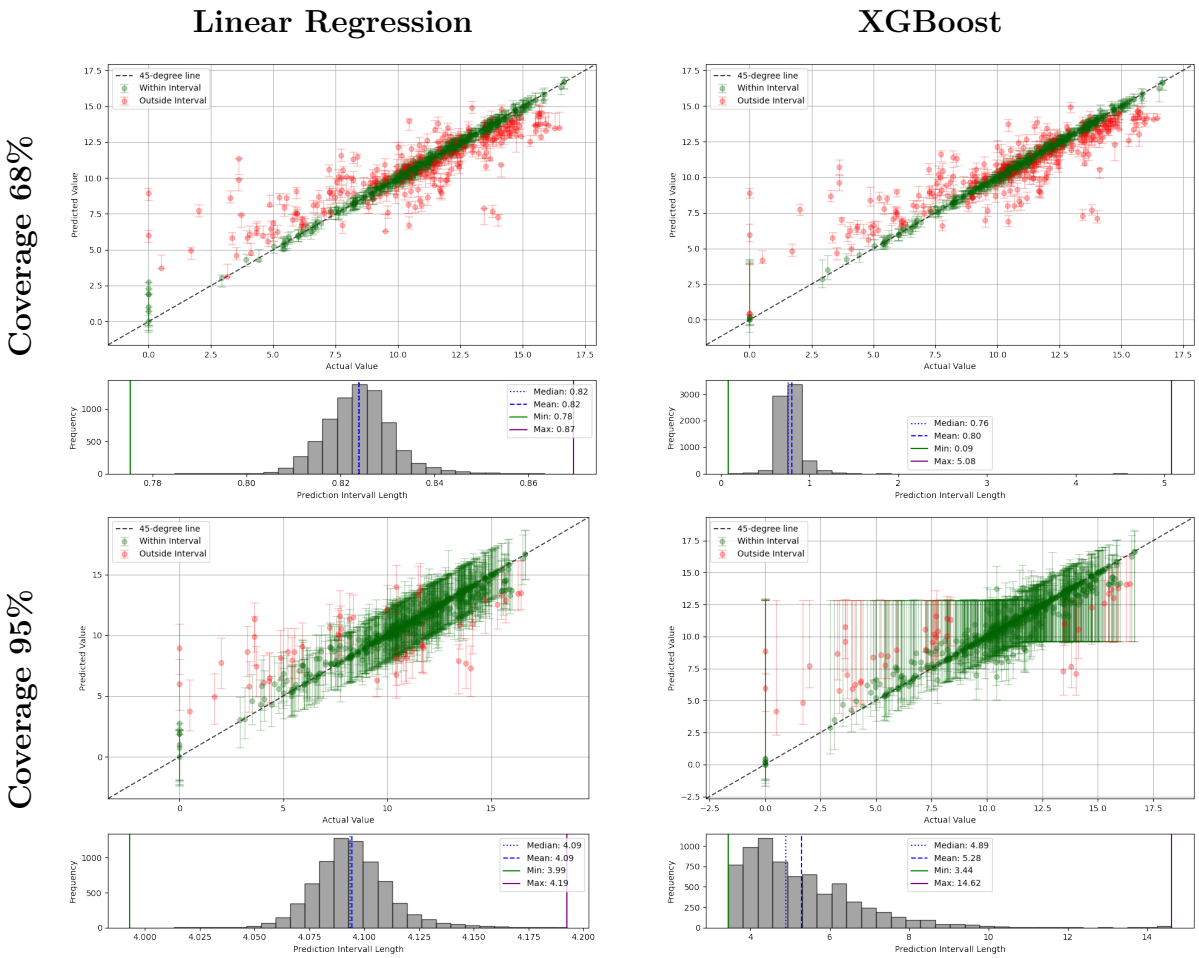


This figure displays the percentage of correctly allocated observations (predicted vs actual) per quintile for the five target variables: Scope 1 emissions, Scope 2 emissions, air pollution, water discharge, and female board share.

Figure 11. Percentage of Quintile Matches

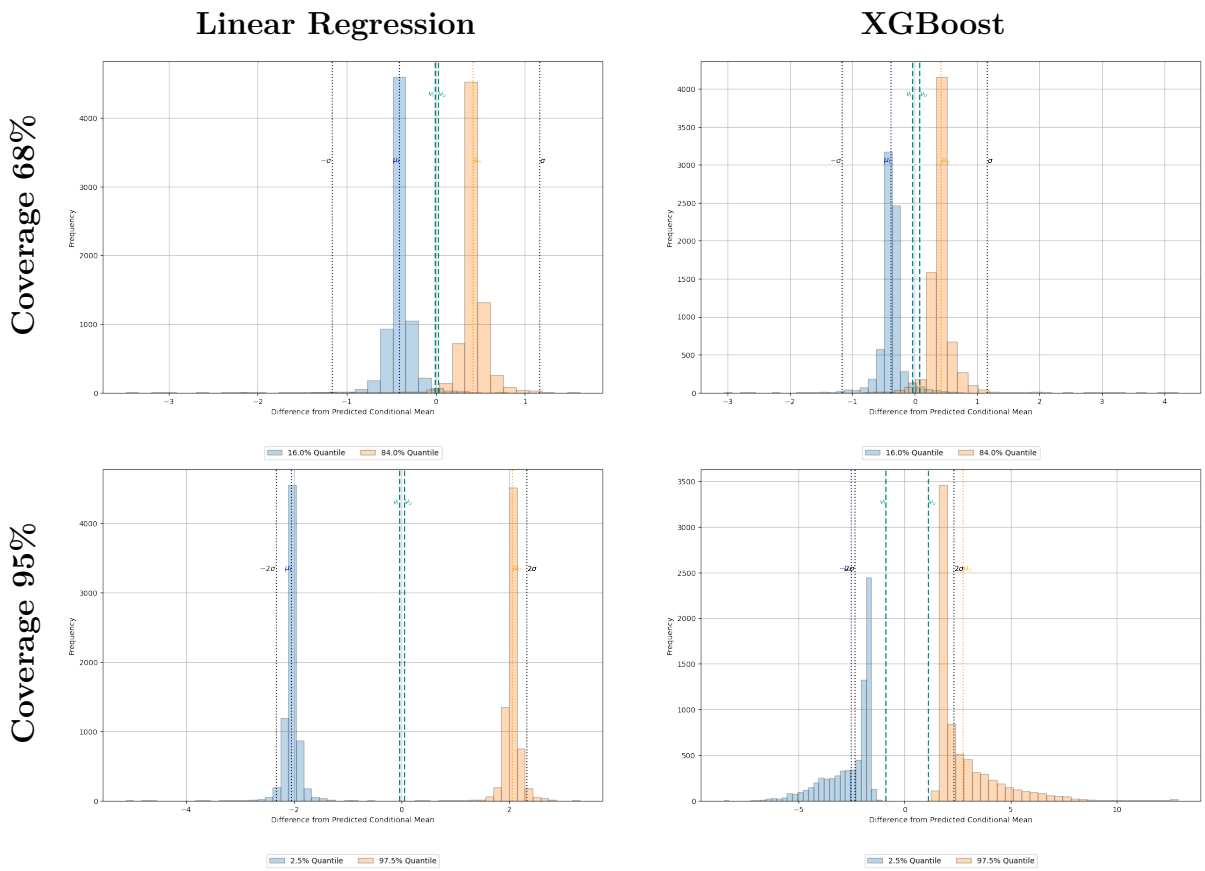
D. Prediction Uncertainty for Remaining Target Variables

Scope 2 Emissions



This figure displays conformalized prediction uncertainty for Scope 2 Emissions. The settings are for the learners linear regression and XGBoost and the coverage of 68% and 95%. For better readability, the figure shows 2% of total observations which are randomly selected.

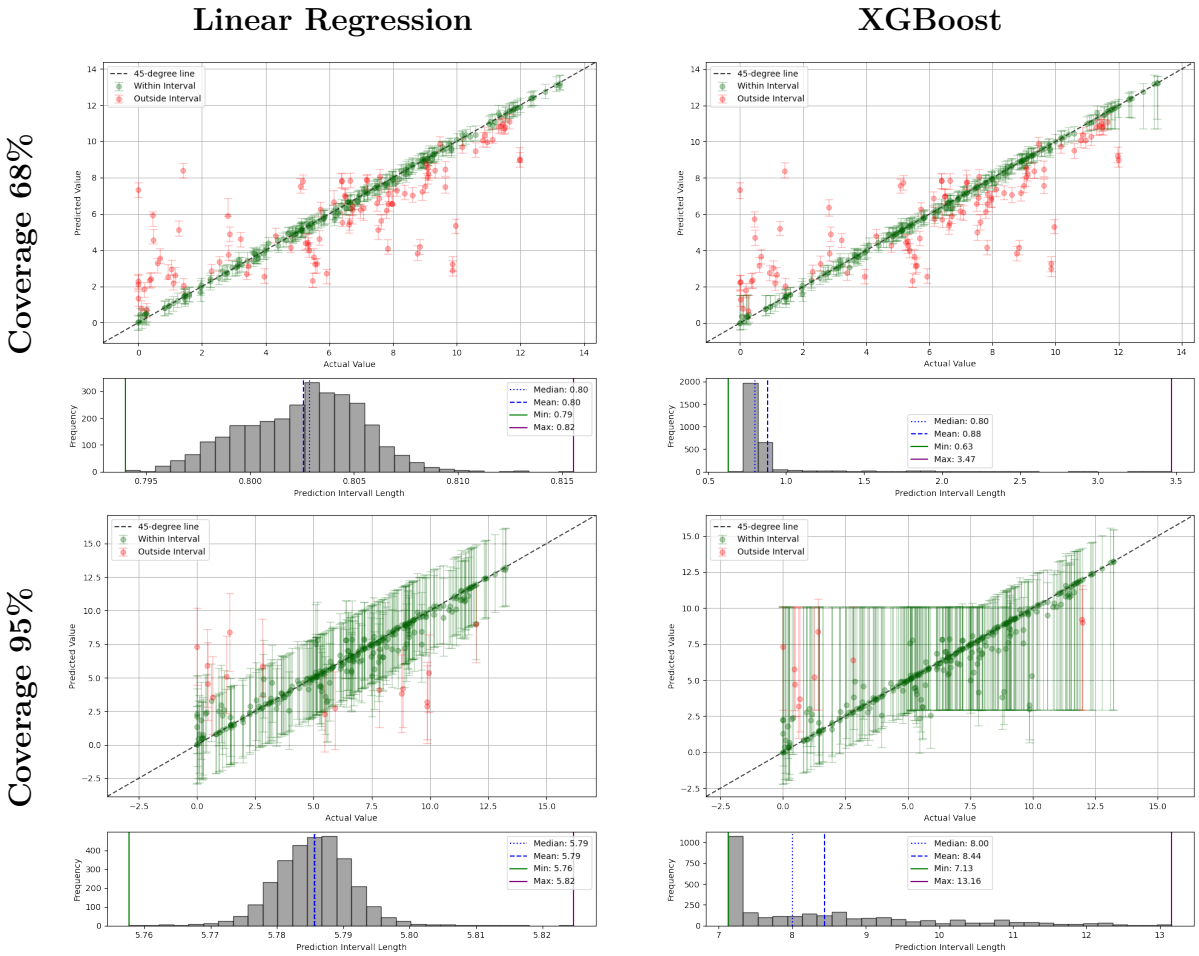
Figure 12.
Conformalized Prediction Uncertainty in Different Settings for Scope 2 Emissions



This figure displays the deviation for the conditional mean for Scope 2 Emissions in different settings. The settings are for uncertainty estimates based on linear regression and XGBoost and for targeted coverage rates of 68% and 95%. In all settings, the conditional mean is predicted by the best-performing XGB meta-model. The black dotted lines represent the prediction intervals using one and two standard deviations (σ), respectively. The green dashed lines represent the mean lower (upper) quantile prediction from standard quantile regression, ν_l (ν_u). Finally, the blue and orange bars show the distribution of quantile predictions for the lower and upper quantile from conformalized quantile regression. The mean of lower (upper) predictions is represented by the blue and orange dotted lines, μ_l (μ_u).

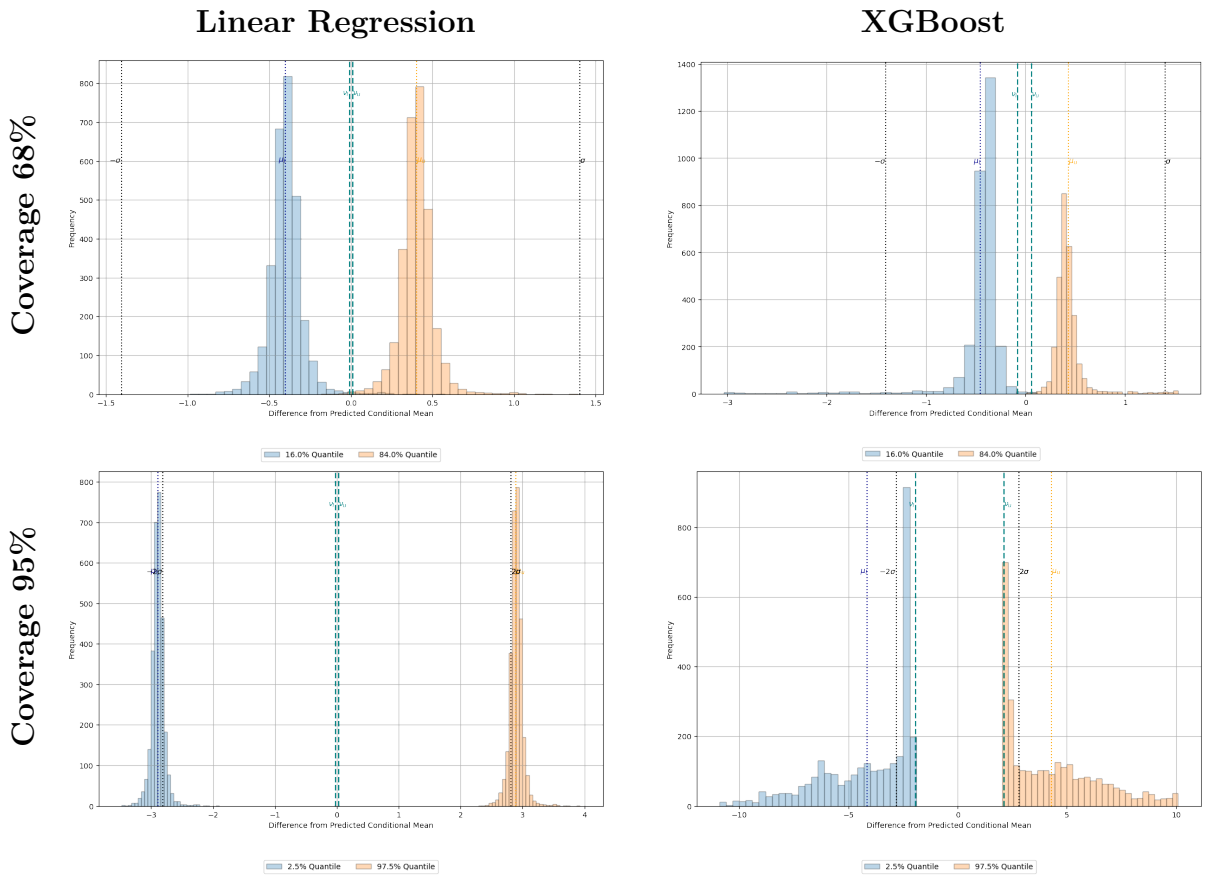
Figure 13. Deviation from Conditional Mean for Scope 2 Emissions

Air Pollution



This figure displays conformalized prediction uncertainty for NOx Emissions. The settings are for the learners linear regression and XGBoost and the coverage of 68% and 95%. For better readability, the figure shows 2% of total observations which are randomly selected.

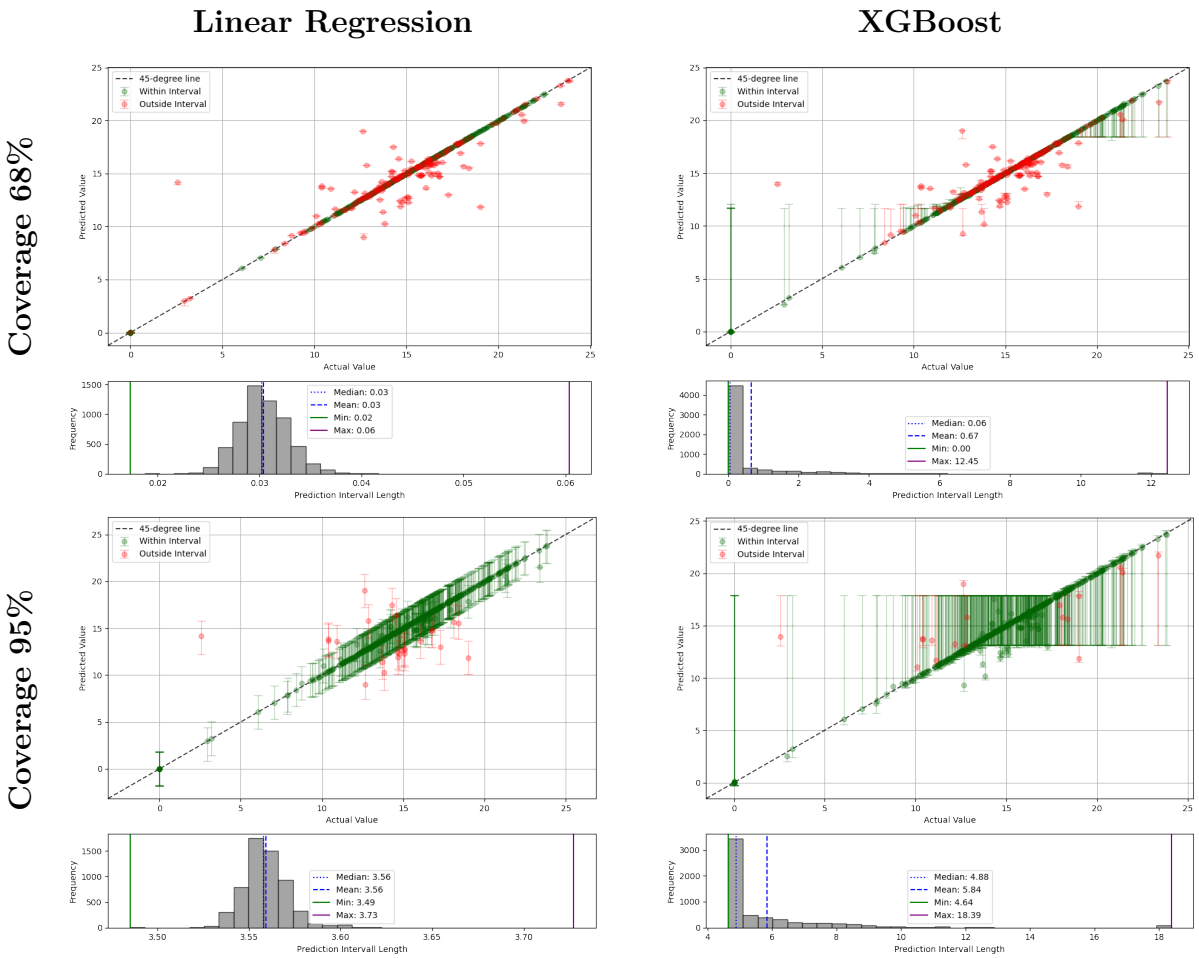
Figure 14.
Conformalized Prediction Uncertainty in Different Settings for NOx Emissions



This figure displays the deviation for the conditional mean for NOx Emissions in different settings. The settings are for uncertainty estimates based on linear regression and XGBoost and for targeted coverage rates of 68% and 95%. In all settings, the conditional mean is predicted by the best-performing XGB meta-model. The black dotted lines represent the prediction intervals using one and two standard deviations (σ), respectively. The green dashed lines represent the mean lower (upper) quantile prediction from standard quantile regression, ν_l (ν_u). Finally, the blue and orange bars show the distribution of quantile predictions for the lower and upper quantile from conformalized quantile regression. The mean of lower (upper) predictions is represented by the blue and orange dotted lines, μ_l (μ_u).

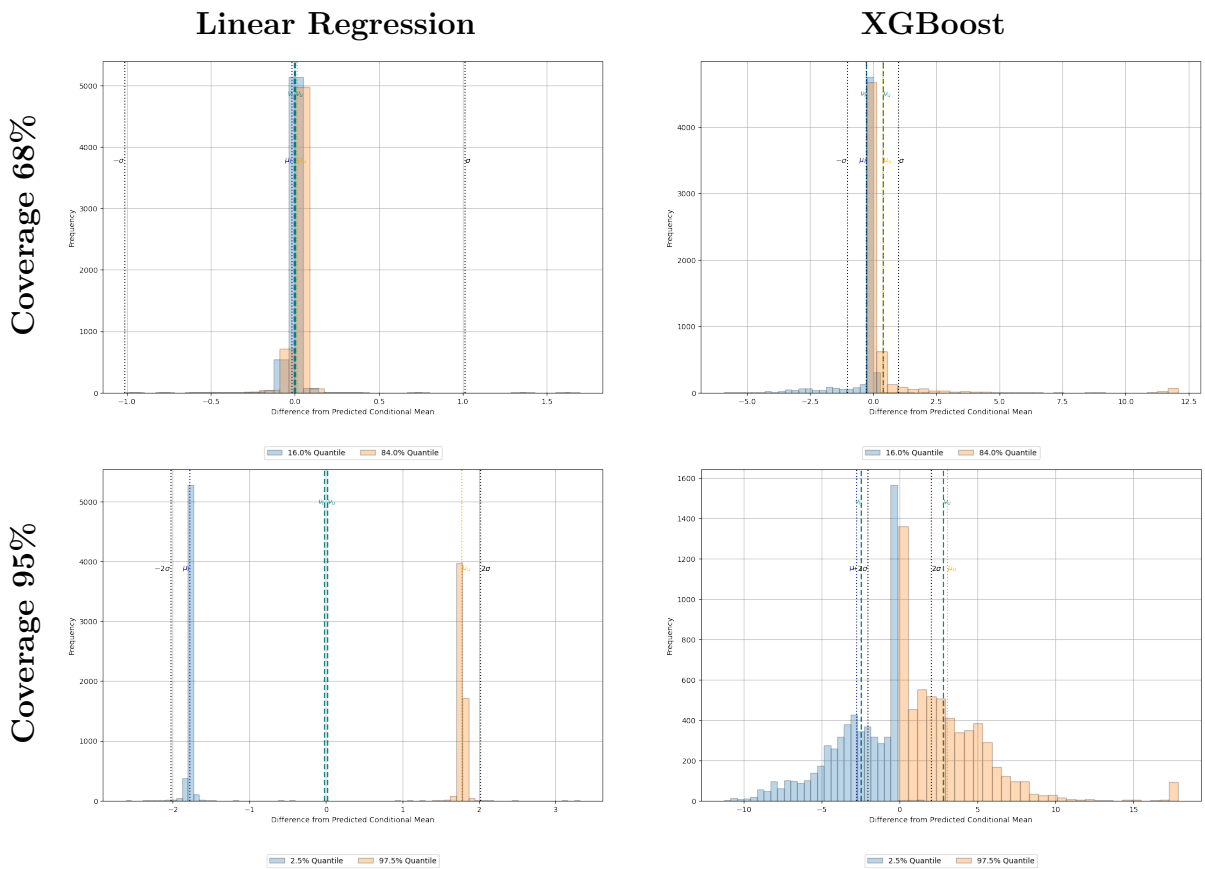
Figure 15. Deviation from Conditional Mean for NOx Emissions

Water Discharge



This figure displays conformalized prediction uncertainty for Water Discharge. The settings are for the learners linear regression and XGBoost and the coverage of 68% and 95%. For better readability, the figure shows 2% of total observations which are randomly selected.

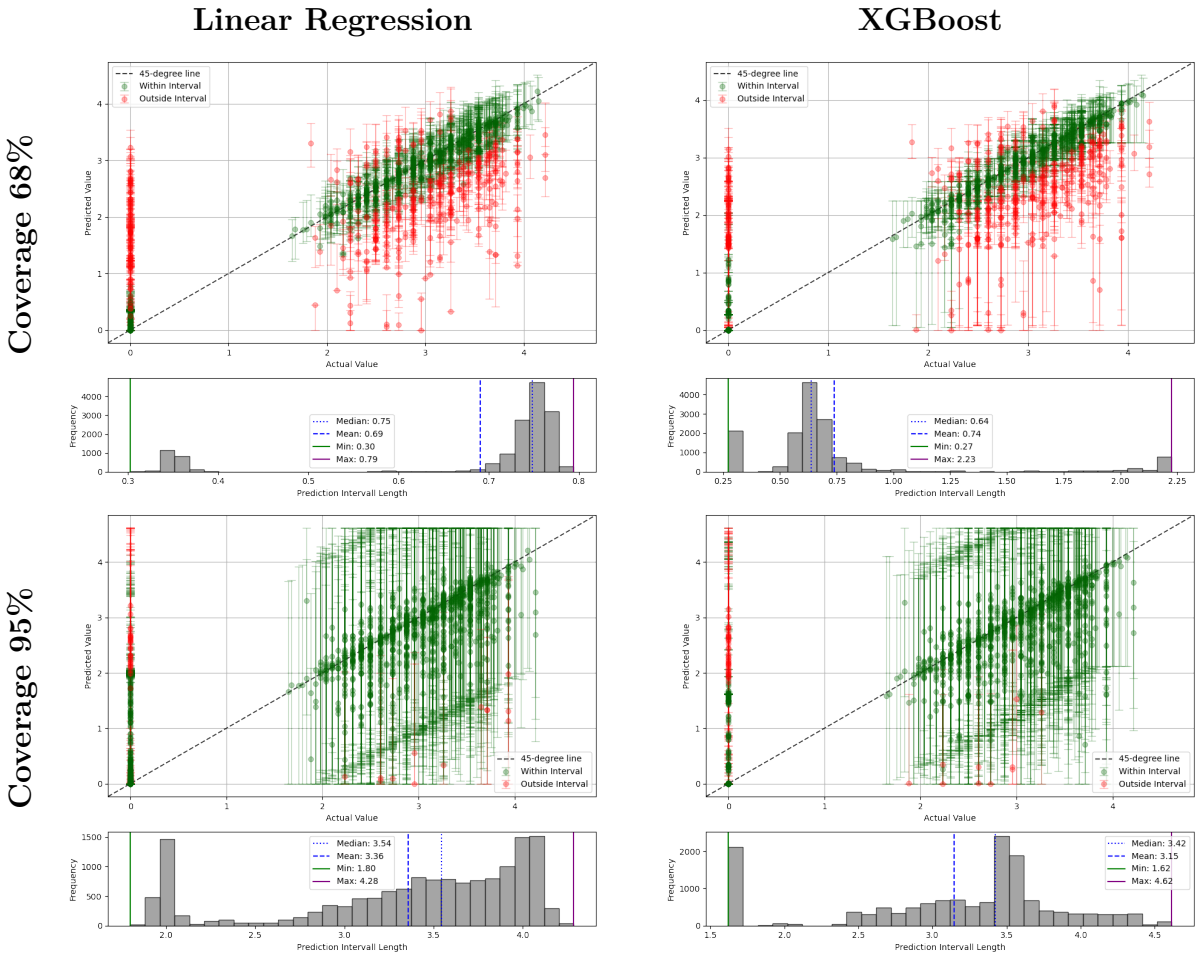
Figure 16.
Conformalized Prediction Uncertainty in Different Settings for Water Discharge



This figure displays the deviation for the conditional mean for Water Discharge in different settings. The settings are for uncertainty estimates based on linear regression and XGBoost and for targeted coverage rates of 68% and 95%. In all settings, the conditional mean is predicted by the best-performing XGB meta-model. The black dotted lines represent the prediction intervals using one and two standard deviations (σ), respectively. The green dashed lines represent the mean lower (upper) quantile prediction from standard quantile regression, ν_l (ν_u). Finally, the blue and orange bars show the distribution of quantile predictions for the lower and upper quantile from conformalized quantile regression. The mean of lower (upper) predictions is represented by the blue and orange dotted lines, μ_l (μ_u).

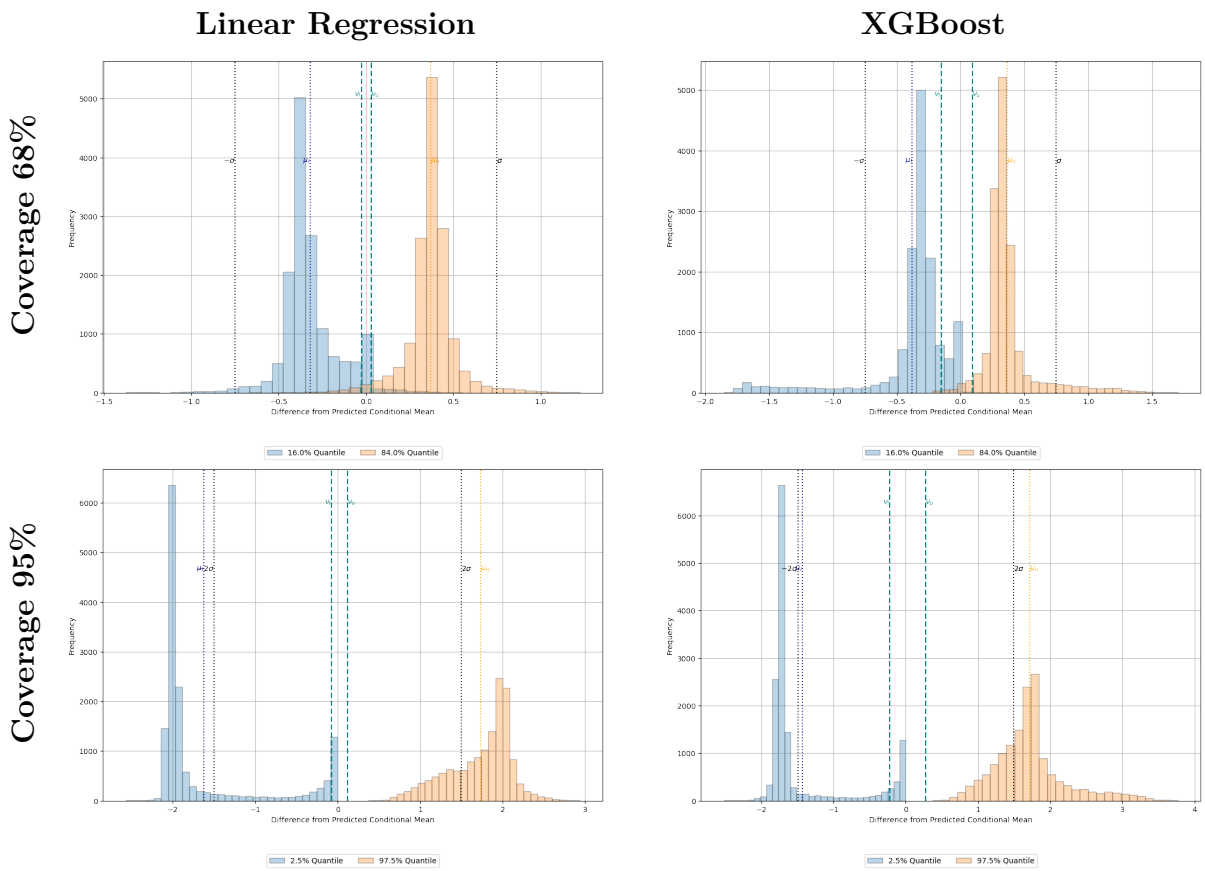
Figure 17. Deviation from Conditional Mean for Water Discharge

Female Board Share



This figure displays conformalized prediction uncertainty for Female Board Share. The settings are for the learners linear regression and XGBoost and the coverage of 68% and 95%. For better readability, the figure shows 2% of total observations which are randomly selected.

Figure 18.
Conformalized Prediction Uncertainty in Different Settings for Female Board Share



This figure displays the deviation for the conditional mean for Female Board Share in different settings. The settings are for uncertainty estimates based on linear regression and XGBoost and for targeted coverage rates of 68% and 95%. In all settings, the conditional mean is predicted by the best-performing XGB meta-model. The black dotted lines represent the prediction intervals using one and two standard deviations (σ), respectively. The green dashed lines represent the mean lower (upper) quantile prediction from standard quantile regression, ν_l (ν_u). Finally, the blue and orange bars show the distribution of quantile predictions for the lower and upper quantile from conformalized quantile regression. The mean of lower (upper) predictions is represented by the blue and orange dotted lines, μ_l (μ_u).

Figure 19. Deviation from Conditional Mean for Female Board Share

B. Early Stopping

In our efforts to improve computational efficiency during model training, we implemented an early stopping mechanism. This mechanism is designed to reduce unnecessary computations during the hyperparameter tuning process, specifically within the ten-fold cross-validation phase of each trial.Figure 20 visually represents the early stopping process.

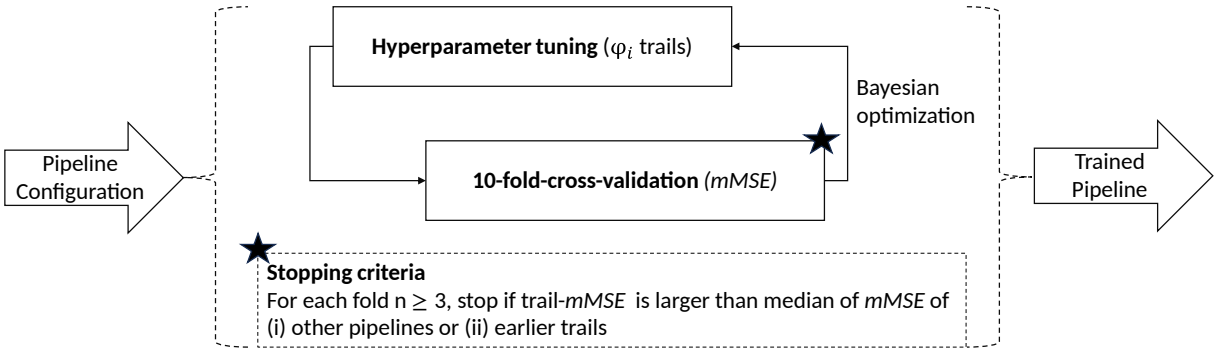
The early stopping approach is structured as follows:

1. Initialization of Trials: During the hyperparameter tuning process, each trial undergoes a tenfold cross-validation to evaluate the model's performance under different parameter settings.

2. Minimum Trial Requirement: To ensure that the model has sufficient opportunity to learn and stabilize, the early stopping mechanism is only considered after the first three folds of the cross-validation process have been completed.

3. Performance Comparison: After the third fold, the mechanism compares the mean squared error (MSE) obtained in the current fold with the median MSEs of previously evaluated models or earlier trials of the same model configuration.

4. Activation of Early Stopping: If the MSE of the current trial exceeds the average MSEs of the previous trials, the early stopping mechanism is triggered. When activated, this mechanism halts the remaining cross-validation folds for the current trial and proceeds directly to the next model configuration. This avoids further computation on a model configuration that is unlikely to outperform existing configurations.



This figure displays the early stopping mechanism that is applied throughout the code.

Figure 20. Early Stopping Approach

C. Computational Efficiency vs. Comprehensiveness Trade-off in Model Development

In this section, we elaborate on the trade-offs we encountered between computational efficiency and the comprehensiveness of our approach in model development.

Our approach to model configuration deliberately incorporates a trade-off between computational demands and the comprehensiveness of the analysis. By treating preprocessing steps as additional hyperparameters and employing a wide range of models and configurations, we place a significant burden on computational resources. This strategy, while resource-intensive, is aimed at enhancing the replicability of our results and ensuring that our methods are broadly applicable, independent of specific institutional contexts in which the data are generated.

The versatility offered by this approach can be valuable to the research community, as it allows for more generalized and adaptable modeling techniques.

However, this versatility comes at the cost of computational efficiency. Despite implementing early stopping mechanisms to reduce unnecessary calculations, the overall process remains resource-demanding. In the following, we provide an overview of the computational resources required for our approach:

- CPU: **2 x AMD EPYC Milan 7713 - 64-Core**
- GPU: **NVIDIA RTX Ada A6000**
- Time Consumption: **19.01 days** *(assuming serial execution)*
- Electricity Consumption: **87.94 kWh** *(estimation)*
- Early Stopping Effect: **active in 37.86% of 59,250 trials**

The above specifications highlight the scale of computational resources needed to execute our models effectively. The use of powerful CPUs and GPUs is essential to manage the large datasets and numerous trials involved in our comprehensive approach to discovering best performing models. Even with these resources, the time required to complete the model runs is considerable, emphasizing the trade-off between the depth of analysis and computational efficiency.

The introduction of early stopping mechanisms plays a crucial role in mitigating some of the computational burdens. As detailed in Annex B, early stopping is activated in approximately 30.02% of the 59,250 trials, significantly reducing the time and energy required to

train the models. This not only improves efficiency but also ensures that computational resources are allocated more effectively towards promising model configurations.

D. Insights and Recommendations for Future Researchers

In this annex, we provide guidance for future researchers who aim to replicate or build on our findings with reduced computational effort. By analyzing the performance of various model configurations and the factors contributing to the best results, we draw several conclusions that can inform more efficient model development in similar studies.

Table V summarizes the configurations of the best-performing models for each target variable based on the mean squared error (MSE) in the test data set. These best performing models reflect the combination of preprocessing steps and learner choices that yielded the most accurate predictions.

Target Variable	Missing Indicator		Imputation			Outlier Removal		Scaler		Transformation		Feature Selection	
	No	Yes	Mean	Median	Iterative	None	Winzorize	Standard	Robust	None	Quantile	None	Lasso
Scope 1 Emissions (tCO2e)	6	6	4	6	2	11	1	6	6	5	7	12	0
Scope 2 Emissions (tCO2e)	2	10	5	7	0	5	7	4	8	5	7	12	0
Air Pollution (tNOx)	4	7	5	6	0	8	3	1	10	8	3	11	0
Water Discharge (cbm)	3	9	8	2	2	10	2	4	8	7	5	12	0
Female Board Share (pct)	4	8	5	6	1	4	8	3	9	10	2	12	0

This table summarizes the best performing model setups along the dimensions missing indicator, imputation, outlier removal, scaler, transformation and feature selection for the five target variables: Scope 1 emissions, Scope 2 emissions, air pollution, water discharge, and female board share.

Table V. Summary of Winning Model Configurations

The analysis of the best performing models reveals several insights that can guide future research.

- Meta Learners and Complex Models: As stated in the main body of the paper, more complex models, such as meta learners, tend to outperform simpler linear models. Researchers should consider including these advanced techniques in their pipelines to achieve better performance.
- Pipeline Options: The study shows that there is no universal best configuration across all preprocessing steps. This suggests that researchers working with different data sets or under different use cases may need to experiment with the full range of pipeline options, rather than relying on a predetermined set of configurations. For example, there is no clear trend favoring a particular missing indicator or scaling method across all target variables.
- Imputation and Transformation Methods: While there are slight trends in the performance of certain imputation and transformation methods, these trends are not strong enough to recommend a specific approach universally. Researchers should evaluate the effectiveness of these methods on a case-by-case basis, as different variables may respond differently to the same preprocessing techniques.

- Feature Selection: The analysis indicates that, in our case, model configurations without feature selection generally performed better. This finding suggests that including more features, rather than reducing them, can enhance model performance. Future researchers should be cautious about overly aggressive feature selection, particularly when working with rich datasets, as it might lead to the loss of valuable information.

In addition to these findings on pipeline buliding, researchers might want to implement early stopping mechanisms, as discussed in Annex B, to reduce computational burden while ensuring that only the most promising configurations are fully explored.

E. Summary Statistics Predictors

Table VI. Summary statistics of predictor variables

The table presents summary statistics for the predictor variables. These statistics include the count of observations, the mean, standard deviation, minimum, maximum, and percentiles (25%, 50%, 75%) for each variable. The units for each variable are indicated in the "Unit" column, with common abbreviations such as B for billions, M for millions, and K for thousands.

	count	mean	std	min	25%	50%	75%	max	Unit
Accounts Payable - Long Term	18316	38.90	467.62	-7.88	0.01	0.11	0.93	18479.00	B
Avg. Payables Payment Days	80527	303.08	23949.38	-32528.25	35.47	56.43	93.25	5769517.74	Unit
Brands, Patents, Trademarks, Marketing (Gross)	27924	8.91	112.41	-0.02	0.01	0.08	0.49	5118.16	B
Capital Expenditures (Total)	101362	79.14	1009.41	-4.17	0.05	0.34	2.94	75161.53	B
Cash & Cash Equivalents	94014	155.92	1722.84	-7.43	0.10	0.66	6.18	110763.21	B
Cash & Cash Equivalents (Total)	103303	223.49	2980.01	-7.43	0.11	0.79	8.87	202104.93	B
Cash & Short-Term Investments	98451	222.83	2579.73	-0.01	0.15	0.95	9.98	124652.84	B
Computer Software (Net)	32490	11.22	113.29	-527.61	0.01	0.07	0.59	5744.00	B
Long Term Debt Issued (Cash Flow)	59479	193.77	2415.98	-111.28	0.08	0.85	8.18	145183.14	B
Long & Short Term Debt Issued (Cash Flow)	19552	1839.72	48455.29	-24.76	0.10	0.95	5.65	1787877.77	B
Short Term Debt Issued (Cash Flow)	11008	471.55	4511.91	-12.11	0.02	0.51	17.66	192778.56	B
Long Term Debt (Total)	96939	342.28	3355.23	-1302.34	0.23	1.95	15.46	183775.07	B
Total Debt	99726	556.41	4952.84	-0.12	0.35	2.78	24.86	205362.30	B
Depreciation & Amortization	13845	30.52	442.09	-6.14	0.03	0.18	1.05	14863.00	B
EBIT	104120	108.47	1348.85	-30329.63	0.08	0.66	5.65	60569.45	B
Part-Time Employees	3661	7.23	23.04	0.00	0.03	0.35	3.59	444.55	K
Equity Earnings/Loss (Pre-Tax, Nonrecurring)	41519	12.19	157.91	-2483.34	-0.00	0.01	0.24	7087.00	B
Short-Term Financial Assets	18291	203.63	2887.98	-738.25	0.01	0.17	1.93	92441.70	B
Net Financing Income/Expense	92886	-6.24	128.64	-10670.09	-0.24	-0.02	0.00	4740.00	B
Goodwill (Gross)	22899	46.16	313.86	-0.40	0.14	0.75	4.26	6958.30	B
Impairment - Financial Investments	14263	6.74	82.03	-459.28	0.00	0.01	0.15	2242.53	B
Impairment - Fixed Assets	50314	4.92	83.93	-1709.81	0.00	0.03	0.39	11072.70	B

Continued on next page

	count	mean	std	min	25%	50%	75%	max	Unit
Income Taxes	101788	28.05	340.50	-1855.99	0.01	0.12	1.23	16990.40	B
Intangible Assets (Accum. Amort. & Impair.)	34239	49.82	424.81	-52.59	0.04	0.26	1.55	18254.57	B
Intangible Assets (Gross)	34632	122.90	907.03	-60.63	0.17	1.08	6.80	37612.29	B
Intangible Assets (Net Cash Flow)	4464	18.25	193.18	-136.28	0.00	0.01	0.11	5615.31	B
Long-Term Investments	66338	105.56	1678.63	-1246.24	0.03	0.52	8.03	126930.42	B
Total Investments	85989	618.23	6893.10	-230.94	0.09	1.27	25.76	442925.68	B
Lending & Long-Term Deposits	7493	714.66	6413.50	-42.61	0.73	10.50	53.20	242431.95	B
Short-Term Loans & Receivables (Net)	40189	267.76	2113.23	-5.18	0.33	1.68	11.65	97072.45	B
Total Loans & Receivables	102232	1126.53	21694.84	-38.63	0.20	1.44	18.53	2531993.14	B
Net Cash Flow - Financing	103774	11.49	961.89	-27753.00	-0.97	-0.05	0.16	121530.63	B
Net Cash Flow - Investing	103526	-115.48	1569.11	-132477.05	-4.01	-0.40	-0.04	25880.94	B
Net Cash Flow - Operating	103874	128.39	1942.81	-80142.33	0.07	0.60	5.42	125791.99	B
Net Financial Income/Expense (Other)	17440	-1.20	29.88	-1517.46	-0.06	-0.01	-0.00	299.66	B
Net Income (After Tax)	51281	101.89	1246.50	-40408.49	0.11	0.86	6.74	44344.86	B
Nonrecurring Income/Expense	86092	-1.47	174.77	-10939.90	-0.24	-0.01	0.00	29391.89	B
Total Operating Expenses	99953	760.55	6541.08	-1143.05	0.68	4.23	31.37	227970.94	B
Total Other Assets	100956	122.21	1323.71	-41897.01	0.04	0.42	5.69	69247.18	B
Total Other Current Assets	30798	29.04	260.52	-14527.45	0.03	0.30	4.92	13300.06	B
Total Other Current Liabilities	89942	102.06	1122.25	-12625.25	0.04	0.40	4.97	150865.89	B
Total Other Noncurrent Liabilities	42108	115.75	2088.40	-1290.92	0.09	0.68	7.05	145414.99	B
Payables & Accrued Expenses	100432	131.24	1145.48	-0.69	0.14	0.94	8.45	77457.92	B
Plant, Machinery & Equipment (Gross)	73124	525.67	6823.00	-105.61	0.15	1.20	11.25	305445.00	B
PPE (Accum. Depreciation & Impairment)	88596	399.03	4681.38	-10168.69	0.16	1.27	11.25	227543.45	B
PPE (Gross)	92792	855.02	8895.91	-4920.35	0.46	3.59	33.30	377471.99	B
PPE (Net)	100907	437.36	4442.67	-991.57	0.20	1.84	17.45	207052.67	B
Other PPE (Gross)	65438	120.92	1761.60	-3128.77	0.03	0.31	3.70	169436.18	B
Total Provisions	89472	79.50	778.60	-166.19	0.06	0.49	4.88	44491.12	B
Long-Term Receivables & Loans	40536	100.80	1385.81	-31.92	0.01	0.14	1.48	46311.56	B
R&D Expense	32492	40.82	630.29	-136.86	0.04	0.20	1.99	22401.73	B
R&D Costs (Gross)	10642	75.39	592.23	-0.00	0.02	0.09	0.73	10374.45	B

Continued on next page

	count	mean	std	min	25%	50%	75%	max	Unit
Short-Term Debt & Notes Payable	54238	267.85	2309.18	-168.57	0.06	1.15	20.76	104779.30	B
Total Short-Term Investments	52532	131.21	1936.84	-3.40	0.03	0.33	3.48	95270.26	B
Total Assets	103976	2732.62	30953.65	0.00	2.05	12.61	137.79	1808429.76	B
Total Book Capital	103954	1154.24	10028.24	-66.81	1.29	7.70	73.41	442390.86	B
Total Current Assets	89457	505.03	4941.26	-57.13	0.50	3.03	26.51	283116.65	B
Total Current Liabilities	44789	516.29	3918.81	0.00	0.80	4.49	40.21	298411.51	B
Total Fixed Assets (Net)	82720	737.86	6722.46	-17420.67	0.59	3.86	31.22	283593.15	B
Total Liabilities	103952	2045.66	26076.03	-20.44	0.99	6.76	75.56	1700262.11	B
Total Shareholders' Equity	99822	689.10	7064.58	-11055.62	0.75	4.17	34.21	304899.93	B
Working Capital Change (Cash Flow)	98107	-26.13	623.51	-31484.37	-0.38	-0.01	0.08	35264.41	B
Avg. Receivables Collection Period (Days)	43635	95.47	919.42	-17463.43	41.87	63.48	90.74	152776.77	Unit
Full-Time Employees	79275	323.43	52542.63	0.00	1.68	6.93	23.05	10785002.89	K
R&D as % of Revenues	43026	287.33	8230.36	-34.17	0.68	2.58	7.06	598611.42	Unit
Turnover	62282	2.12	108.80	-17573.65	0.00	0.02	0.23	19941.57	B
Date of Incorporation	105929	1.98	0.03	1.81	1.97	1.99	2.00	2.02	K
Year	109085	2.01	0.01	2.00	2.01	2.02	2.02	2.02	K
Parent Shareholders' Equity (Total)	4123	482.85	1871.17	-63.52	6.26	55.35	326.36	37441.42	B
Business Financing Revenue (Other)	4278	45.33	542.50	-8261.00	0.00	0.04	1.33	11075.05	B
Inventories (Finished Goods)	22867	48.84	497.25	-4.83	0.03	0.22	1.57	17537.84	B
Long-Term Loans	9204	62.72	938.79	-0.00	0.01	0.19	1.91	28704.25	B
Short-Term Loans	10735	33.42	370.27	-0.00	0.01	0.22	1.96	21022.04	B
Total Noncurrent Liabilities	41426	202.81	2583.53	-99.82	0.10	0.85	6.30	131277.39	B

It’s in the Financials, Stupid! But is it certain?

New Insights from Financial Statements

Christian Haas, Ulf Moslener & Sebastian Rink

October 2024

Frankfurt School of Finance & Management

Motivation

- Company-level sustainability data are **complex and non-linear**
- Machine Learning is only applied to GHG emissions so far
- **Increasing amounts of sustainability data** reported by companies

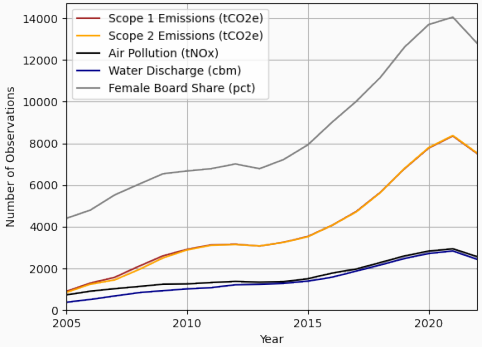


Figure 1: Reported Sustainability Data by Companies

Research Questions

1. To what extent can we **derive corporate sustainability data from corporate financial data only** using ML?
2. How does the prediction performance **change for different dimensions**?
3. **How certain are the point estimates** of the prediction models?

Methodology & Data

Target Variables

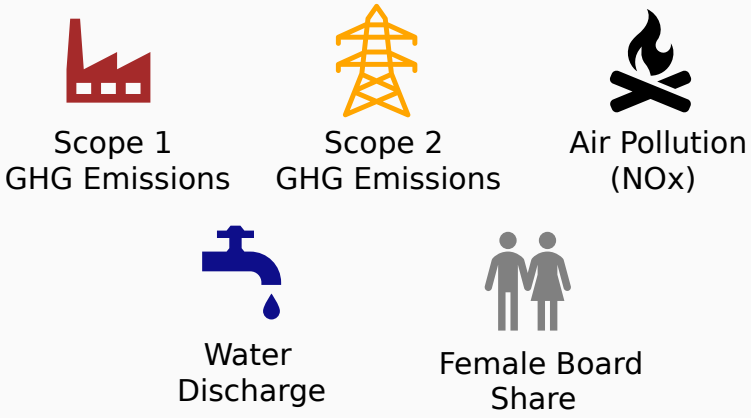


Figure 2: Target Variables

Training Approach for Point Estimates

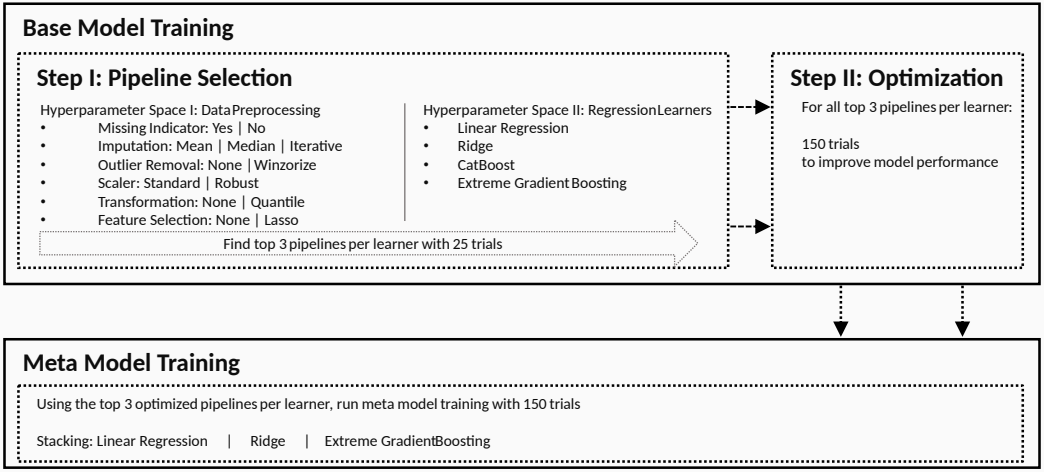


Figure 3: Point Estimate Training Approach using MSE

Stacking Stopping

Uncertainty Quantification

1. Quantile Regression:
   $L_\alpha(\hat{y}_\alpha, y) = (y - \hat{y}_\alpha)\,\alpha \mathbb{1}\{y > \hat{y}_\alpha\} + (\hat{y}_\alpha - y)\,(1-\alpha)\,\mathbb{1}\{y \leq \hat{y}_\alpha\}$
2. Quantile Prediction:
   $\hat{y}_\alpha(x)$
3. Conformal Scores:
   $c(x, y) = \max\{\hat{y}_{\tau/2}(x) - y, y - \hat{y}_{1-\tau/2}(x)\}$
4. Rectifying Quantiles for Intervals:
   $I(x) = \left[\hat{y}_{\tau/2}(x) - \hat{r}, \hat{y}_{1-\tau/2}(x) + \hat{r}\right]$, where
   $\hat{r} = \text{Quantile}\left(\frac{\lceil (n_{cal}+1)(1-\alpha) \rceil}{n_{cal}}, \{c_1, ..., c_{n_{cal}}\}\right)$

Dataset

Dataset	Scope 1 Emissions	Scope 2 Emissions	Air Pollution	Water Discharge	Female Board Share
General Information					
Number of observations	47685	47320	20980	18426	108834
Number of sectors	83	83	76	74	86
Number of countries	83	83	63	63	95
Number of companies	8391	8335	3389	3098	14406
Start year	2005	2005	2005	2005	2005
End year	2022	2022	2022	2022	2022
Number of predictor variables	240	240	213	211	255
Data completeness (in %)	63.86	63.77	65.72	65.62	57.92
Target Variable Information					
Log (1+value) Mean	10.75	11.07	6.31	14.98	2.17
Log (1+value) Std	3.55	2.72	3.28	3.60	1.40
Log (1+value) Min	0.00	0.00	0.00	0.00	0.00
Log (1+value) Max	22.21	22.72	16.46	24.01	4.62

Table 1: Summary Statistics

Results & Discussion

It's in the financials, stupid!

Target Variable	Metric	Base Learner				Meta Learner		
		Linear Regression	Ridge	CatBoost	XGBoost	Linear Regression	Ridge	XGBoost
Scope 1 Emissions (tCO2e)	MAE	1.261	1.240	0.648	0.662	0.610	0.609	0.589
	MSE	3.401	3.252	1.579	1.617	1.622	1.622	1.589
	R2	0.713	0.726	0.867	0.864	0.863	0.863	0.866
Scope 2 Emissions (tCO2e)	MAE	1.252	1.118	0.513	0.627	0.556	0.556	0.527
	MSE	3.400	2.959	1.163	1.269	1.275	1.275	1.204
	R2	0.540	0.599	0.842	0.828	0.827	0.827	0.837
Air Pollution (tNOx)	MAE	1.501	1.444	0.695	0.905	0.672	0.672	0.650
	MSE	4.598	4.320	1.923	2.262	1.907	1.907	1.909
	R2	0.564	0.590	0.818	0.786	0.819	0.819	0.819
Water Discharge (cbm)	MAE	1.782	1.782	0.278	0.623	0.275	0.275	0.273
	MSE	7.528	7.522	1.055	1.526	1.063	1.063	1.053
	R2	0.443	0.443	0.922	0.887	0.921	0.921	0.922
Female Board Share (pct)	MAE	0.876	0.876	0.473	0.501	0.448	0.448	0.428
	MSE	1.253	1.249	0.598	0.638	0.595	0.595	0.592
	R2	0.347	0.349	0.688	0.667	0.690	0.690	0.691

Table 2: Global Model Performance

Variation by Quintile

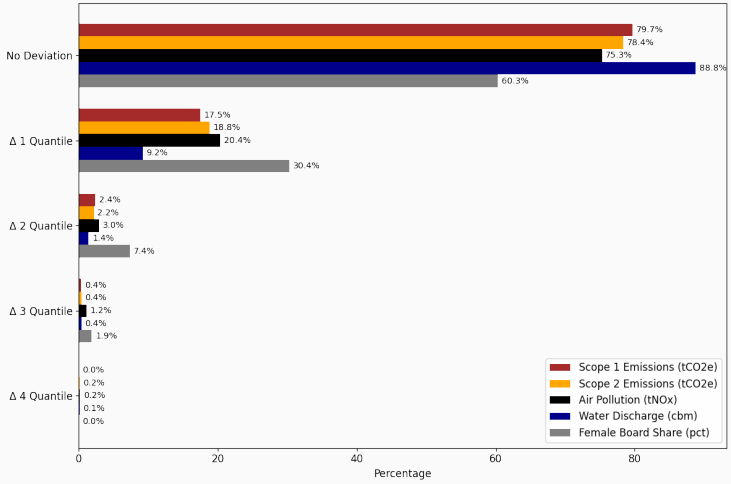


Figure 4: Number of Deviating Quintiles

Temporal Variation

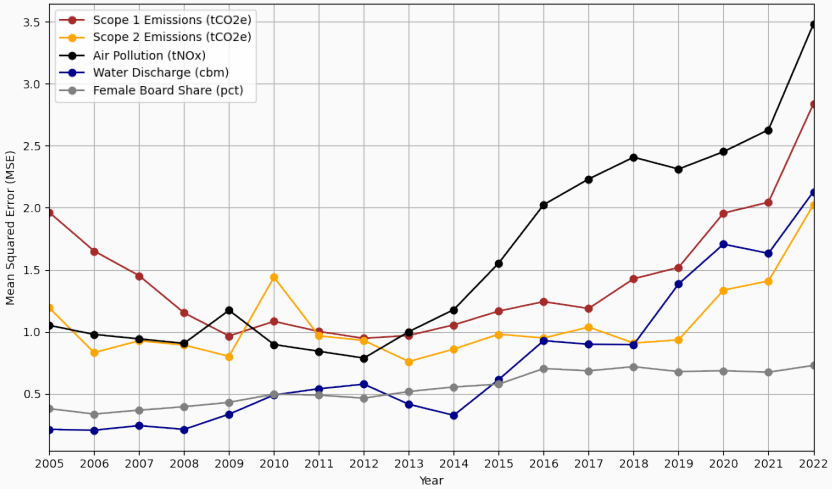


Figure 5: Temporal Model Performance

Spatial Variation

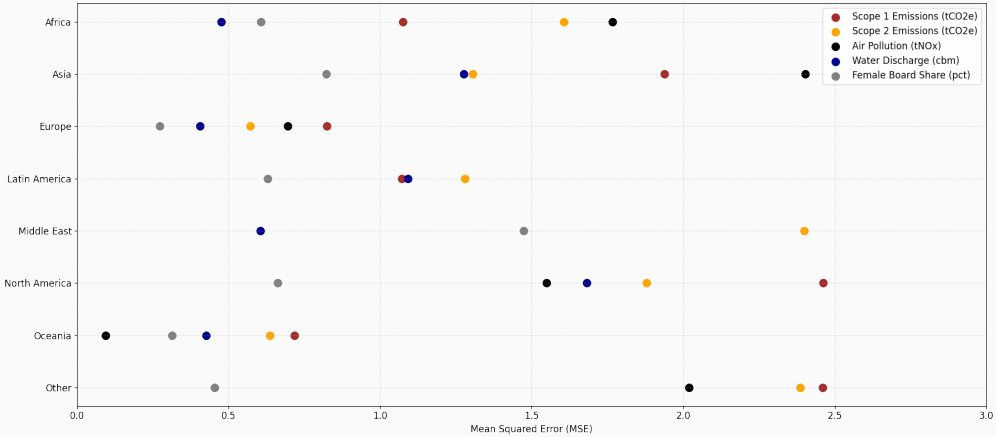


Figure 6: Spatial Model Performance

Sectoral Variation

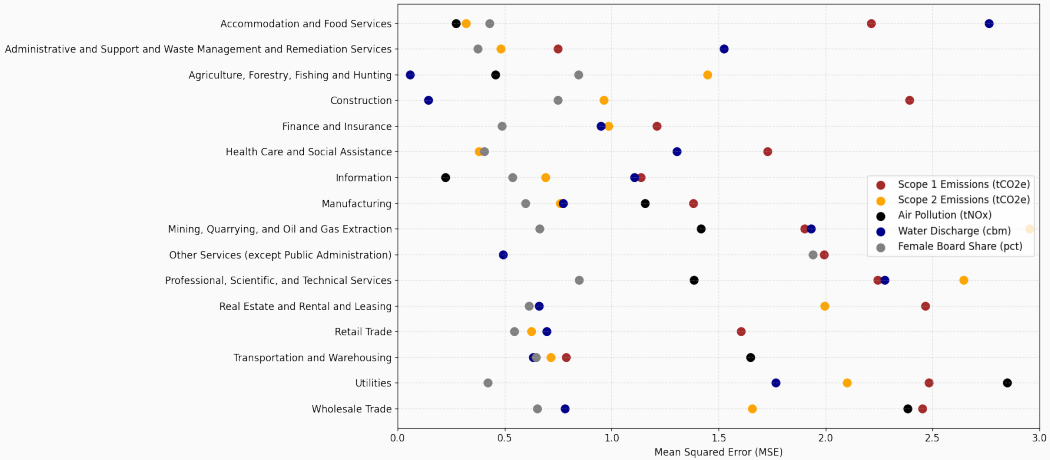


Figure 7: Sectoral Model Performance

Beware of Prediction Uncertainty!

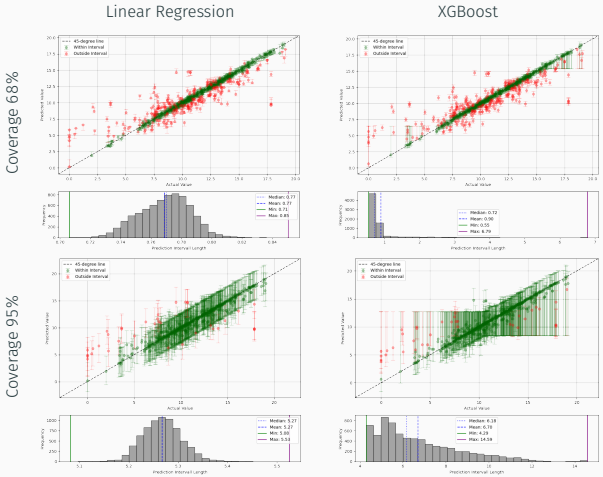


Figure 8: Prediction Uncertainty in Different Settings for Scope 1 Emissions

Reflect On Uncertainty Measure

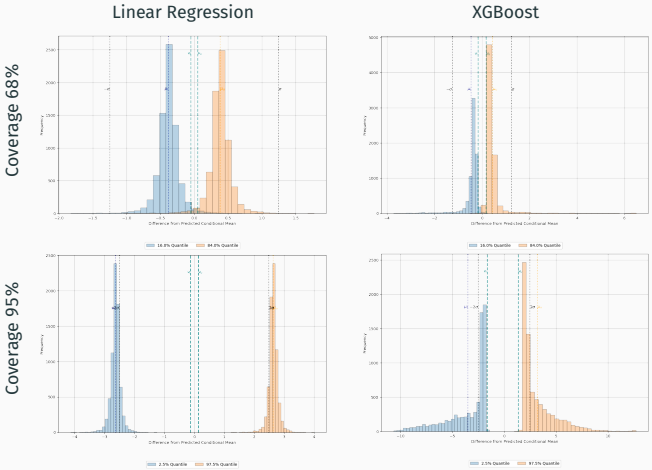


Figure 9: Deviation from Conditional Mean for Scope 1 Emissions

Conclusion

- **It's actually in the financials**, stupid!
- **Beware of prediciton variation and uncertainty!**
- Policy makers may take a more open stance in ML in sustainable finance, but the exact area of application matters
- Future research: local models & little reported sustainability metrics

Takeaways For Policy Makers

1. Machine learning can support financial institutions and companies in assessing sustainability risks and impacts with a high degree of predictive performance. As such, these **institutions should be allowed to use ML to predict sustainability data** as a supplement to the available reported data, especially if the costs of accessing the raw data are high.
2. Users of ML-predicted sustainability data should be required to **increase transparency on the quality of the predictions**, not only at the global level, but also in the dimensions time, space, and sector.
3. **Prediction uncertainty should be considered** by users of ML-predicted sustainability data and the respective assumptions / risk appetite should be made transparent.

References

Ismail, Shereen, Zakaria El Mrabet, and Hassan Reza (Dec. 2022). **“An Ensemble-Based Machine Learning Approach for Cyber-Attacks Detection in Wireless Sensor Networks”.** In: *Applied Sciences* 13.1, p. 30. ISSN: 2076-3417. DOI: *10.3390/app13010030*.

Appendix

Early Stopping

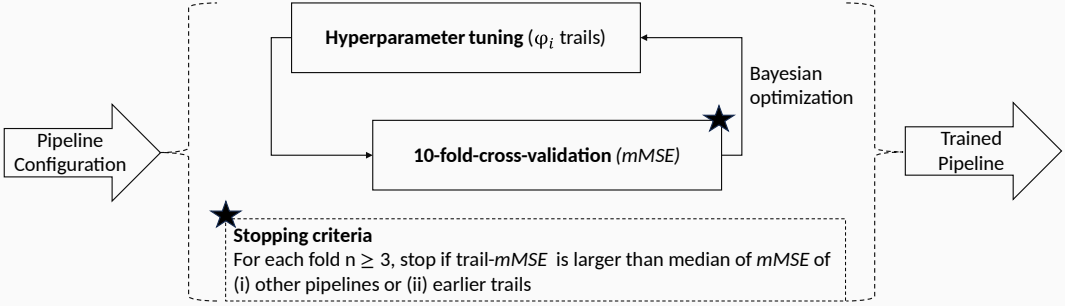


Figure 10: Early Stopping Approach

Efficiency

- CPU: **2 x AMD EPYC Milan 7713 - 64-Core**
- GPU: **NVIDIA RTX Ada A6000**
- Time Consumption: **19.01 days** *(assuming serial execution)*
- Electricity Consumption: **87.94 kWh** *(estimation)*
- Early Stopping Effect: **active in 37.86% of 59,250 trials**

Boosting, Bagging, Stacking

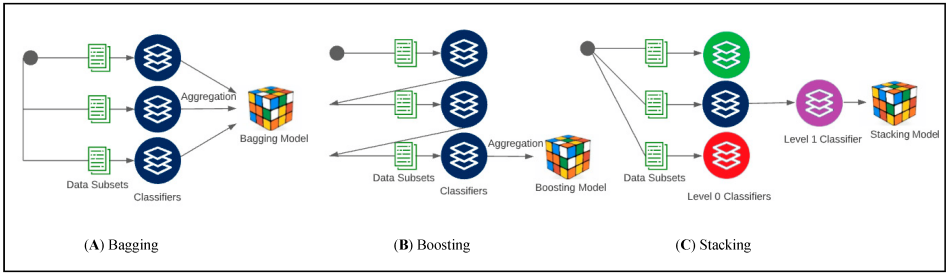


Figure 11: Approach to Boosting, Bagging, and Stacking by Ismail, El Mrabet, and Reza 2022

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Quantifying material risks from textual disclosures in financial statements using LLMs and LLM agents – A case study on physical risks[1]

Taejin Park,
Bank for International Settlements

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

Quantifying Material Risks from Textual Disclosures in Financial Statements using LLM Agents – A Case Study on Physical Risks

Taejin Park[1]

Abstract

This paper highlights the novel application of large language models (LLMs) in context-aware materiality assessment of risk disclosures, using physical risks as an example. As physical risks like natural disasters increasingly affect financial stability, accurately capturing and evaluating these risks in corporate disclosures is critical. Traditional NLP methods based on keyword analysis struggle to detect risks embedded in complex or indirect language. In contrast, LLMs, through their ability to understand context and semantics, offer significant improvements in extracting and evaluating disclosures. The key innovation explored in this study is the LLM agent's capacity to assess materiality by integrating company-specific financial, operational, and geographic data, providing a more accurate and tailored evaluation of risks. This approach enhances the ability to differentiate between material and non-material disclosures, aiding financial supervisors and investors in more effective decision-making.

Keywords: Physical risks, risk assessment, large language models (LLM), Natural language processing (NLP), language agent, corporate disclosures

JEL classification: G32, C45, Q54

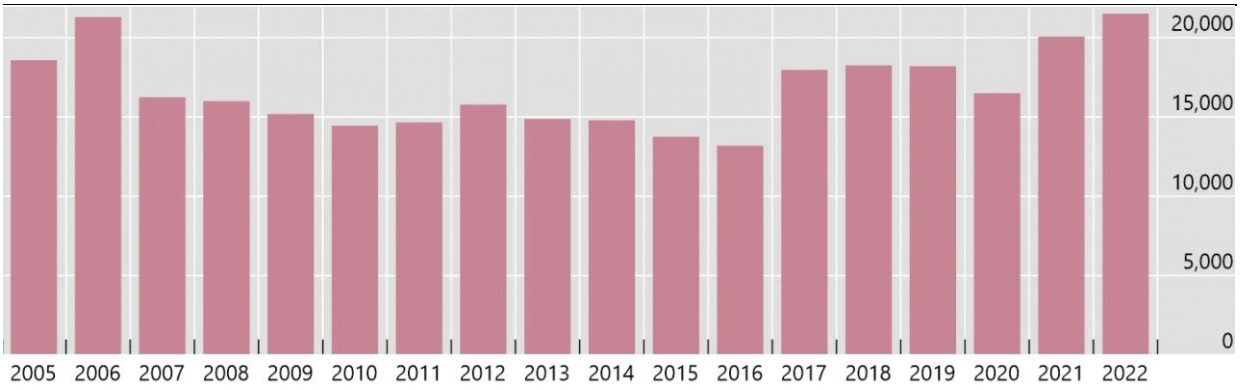
[1] Bank for International Settlements (taejin.park@bis.org). The views expressed here are those of the author only and not necessarily those of the Bank for International Settlements.

1. Introduction

The timely and reliable capture of physical risks is crucial for both investors and financial supervisors. Physical risks, such as natural disasters, pose significant financial stability risks due to their sudden and potentially severe impacts on companies. Unlike transition risks, which develop gradually—such as policy changes or shifts in consumer preferences—physical risks can occur unexpectedly and cause substantial damage.

An increasing number of companies are disclosing their exposure to physical risks (Graph 1). However, extracting relevant information from these disclosures presents challenges. Unlike transition risks, which can often be quantified through metrics such as carbon emissions or energy mix, physical risk disclosures are typically qualitative. As a result, if these disclosures are not thoroughly analyzed, important risks may be overlooked. Furthermore, assessing materiality requires subjective judgment, as readers must weigh short-term versus long-term impacts and relate the risks with various aspects of a company's operation. Given the inherent uncertainty of natural disasters, even well-disclosed risks may be underestimated depending on the risk tolerance of the reader.

Frequency of physical risk-related keywords in SEC 10-K filings — Graph 1



Included keywords: flood, wildfire, hurricane, Flood, typhoon, Hurricane, drought, Wildfire, heat wave, extreme heat, Typhoon, extreme cold, Drought, Extreme heat, heatwave, FLOOD, WildFire, WILDFIRE, HeatWave, HURRICANE, cold wave, Extreme cold, coldwave, Heat Wave, TYPHOON, Heat wave, Extreme Heat, Heatwave, Extreme Cold, DROUGHT and HEAT WAVE.

Sources: S&P; author's calculations.

Recent developments in natural language processing (NLP) and machine learning, especially large language models (LLMs), offer significant potential for efficiently scaling up the capture of qualitative risk exposures from disclosures (BIS, 2024). Despite this progress, challenges remain. Accurately capturing information relevant to a specific user's needs is particularly difficult. The process involves multiple steps: first, capturing the risk information using NLP; second, analyzing the sentiment of that risk; and finally, contextualizing the risk sentiment to ensure it is relevant to the reader's specific concerns. Proper contextualization, like human expert's judgment, is critical for effective risk assessment (Graph 2).

This paper explores how various NLP techniques can be applied to analyze the disclosure of physical risks and presents real-world examples to illustrate their practical applications.

LLM-based risk analysis process from company disclosures Graph 2



Source: Author's elaboration.

2. Extracting Relevant Information

1. Traditional keyword-based approaches

Traditional techniques for identifying keywords or topics from large volumes of text primarily rely on word frequency analysis. Methods like Term Frequency-Inverse Document Frequency (TF-IDF; Jones, 1972) and Bag of Words (Harris, 1954) rank words based on how often they appear within a document or across a collection of texts. More advanced techniques, such as Latent Dirichlet Allocation (LDA; Blei et al, 2003) and Latent Semantic Analysis (LSA; Deerwester et al, 1990), group words into topics by identifying patterns in their co-occurrence. Approaches like TextRank (Mihalcea and Tarau, 2004) and basic word frequency counting also emphasize the repetition of words or phrases. Overall, these methods focus on word frequency to extract key terms or themes from text.

Such traditional keyword extraction techniques offer the advantage of simplicity, transparency and computational efficiency, making them accessible for basic text analysis tasks. They are effective for identifying frequently occurring terms and clustering related words into topics, which is useful for quick insights or when processing large text corpora.

However, these methods have notable limitations. They often fail to capture the semantic meaning of words, as they treat words in isolation or based solely on their co-occurrence patterns. Additionally, they struggle with handling synonyms, polysemy (words with multiple meanings), and contextual nuances. Another major limitation is the difficulty in pre-defining an exhaustive list of keywords related to complex topics where descriptions can vary widely. Moreover, risks may be described indirectly, without mentioning any of the expected keywords, making it easy for these methods to overlook critical information (Box 1). As a result, they may miss deeper relationships within the text, especially in more complex language structures or when a deeper understanding of context is required. This restricts their accuracy and limits their ability to provide comprehensive insights in advanced text analysis tasks.

Box 1

Examples of disclosures that do not include typical keywords related physical risks

- "We anticipate a modest decrease in harvest volumes versus the prior year as logging conditions normalize following a period of relatively **dry weather**." (Rayonier Inc., 8-K, 31-01-2024)
- "Because substantially all of our products are manufactured and assembled in and distributed from a few locations, our operations could be interrupted by events beyond our control, including: … **water damage** or **water shortage** …" (Acushnet Holdings Corp., 10-K, 01-03-2023)
- "Both parks experienced unusually **inclement weather** during that quarter." (Cedar Fair, L.P., 10-K, 16-02-2024)
- "FY23 Saleable production levels are lower than prior year due to the impacts from these **poor geotechnical conditions** and the impacts from the rock intrusion in the September quarter, that slowed production rates and impacted yield." (Coronado Global Resources Inc., 8-K, 22-01-2024)
- "…**warmer weather in the heating season** and **cooler weather in the cooling season** will reduce our customers' energy demand and our retail operating revenues." (Avista Corporation,10-K, 2024-02-21)

2. Recent context-based approaches

Recent advances in NLP, particularly in semantic search using embeddings, have significantly improved the ability to overcome the limitations of traditional keyword-based methods. Embeddings, such as those generated by models like Word2Vec (Mikolov et al, 2013), GloVe (Pennington et al, 2014), and more recently, transformer-based models like BERT (Devlin et al, 2019) and GPT (Brown et al, 2020), represent words and phrases as continuous vectors in a high-dimensional space (Box 2). This allows the models to capture the semantic relationships between words, even if they do not explicitly share the same surface form. As a result, semantic search can identify relevant information based on meaning, rather than relying solely on exact word matches. For example, in the context of physical risks, embeddings can recognize descriptions of risks even when they are phrased differently or do not include predefined keywords. This shift from word frequency to meaning-based search allows NLP systems to better handle synonyms, polysemy, and context, leading to more accurate and comprehensive extraction of relevant information from large and complex texts.

Box 2

Example of OpenAI's 'text-embedding-3-small' model: translating text into a 1,536-dimension vector

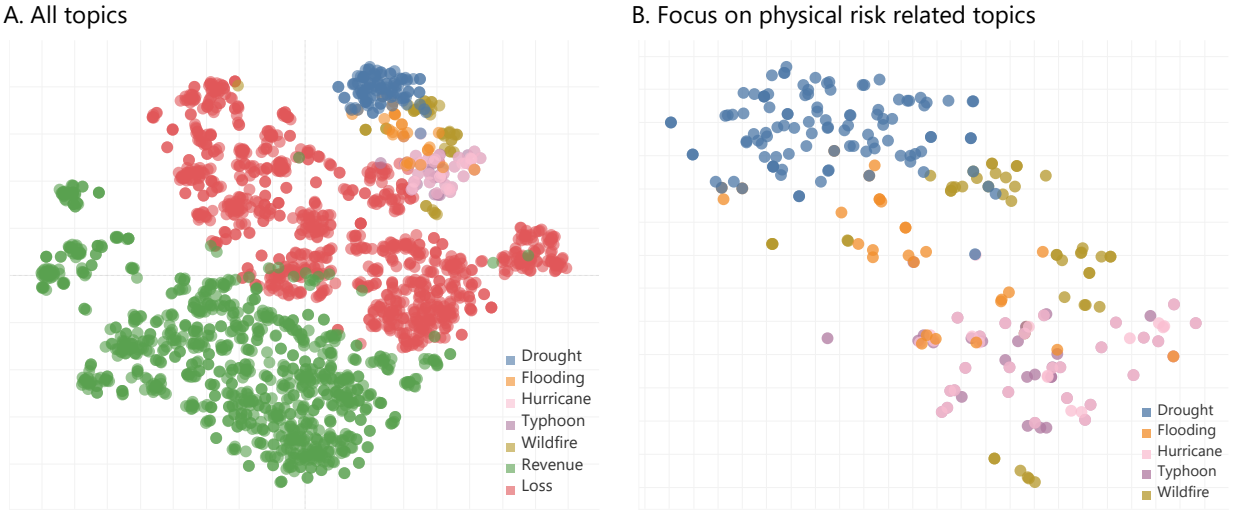
"Prolonged drought conditions increase the intensity of competition for fresh water." (Energy Transfer LP, 10-K, 2023)

[0.03909442201256752, 0.025339428335428238, …, 0.03553372994065285, 0.012559967115521431]

To illustrate the use of embeddings, I extracted 2,272 sentences from SEC 10-K and 8-K filings over the past year that are semantically close to physical risk-related keywords such as "drought," "flooding," "hurricane," "typhoon," and "wildfire."

Additionally, I included sentences related to general financial topics, such as "revenue" and "loss," to compare them with the physical risk-related topics. These sentences were then translated into 1,536-dimensional vectors, representing their semantic meaning based on OpenAI's 'text-embedding-3-small' model. However, because this high-dimensional space is difficult to interpret visually, I reduced the dimensionality to two using the t-SNE (t-Distributed Stochastic Neighbor Embedding; van der Maaten and Hinton, 2008)[2] method for visualization purposes. After applying t-SNE, the embeddings were plotted in 2D space (Graph 2), enabling a clearer visual comparison between physical risk-related topics and general financial topics, as well as distinctions among the physical risk-related topics.

2D visualization of embeddings for physical risks and financial topics using t-SNE Graph 2



The embeddings are generated using OpenAI's 'text-embedding-3-small' model.

Sources: OpenAI; ProntoNLP; S&P; author's calculations.

Graph 2.A reveals several interesting patterns in the sentence embeddings. Clear clusters emerge, with distinct separation by topic, highlighting the effectiveness of the embeddings in capturing semantic relationships. Notably, the sentences related to physical risks form a well-defined cluster, distinct from the general financial topics like "revenue" and "loss." This suggests that the embeddings effectively distinguish physical risk-related content from broader financial discussions, even after significant dimensionality reduction. Additionally, the cluster of physical risk-related sentences is positioned close to the cluster of "loss"-related sentences, which aligns with expectations, as both topics likely convey negative impacts on a company's performance. This proximity suggests a semantic connection, where physical risks and losses are understood to be closely related in their potential consequences for businesses.

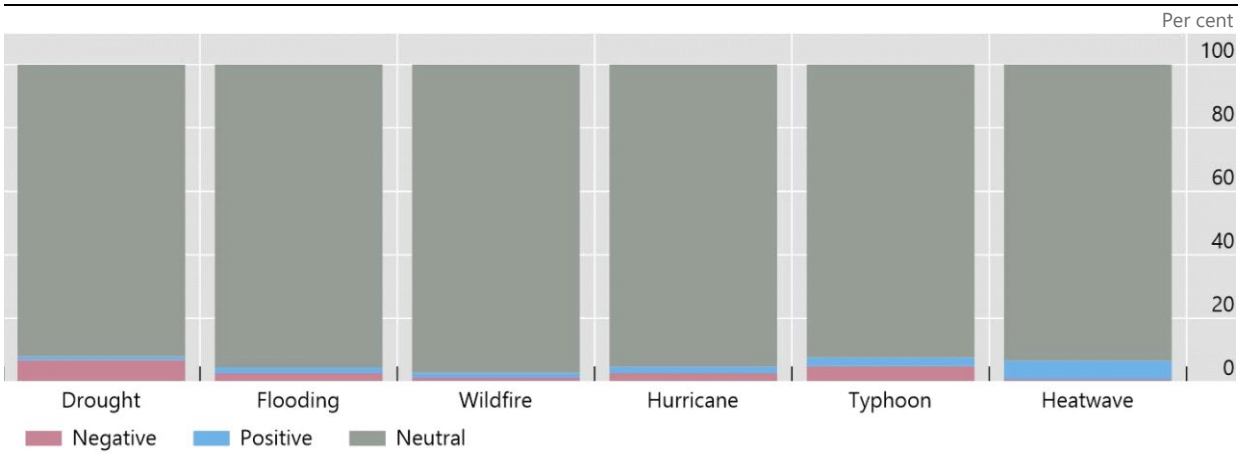
[2] t-SNE is a non-linear dimensionality reduction technique commonly used to visualize high-dimensional data. It works by preserving the local structure of the data, meaning that points close to each other in the high-dimensional space remain close in the lower-dimensional space, making it well-suited for visualizing clusters or relationships in embeddings.

When zooming in on only the physical risk-related sentences, distinct clusters emerge corresponding to different types of events (Graph 2.B). Notably, the embeddings for "hurricane" and "typhoon" show a significant overlap, despite the differences in their surface forms. This highlights the strength of the embeddings in capturing semantic similarity, as these two terms, though different in wording, represent nearly identical meteorological phenomena. Furthermore, their cluster is positioned opposite the "drought" cluster, which also makes sense. While hurricanes and typhoons involve excessive water, droughts represent water scarcity, placing them at opposite extremes. This clear separation demonstrates the model's ability to distinguish between different types of physical risks while effectively grouping semantically similar events, recognizing and organizing concepts based on meaning rather than just word forms.

3. Assessing Risk Sentiment

Once information related to physical risks is captured, it is important to assess whether the disclosure is positive, negative, or neutral for the company, as not all disclosures have a negative impact. Some may be neutral descriptions, or even positive if accompanied by negations. In fact, my analysis of 6,000 randomly selected physical risk-related sentences across six different event types from SEC 8-K and 10-K filings over the past year revealed that the vast majority of the captured information was classified as "neutral" according to sentiment analysis by ProntoNLP, an NLP service provider (Graph 3). This underscores the need for sentiment analysis following the capture of physical risk-related text disclosures.

Sentiment analysis of physical risk disclosures — Graph 3



The analysis is based on a random selection of 1,000 sentences per category related to physical risk disclosures.

Sources: ProntoNLP; S&P; author's calculations.

Traditional NLP approaches for sentiment classification often rely on counting positive or negative words in a text. While simple, this method tends to miss a lot of nuance, as it fails to account for context, negations, or the overall structure of the sentence. More advanced approaches have shifted toward supervised machine learning, where sentiment classification occurs in the embedding space, allowing

models to capture contextual sentiments. These methods, however, typically require large amounts of labelled data for training, making human annotation important for tasks like fine-tuning. Recent generative LLMs, such as those equipped with zero-shot or few-shot learning capacity, have reduced the need for extensive fine-tuning. These models can classify sentiment with little to no task-specific training, making sentiment analysis more efficient and adaptable, particularly in cases where labelled data is limited.

4. Assessing Materiality

Even if risk sentiment is accurately captured, not all negative sentiments are necessarily material to users. The materiality of a given risk is highly subjective and depends on the specific priorities and needs of the information users. Box 3 illustrates examples of disclosures where understanding the broader context is essential. These disclosures cannot be fully understood in isolation, as their significance often hinges on the circumstances surrounding the company, industry, or location. Therefore, assessing the materiality of information relevant to a particular user is critical. However, determining materiality is inherently subjective, varying based on the reader's perspective, objectives, and risk tolerance.

Box 3

Examples of disclosures that require understanding of the context

- Approximately $0.01 per diluted share of net expense due to fire related costs.
- As more fully discussed in Note 3, in December 2021, the Company wrote off approximately $8.7 million of inventory which is held at the Solar Park contract manufacturer's site in Korea and is not expected to be recoverable.
- Current accident year losses in 2022 included $69.2 million of net losses and loss adjustment expenses attributed to Hurricane Ian and the Russia-Ukraine conflict.
- In recent years, California and Utah have experienced persistent and severe drought conditions.
- We have experienced short-term impacts on our sales due to closures from weather events in recent years, including Hurricane Ian in Florida in 2022.
- The long-term trend is that wildfires are increasing resulting from drought conditions.

Understanding materiality requires a high level of language proficiency and the ability to grasp the overall context, much like a human expert. Recent technologies in LLMs have made this more feasible. Developing LLM agents, especially those specialized in identifying the materiality of risks, could offer a promising solution. An LLM agent is defined by its ability to access and integrate external knowledge provided by users, allowing it to broadly capture the context and interpret information in a human-like manner.

However, just as with human experts, fully understanding the materiality of risks requires extensive knowledge about the specific company. Factors like industry,

location, financials, and business description all play a crucial role in determining materiality. A comprehensive assessment that integrates this information is essential to making informed judgments about the significance of disclosed risks.

To test the capabilities of LLM agents in such tasks, I conducted an analysis using the sample sentences from 8-K and 10-K filings listed in Box 3. In addition to the disclosed text, I provided financial and contextual information to support the LLM agent's assessment of materiality. The additional data included disclosure date, total assets, total revenue, EBITDA, EBITDA margin, earnings-per-share, price-to-book value, return on assets, net income, business description, industry classification, and location. With this context, the LLM agent was then tasked with evaluating the materiality of each disclosure, as illustrated in Box 4.

Box 4

Example of prompt given to LLM Agent for assessing materiality of physical risk disclosure

Instruction: You are a risk analyst assessing firm disclosures for material risks arising from physical climate change risks. When I give you an extract from a firm's disclosure, please assess whether the stated risk is material for the company and investors. Materiality is the major events that shareholders should immediately know about for timely decisions. When you make the assessment, consider all available factors, not only the disclosure in isolation, but also all the relevant information available, such as the nature of the company, business areas, financials, location, etc. The assessment is based on current or future implications.

Risk statement: In recent years, California and Utah have experienced persistent and severe drought conditions.

Company background: [Company Name] operates as an independent upstream energy company in the western United States. It operates through Exploration and Production (E&P), and Well Servicing and Abandonment segments. The E&P segment engages in the development and production of onshore, low geologic risk, and long-lived conventional oil and gas reserves primarily located in California and Utah. The Well Servicing and Abandonment segment provides wellsite services in California to oil and natural gas production companies with a focus on well servicing, well abandonment services, and water logistics. The company was founded in 1909 and is headquartered in Dallas, Texas. The company is located in Dallas, TX and operates in the Energy; Energy; Oil, Gas and Consumable Fuels; Oil and Gas Exploration and Production; Oil and Gas Production industry. As of 2023-12-31, it had total assets of 1613.925 million USD and total revenue of 222.084 million USD for the quarter. The net income for the quarter was 37.4 million USD. The EBITDA for the quarter was 401.104 million USD with an EBITDA margin of 38.003%. The Basic EPS was 3.186, the price-to-book value was 0.751229, and ROA was 9.889%.

Sources: S&P; author's elaboration.

Table 1 summarizes the materiality assessments of the LLM agent, which demonstrates a capability to assess the materiality of physical risk disclosures within a given context. It effectively evaluates the implications of risks or damages by considering the company's financial position, the nature of its business, operational processes, and geographic locations.

The agent is able to differentiate between material and non-material impacts depending on the company's size and financial health. For instance, it assessed a $69.2 million damage as non-material for one company, while determining that an $8.7 million loss was material for another, reflecting their differing financial circumstances.

Additionally, the LLM agent showed an ability to connect events to a company's operational locations, identifying if risks are more severe based on whether the company operates in a primary or diversified location. It also linked specific risks to business processes. For example, it considered drought as material for companies where water is essential for drilling or production.

The agent also accounts for both short-term and long-term impacts on the business. A generic statement like "The long-term trend is that wildfires are increasing due to drought conditions" was assessed as material for a winemaker operating in fire-prone areas, demonstrating the agent's ability to recognize the significant implications of ongoing risks given the specific nature of business (see Annex for the full assessment).

Summary of LLM agent's materiality assessments for physical risk disclosures Table 1

Disclosure	Materiality	Main reason
"Approximately $0.01 per diluted share of net expense due to fire related costs."	No	• **Strong financial position** (eg EPS $1.43) • **Isolated incident**
"As more fully discussed in Note 3, in December 2021, the Company wrote off approximately $8.7 million of inventory which is held at the Solar Park contract manufacturer's site in Korea and is not expected to be recoverable."	Yes	• The **relatively large size of the write-off** 12% of total assets and 35% of quarterly revenue • Weak financial position • **Operational inefficiencies and supply chain management issues**
"Current accident year losses in 2022 included $69.2 million of net losses and loss adjustment expenses attributed to Hurricane Ian and the Russia-Ukraine conflict."	No	• This loss accounts for approximately **3.3% of the company's quarterly net income** • Isolated incident
"In recent years, California and Utah have experienced persistent and severe drought conditions."	Yes	• The company's **primary operation locations**. • **The need for water for drilling and production processes**
"We have experienced short-term impacts on our sales due to closures from weather events in recent years, including Hurricane Ian in Florida in 2022."	No	• Short-term sales impacts • **Diversified operations and geographical spread**
"The long-term trend is that wildfires are increasing resulting from drought conditions."	Yes	• The company's **reliance on vineyards in fire-prone regions** • **Financial and operational vulnerability**

Sources: S&P; OpenAI; author's calculations.

5. Conclusion

This paper outlines how recent LLM technology can present a valuable opportunity to streamline the analysis of financial disclosures, from identifying physical risk-related events to sentiment classification and materiality assessment. The ability of LLM agents to capture context and assess the materiality of risks can greatly enhance the detection of realized or potential risks, including those linked to climate change.

While assessing materiality is inherently subjective—given that the accounting definition of materiality is loosely defined and one must consider the unique needs of information users—customizing LLM agents to address specific requirements or

interests is essential. This ensures that the assessment is aligned with the particular concerns of stakeholders.

Ultimately, the final judgment on materiality remains with human experts. However, LLM technology can significantly aid this process by providing an initial evaluation, including a contextual summary tailored to the needs of the user. This enables professionals to focus on areas of particular interest, improving both the efficiency and accuracy of the risk assessment process.

Reference

BIS (2024): *Project Gaia: Enabling climate risk analysis using generative AI*, March.

Blei, DM, AY Ng and MI Jordan (2003): "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, vol 3, pp 993–1022.

Brown, T, B Mann, N Ryder, M Subbiah, J Kaplan, P Dhariwal, A Neelakantan, P Shyam, G Sastry, A Askell, S Agarwal, A Herbert-Voss, G Krueger, T Henighan, R Child, A Ramesh, D Ziegler, J Wu, C Winter, C Hesse, M Chen, E Sigler, M Litwin, S Gray, B Chess, J Clark, C Berner, S McCandlish, A Radford, I Sutskever and D Amodei (2020): "Language models are few-shot learners", *NeurIPS*.

Deerwester, S, ST Dumais, GW Furnas, TK Landauer and R Harshman (1990): "Indexing by latent semantic analysis", *Journal of the American Society for Information Science*, vol 41, no 6, pp 391–407.

Devlin, J, MW Chang, K Lee and K Toutanova (2019): "BERT: Pre-training of deep bidirectional transformers for language understanding", *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, June, pp 4171–4186.

Harris, Z (1954): "Distributional structure", *Word*, vol 10, no 2-3, pp 146–162.

Jones, K Sparck (1972): "A statistical interpretation of term specificity and its application in retrieval", *Journal of Documentation*, vol 28, no 1, pp 11–21.

Mihalcea, R and P Tarau (2004): "TextRank: Bringing order into texts", *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, July, pp 404–411.

Mikolov, T, C Corrado, K Chen and J Dean (2013): "Efficient estimation of word representations in vector space", *Proceedings of the International Conference on Learning Representations (ICLR)*, May.

Pennington, J, R Socher and C Manning (2014): "GloVe: Global vectors for word representation", *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, October, pp 1532–1543.

van der Maaten, L and G Hinton (2008): "Visualizing data using t-SNE", *Journal of Machine Learning Research*, vol 9, pp 2579–2605.

Annex – Materiality Assessment Example by LLM Agent

Material Risk Assessment: Increasing Wildfires Due to Drought Conditions

The risk statement highlights a **long-term trend of increasing wildfires**, exacerbated by drought conditions, which could affect [Company Name]. To assess whether this poses a **material risk**, we will examine the potential operational, financial, and strategic implications of this risk in the context of the company's performance and business model.

1. Company Background and Operations:

- **Industry**: [Company Name] operates in the **wine and spirits industry**, producing and selling wines under various brands, including [Brand A], [Brand B], and [Brand C], among others. In addition to producing wines, the company owns and operates **hospitality facilities**, providing services like bottling, fulfillment, and storage to other companies.
- **Geographical Presence**: The company operates primarily in the **United States** and **Canada**, regions that have seen increasing wildfire activity, particularly in **California** and other key wine-producing areas. Wine production is highly dependent on climatic conditions, and **wildfires** are a known threat to vineyards, wine quality, and overall operations.
- **Financial Position**: [Company Name] has filed for **Chapter 11 bankruptcy** (as of July 24, 2024), indicating financial distress. The company's **net income of -$190.229 million** for the quarter and **low EBITDA margin of 3.756%** suggest it is operating with very limited profitability and is highly vulnerable to additional financial shocks. Any significant disruption, like wildfires, could further strain its financial position.

2. Nature of the Risk (Wildfires Due to Drought):

- **Operational Impact**:
    - **Damage to Vineyards**: Wildfires pose a direct threat to **vineyards**, which are highly susceptible to damage from fire, heat, and smoke. Even if vineyards are not directly burned, **smoke taint** can spoil grape quality, leading to reduced production volumes or lower-quality wine.
    - **Disruption to Harvesting and Production**: Wildfires can disrupt vineyard operations by **forcing evacuations**, limiting access to properties, and damaging equipment and infrastructure (e.g., irrigation systems). This can lead to delays or cancellations of **harvests** and interruptions in wine production, directly affecting revenues.
    - **Hospitality Facilities**: The company also operates **hospitality facilities**, which could face closures or reduced visitor numbers during wildfire seasons, particularly if air quality deteriorates or if there are safety concerns in affected areas.
    - **Supply Chain Disruptions**: Wildfires could disrupt the **logistics and supply chains** for bottling and distribution, particularly in California and other regions with key infrastructure.
- **Financial and Operational Flexibility**:
    - **Bankruptcy Proceedings**: The company is already in **Chapter 11 bankruptcy**, meaning it is undergoing reorganization to address its financial challenges. Wildfires and the resulting disruptions could significantly hinder

recovery efforts by causing additional **revenue losses** or increasing operational costs (e.g., for repairing damaged vineyards or investing in smoke-taint mitigation technologies).

- **Limited Financial Resilience**: With a **low EBITDA margin of 3.756%**, the company is operating on thin margins. Any additional operating costs due to wildfires (e.g., damaged crops, reduced yields, or lower wine quality) could further reduce profitability, complicating its ability to stabilize financially.

- **Long-Term Climate Change Implications**: Wildfires, exacerbated by drought conditions, are expected to become more frequent and severe due to **climate change**. For a wine producer like [Company Name], this represents a significant long-term risk to vineyard operations, product quality, and regional tourism, all of which are critical to the company's business model.

3. Materiality of the Risk:

- **Direct Financial Impact**:
  - **Revenue and Production**: The company's quarterly revenue of **$64.651 million** is already under pressure, and any reduction in production due to wildfires could further decrease revenues. **Vineyards are capital-intensive**, and rebuilding or rehabilitating vineyards after a wildfire can take years, making the risk of wildfires particularly costly.
  - **Net Income**: With a **net income of -$190.229 million**, [Company Name] is already facing significant financial challenges. Even minor disruptions from wildfires could exacerbate these losses, given the thin profit margins and ongoing bankruptcy proceedings.
  - **Asset Impairment**: The company has **total assets of $626.446 million**, and wildfires could lead to asset impairments if vineyards, equipment, or hospitality facilities are damaged. This could reduce the value of the company's assets further, negatively impacting its already low **price-to-book ratio of 0.23**, indicating the market already perceives significant risk.
- **Operational Disruptions**:
  - **Wildfire Frequency**: The risk statement emphasizes that **wildfires are increasing**, suggesting that operational disruptions are not limited to a single event but represent an ongoing and growing threat. This increased frequency could lead to recurring disruptions in production cycles and long-term damage to grapevines, impacting both short-term revenue and long-term viability.
  - **Wine Quality**: Smoke taint can severely affect the **quality of wine**, leading to potential reputational damage if the company produces wine that fails to meet customer expectations. Lower-quality wine might need to be sold at a discount or discarded altogether, further impacting profitability.

4. Long-Term Outlook and Strategic Considerations:

- **Climate Change and Sustainability**: Given that wildfires are likely to increase due to **climate change**, [Company Name] may need to invest in **mitigation strategies**, such as enhanced fire protection for its vineyards, more robust irrigation systems to combat drought, or **smoke-taint mitigation technologies** (e.g., air filtration, protective coverings). However, given the company's financial situation, it may lack the capital required to make such investments.
- **Reputation and Market Position**: [Company Name] may face reputational challenges if it is unable to consistently deliver high-quality wines due to wildfire-

related disruptions. This could lead to **loss of market share** in an already competitive industry where brand loyalty and product quality are critical to success.

- **Long-Term Financial Viability**: The combination of bankruptcy proceedings and increased wildfire risk places significant pressure on the company's long-term viability. The ability to successfully emerge from Chapter 11 will depend on how well the company can mitigate climate risks while stabilizing its finances.

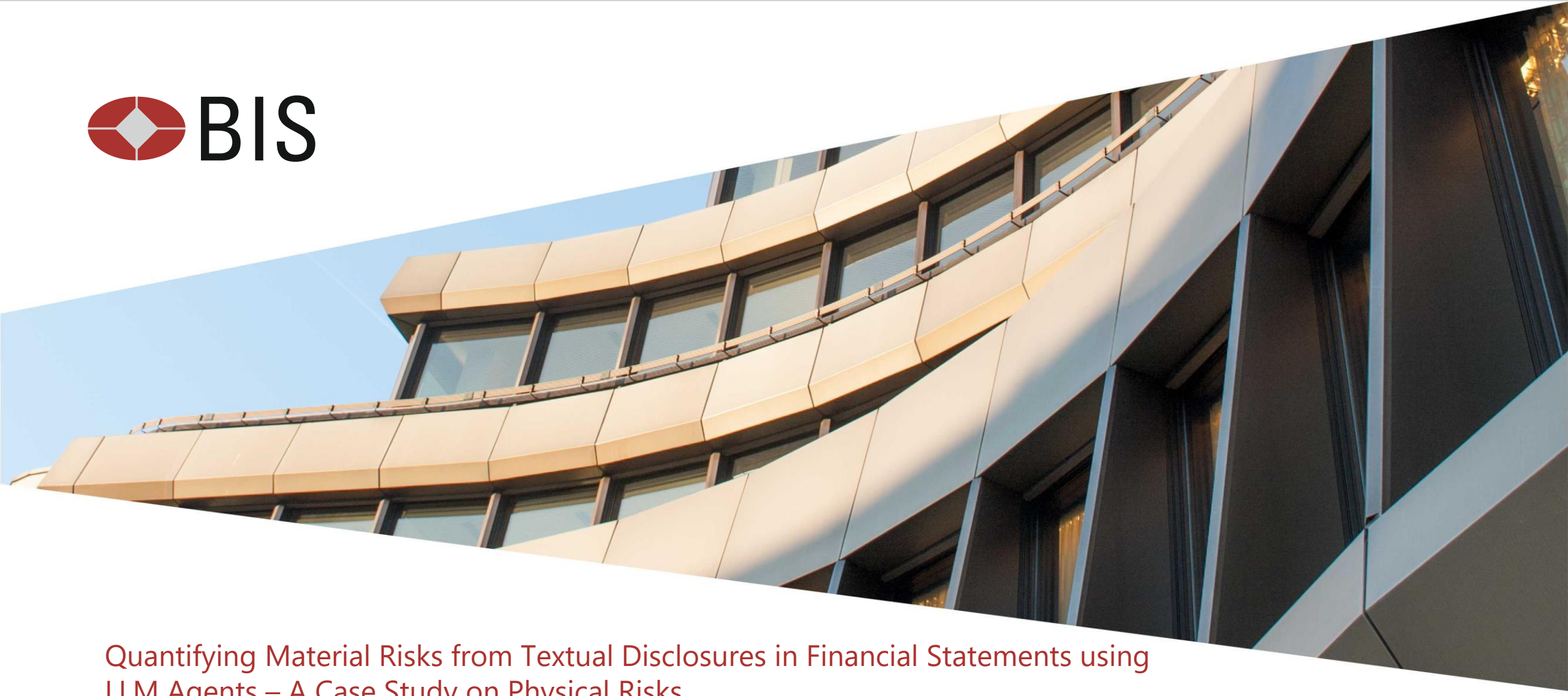
5. Investor Considerations:

- **Material Risk for Investors**: The **increasing frequency of wildfires** due to drought conditions represents a **material risk** for [Company Name]. The company's fragile financial position, combined with its reliance on climate-sensitive assets (vineyards), makes it highly vulnerable to wildfire disruptions. Investors should be concerned about the potential for continued operational and financial deterioration if the company cannot manage wildfire risks effectively.
- **Bankruptcy and Recovery**: The company's ongoing **Chapter 11 bankruptcy** adds another layer of complexity. Any significant operational disruptions caused by wildfires could delay or derail the company's recovery plan, potentially leading to further financial distress or even liquidation if the reorganization is unsuccessful.
- **Climate Risk as a Long-Term Threat**: Investors focused on **ESG** factors may see the increasing wildfire risk as part of a broader **climate risk** concern. [Company Name] may need to prioritize sustainability and resilience strategies in order to remain competitive in an industry increasingly impacted by climate change.

Conclusion: Material Risk Assessment

- **Material Risk**: The **long-term trend of increasing wildfires** due to drought conditions poses a **material risk** to [Company Name]. Given the company's reliance on vineyards in fire-prone regions, any damage to vineyards, disruptions in production, or quality issues caused by smoke taint could significantly impact its operations and financial performance.
- **Financial and Operational Vulnerability**: [Company Name] is already under financial strain, with a **negative net income** and **low EBITDA margin**. Wildfires could exacerbate this situation, leading to further losses, operational disruptions, and a delayed recovery from Chapter 11 bankruptcy.
- **Long-Term Considerations**: Wildfires and climate risks will likely continue to challenge the company's long-term viability, requiring strategic investments in mitigation efforts that may be difficult to implement given its financial constraints.

In conclusion, **increasing wildfires** due to drought conditions represent a **material risk** for [Company Name], particularly given its current financial instability and reliance on climate-sensitive assets.



Quantifying Material Risks from Textual Disclosures in Financial Statements using LLM Agents – A Case Study on Physical Risks

Taejin Park

Workshop on New insights from Financial Statements

Madrid, 17 October 2024

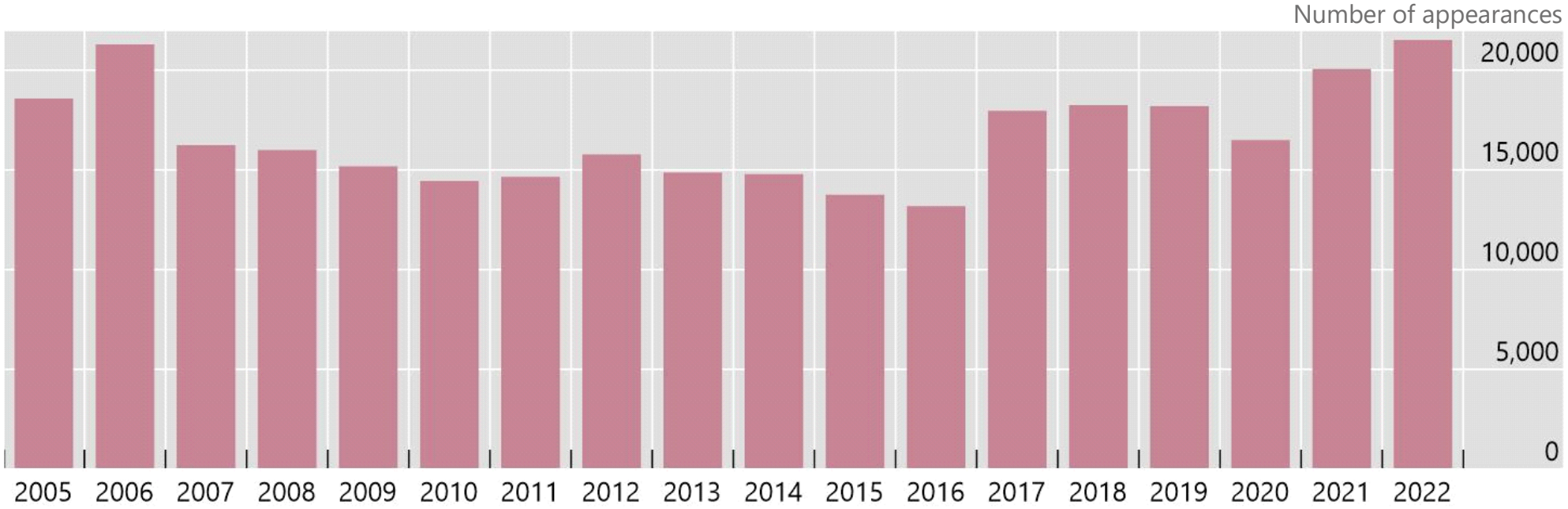
Introduction

- Data on physical risks largely rely on **disclosures**.
- Gathering and analysing them can be **overwhelming**.
- **Recent LLM technologies** offer scalable solutions.



Rising Trends in Disclosure

Frequency of physical risk-related keywords in 10-K filings



Included keywords: flood, wildfire, hurricane, Flood, typhoon, Hurricane, drought, Wildfire, heat wave, extreme heat, Typhoon, extreme cold, Drought, Extreme heat, heatwave, FLOOD, WildFire, WILDFIRE, HeatWave, HURRICANE, cold wave, Extreme cold, coldwave, Heat Wave, TYPHOON, Heat wave, Extreme Heat, Heatwave, Extreme Cold, DROUGHT and HEAT WAVE.

Sources: S&P; author's calculations.

Key Challenges

How to detect info embedded in complex or indirect language?

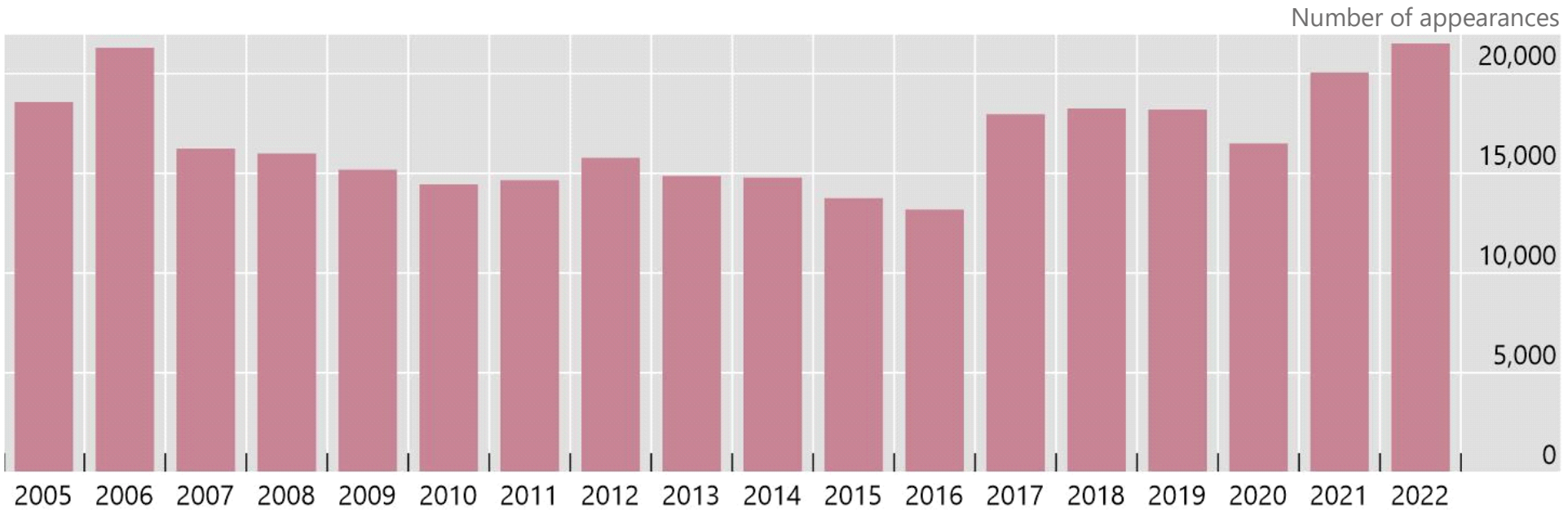
Reflecting the **specific context** and relevance to each user.



Not all risk disclosures are negative.

Extracting Relevant Information: Traditional Keyword-Based Approach

Frequency of physical risk-related keywords in 10-K filings



Included keywords: flood, wildfire, hurricane, Flood, typhoon, Hurricane, drought, Wildfire, heat wave, extreme heat, Typhoon, extreme cold, Drought, Extreme heat, heatwave, FLOOD, WildFire, WILDFIRE, HeatWave, HURRICANE, cold wave, Extreme cold, coldwave, Heat Wave, TYPHOON, Heat wave, Extreme Heat, Heatwave, Extreme Cold, DROUGHT and HEAT WAVE.

Sources: S&P; author's calculations.

Extracting Relevant Information: Traditional Keyword-Based Approach

- Main Challenges:
  - How do you create a truly **exhaustive list** of keywords for risk identification?
    - What about synonyms, polysemy... (eg bank, bond)
  - One can **describe a risk without explicitly mentioning those keywords**.

Extracting Relevant Information: Traditional Keyword-Based Approach

- Examples of complex language
  - "We anticipate a modest decrease in harvest volumes versus the prior year as logging conditions normalize following a period of relatively **dry weather**." (Rayonier Inc., 8-K, 31-01-2024)
  - "Because substantially all of our products are manufactured and assembled in and distributed from a few locations, our operations could be interrupted by events beyond our control, including: … **water damage** or **water shortage** …" (Acushnet Holdings Corp., 10-K, 01-03-2023)
  - "Both parks experienced unusually **inclement weather** during that quarter." (Cedar Fair, L.P., 10-K, 16-02-2024)
  - "FY23 Saleable production levels are lower than prior year due to the impacts from these **poor geotechnical conditions** and the impacts from the rock intrusion in the September quarter, that slowed production rates and impacted yield." (Coronado Global Resources Inc., 8-K, 22-01-2024)
  - "…**warmer weather in the heating season** and **cooler weather in the cooling season** will reduce our customers' energy demand and our retail operating revenues." (Avista Corporation,10-K, 2024-02-21)

Extracting Relevant Information: Recent Embedding-based Approach

“Prolonged drought conditions increase the intensity of competition for fresh water.”

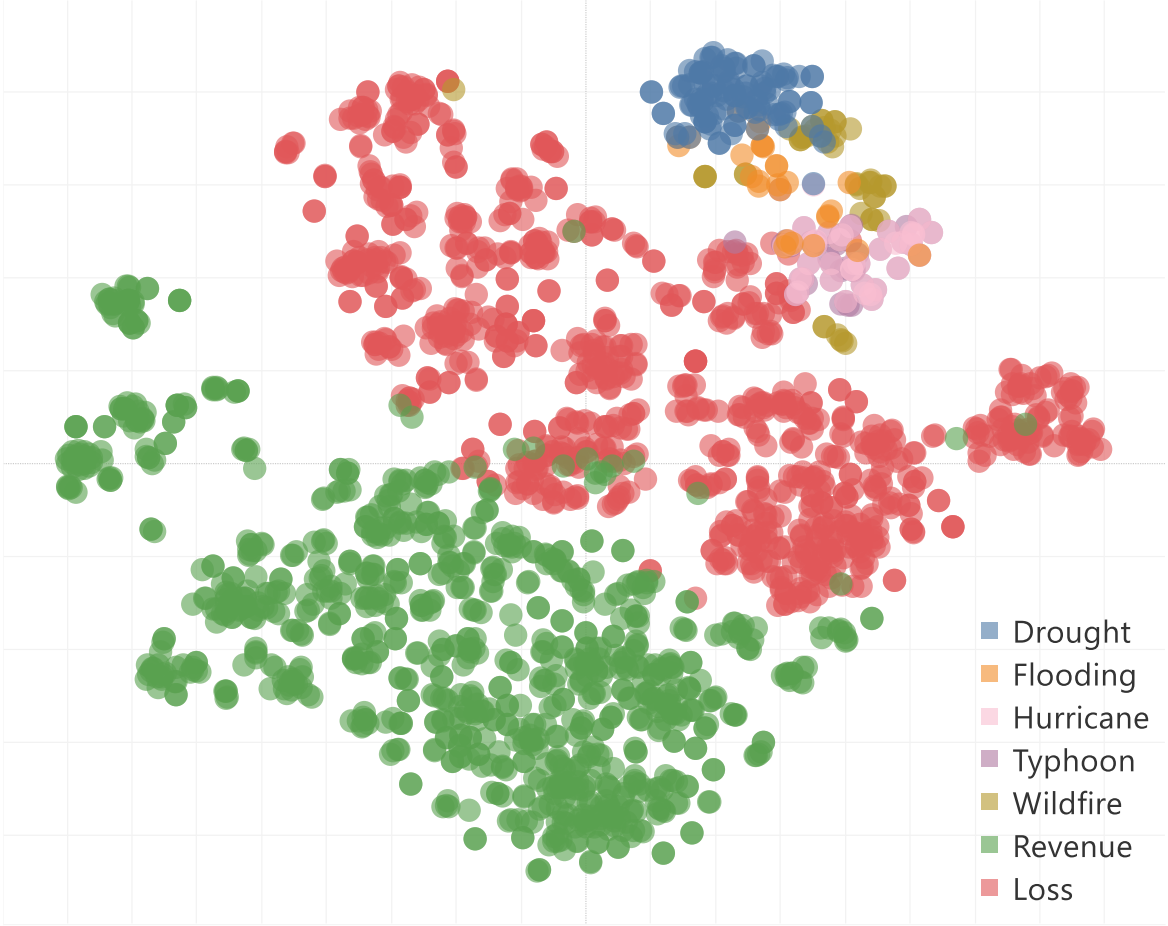


[0.039094422012567 52, 0.025339428335428238, 0.07760352641334407, … 0.012559967115521431]

→1,536 dimensions

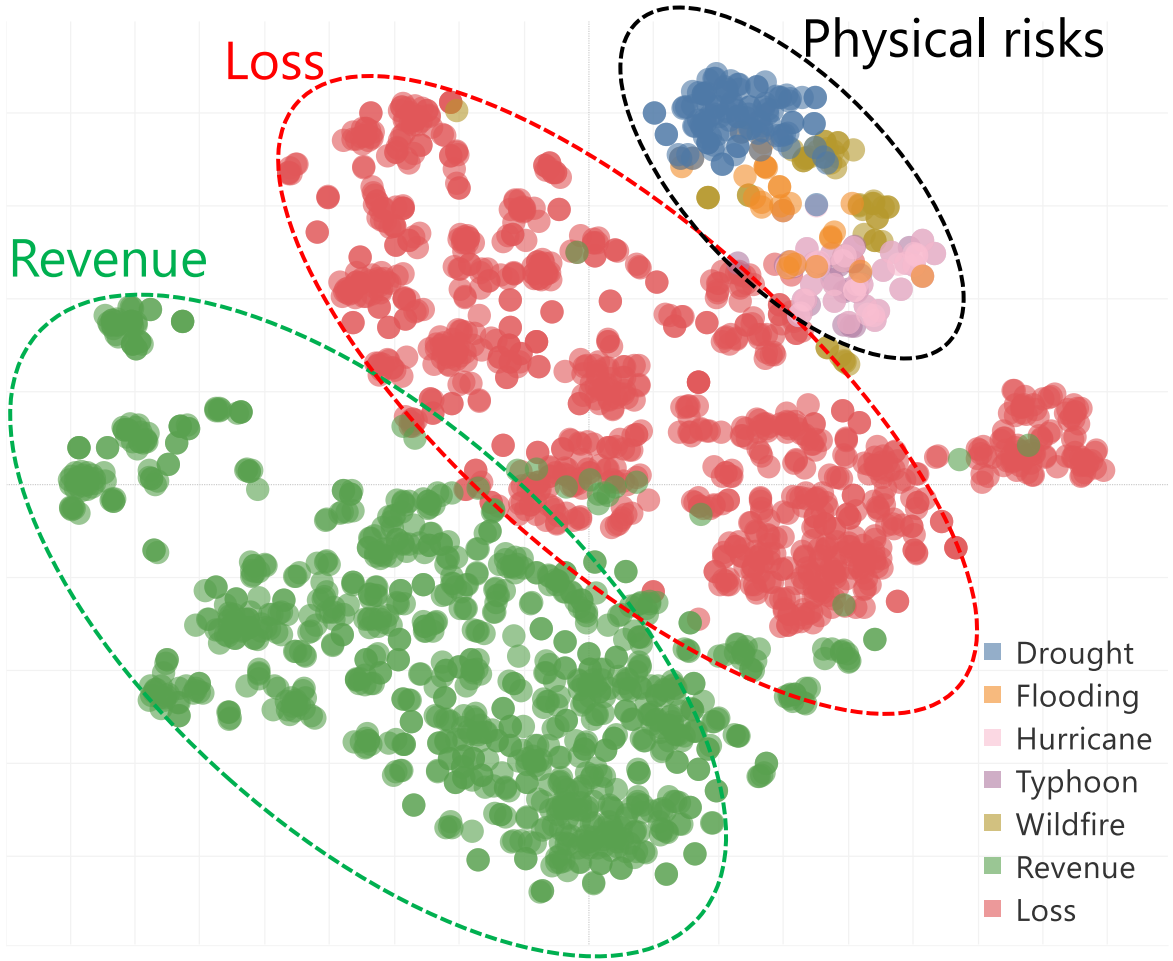
- Enabling semantic search

Recent NLP approach – Embedding-based approaches



- Physical risks & general financial topics
- Dimension reduction: 1,536 → 2 using t-SNE

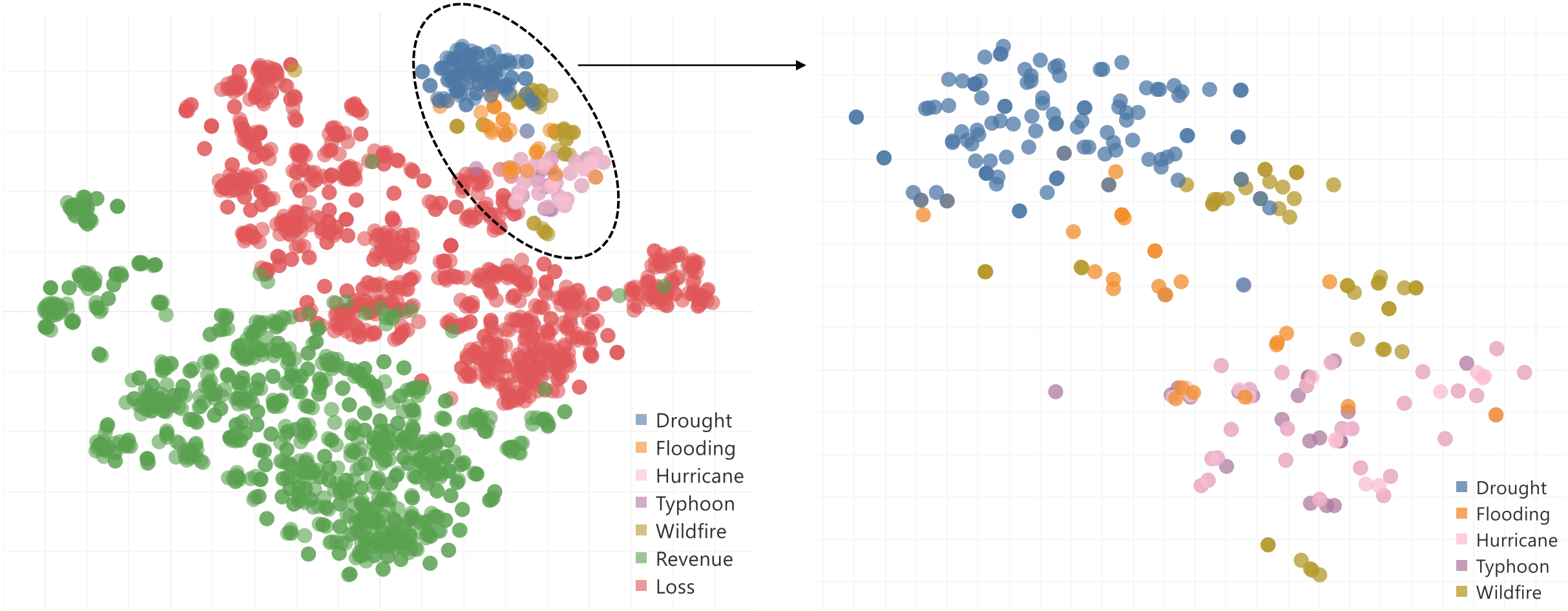
Recent NLP approach – Embedding-based approaches



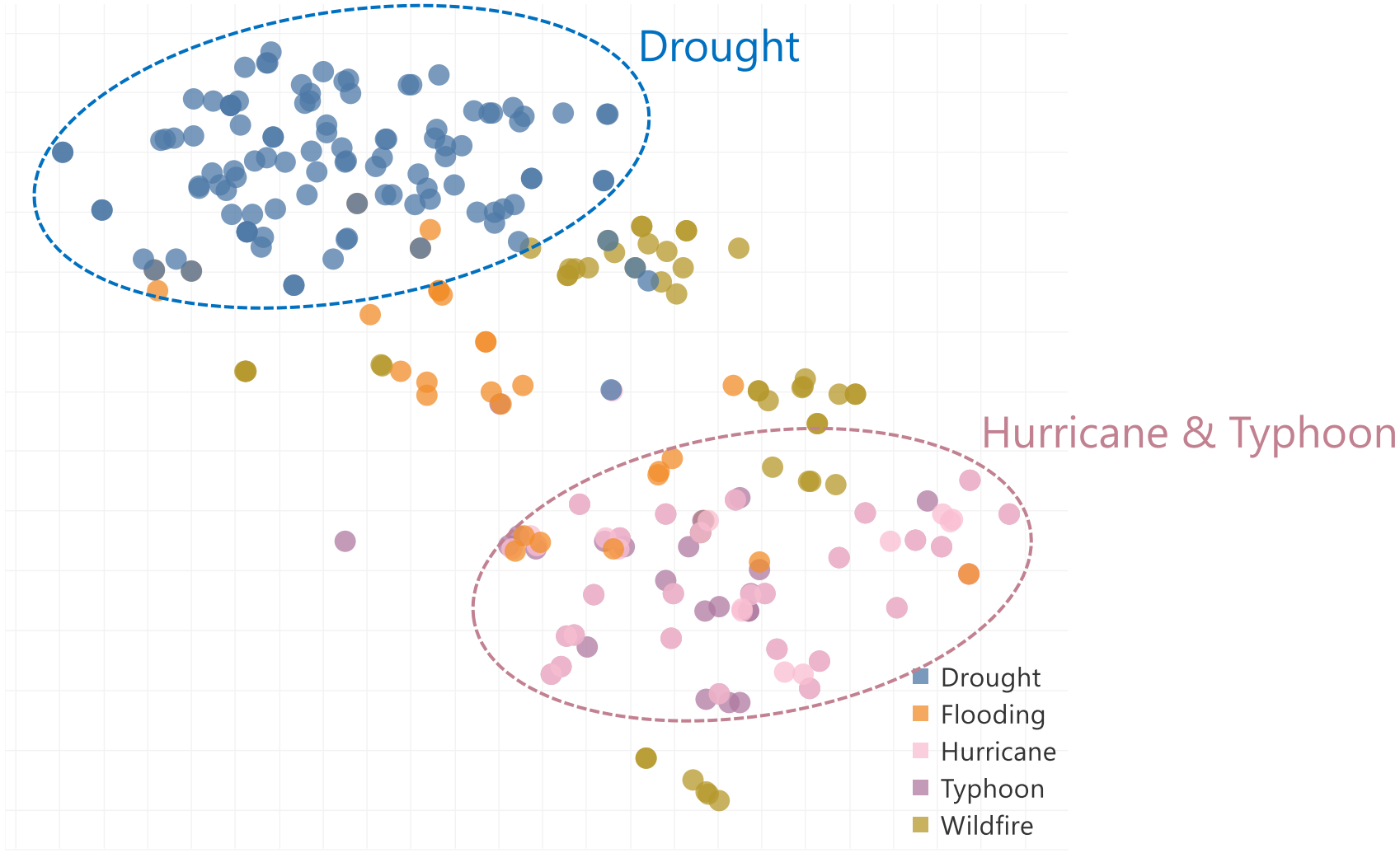
Even with the highly compressed dimensions,

- Clear emergence of clusters by topic
- Physical risk-related cluster
- Physical risks (semantically) close to "loss" cluster

Recent NLP approach – Embedding-based approaches

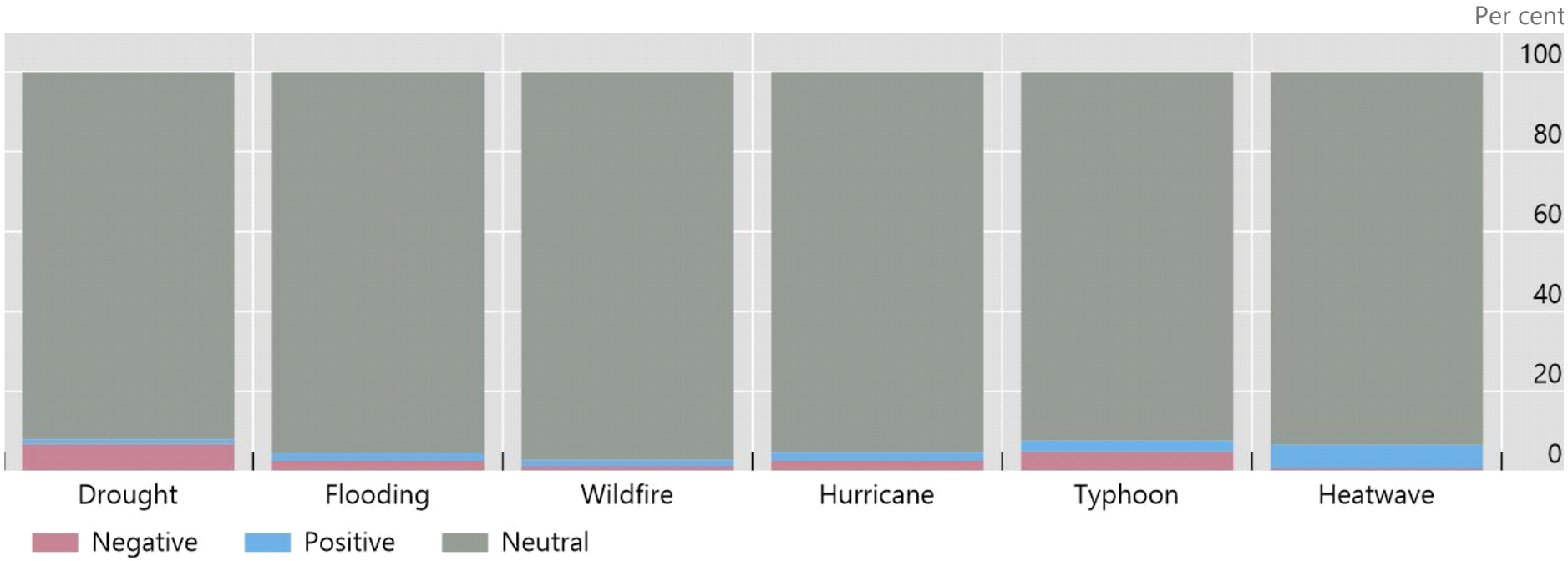


Recent NLP approach – Embedding-based approaches



Assessing Risk Sentiment is Essential

Sentiment analysis of physical risk disclosures



The analysis is based on a random selection of 1,000 sentences per category related to physical risk disclosures.

Sources: S&P; ProntoNLP; author's calculations.

- Fine-tuned LLMs or In-context learning can be used.

Contextual and Subjective Nature of Materiality Assessment: Examples

- "Approximately **$0.01 per diluted share** of net expense due to fire related costs."
- "As more fully discussed in Note 3, in December 2021, the Company wrote off approximately **$8.7 million of inventory** which is held at the Solar Park contract manufacturer's site in Korea and is not expected to be recoverable."
- "Current accident year losses in 2022 included **$69.2 million of net losses** and loss adjustment expenses attributed to Hurricane Ian and the Russia-Ukraine conflict."
- "In recent years, **California and Utah** have experienced persistent and severe drought conditions."
- "We have experienced **short-term impacts on our sales** due to closures from weather events in recent years, including Hurricane Ian in Florida in 2022."
- "The **long-term trend** is that wildfires are increasing resulting from drought conditions."

The need for context-aware analysis to assess the materiality

- Disclosure, Disclosure date, Filing type



- Business description
- Industry classification
- Location
- Total assets
- Total revenue
- EBITDA
- EBITDA margin
- Earnings-per-share
- Price-to-book value
- Return on Assets
- Net income, …



LLM Agent
to assess materiality

```
def generate_description(row):
    description = (
        f"{row['Business Description']} "
        f"The company is located in {row['Location']} and operates in the {row['Industry Classification']} industry. "
        f"As of {row['Date']}, it had total assets of {row['Total Assets (mn USD)']} million USD and total revenue of {row['Total Revenue (Qtr; mn USD)']} million USD for the quarter. "
        f"The net income for the quarter was {row['Net Income (Qtr; mn USD)']} million USD."
        f"The EBITDA for the quarter was {row['EBITDA (Qtr)']} million USD with an EBITDA margin of {row['EBITDA Margin']}%. "
        f"The Basic EPS was {row['EPS']}, the price-to-book value was {row['Price-to-book value']}, and ROA was {row['ROA']}%. "
    )
    return description
```

Context-aware analysis

Disclosure	Summary of responses by the LLM Agent	
	Materiality	**Main reason**
"Approximately $0.01 per diluted share of net expense due to fire related costs."	**No**	• Strong financial position (eg EPS $1.43) • Isolated incident
"As more fully discussed in Note 3, in December 2021, the Company wrote off approximately $8.7 million of inventory which is held at the Solar Park contract manufacturer's site in Korea and is not expected to be recoverable."	**Yes**	• The relatively large size of the write-off 12% of total assets and 35% of quarterly revenue • Weak financial position • Operational inefficiencies and supply chain management issues
"Current accident year losses in 2022 included $69.2 million of net losses and loss adjustment expenses attributed to Hurricane Ian and the Russia-Ukraine conflict."	**No**	• This loss accounts for approximately 3.3% of the company's quarterly net income. • Isolated incident
"In recent years, California and Utah have experienced persistent and severe drought conditions."	**Yes**	• The company's primary operation locations. • The need for water for drilling and production processes.
"We have experienced short-term impacts on our sales due to closures from weather events in recent years, including Hurricane Ian in Florida in 2022."	**No**	• Short-term sales impacts • Diversified operations and geographical spread
"The long-term trend is that wildfires are increasing resulting from drought conditions."	**Yes**	• The company's reliance on vineyards in fire-prone regions. • Financial and operational vulnerability

Conclusion

- **Streamlining Analysis**: LLMs help automate financial disclosure analysis—from identifying physical risks to sentiment and materiality assessment.
- **Context Matters**: LLMs' ability to capture context boosts accuracy in detecting risks, including climate-related ones.
- **Subjectivity in Materiality**: Tailoring LLMs to users' needs is crucial, given the subjective nature of materiality assessments.
- **Augmenting Human Expertise**: LLMs provide initial evaluations, allowing experts to focus on key risks and enhance decision-making.

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

The allocation of cash flow by Spanish firms: new evidence on the impact of financial frictions[1]

Martin Farias,
Center for Monetary and Financial Studies (CEMFI)

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

The Allocation of Cash Flow by Spanish Firms: New Evidence on the Impact of Financial Frictions*

Martin Farias †
CEMFI

Abstract

This paper studies the impact of financial frictions on the allocation of cash flow using administrative data of Spanish non-financial corporations from 2003 to 2019. Employing an analytical framework based on the uses and sources of funds identity, I estimate regression models to examine the allocation of cash flow across its competing uses. My findings reveal that larger financial frictions are, on average, associated with a higher proportion of marginal cash flow allocated to debt repayment and lower proportions allocated to cash savings, investment, and dividend distribution. The analysis also highlights that the effect of financial frictions on the allocation of marginal cash flow varies significantly with variables capturing the economic and financial situation of each firm such as leverage, cash holdings, capital, and the availability of investment opportunities.

JEL codes: G30, G32, G35.
Keywords: Financial frictions, cash flow, investment, dividends, debt, cash savings.

*I would like to extend my gratitude to my advisor, Javier Suarez, for his support and invaluable guidance. Special thanks are also due to Samuel Bentolila and Rafael Repullo for their contributions at every stage of this project. I want to express my gratitude to the participants of the Banking and Finance seminar at CEMFI, with special thanks to Vedant Agarwal for our insightful discussions and exchanges. I acknowledge financial support from the Maria de Maeztu Unit of Excellence CEMFI MDM-2016-0684, funded by MCIN/AEI/10.13039/501100011033, the Santander Research Chair, Fundación Carolina and CEMFI.

†Contact: Email: martin.farias@cemfi.edu.es

1 Executive Summary

This paper is motivated by the fact that cash flow is one of the primary funding sources for firms, enabling them to allocate funds towards cash savings, investment, debt repayment, and net dividends. Moreover, financial frictions limit the availability and increase the cost of funding coming from external debt and equity issuance, which makes cash flow a more relevant determinant of these uses. In this paper, I use administrative data on the quasi-universe of Spanish firms to show empirically how these firms allocate their marginal cash flow to alternative uses, and how this allocation is affected by the presence of financial frictions.

I build the empirical analysis on a cash flow identity that equates cash flow to the sum of all its alternative uses. My baseline specifications consist of regressing the different uses of cash flow on cash flow, control variables, and firm and year fixed effects. The coefficients associated with cash flow, which are deemed *cash flow sensitivities*, measure how much of a marginal unit of cash flow is allocated to the corresponding use. I find that, at the margin, an additional euro of cash flow is followed by average increases of 29 cents in cash savings, 44 cents in investment, 23 cents in debt repayment, and 4 cents in dividend distribution.

To study how financial frictions affect the allocation of marginal cash flow by firms, I extend the baseline specifications with the interaction of the cash flow variable with a proxy for the potential importance of financial frictions. The proxy I use is the *Size-Age* (*SA*) index, which captures the idea that younger and smaller firms are more likely to face financial frictions.

My findings for Spanish firms during the period 2003-2019 indicate that higher values of the *SA* index are associated, on average, with higher cash flow sensitivities of debt repayment and lower cash flow sensitivities of each of the other uses. Specifically, a one-unit increase in the *SA* index (whose standard deviation is 0.85) is associated with an average increase of 6 cents in the share of marginal cash flow allocated towards debt repayment and average decreases of 2 cents, 3 cents, and 1 cent in the shares allocated to cash savings, investment, and net dividends, respectively. These results support the notion that firms facing larger financial frictions in Spain during the analyzed period employ a larger portion of their cash flow to mitigate these frictions by reducing their reliance on external debt.

Next, I explore whether the impact of potential financial frictions on cash flow sensitivities further varies with variables that are typically treated as *state variables* in structural dynamic models of firms' behavior: leverage, cash holdings, capital, and availability of investment opportunities. The empirical analysis consists of splitting the observations in my sample based on lagged values of these variables and comparing the effects of financial frictions on the allocation of cash flow across the resulting subsamples.

The main results from this exercise are the following. First, changes in the *SA* index imply minor or insignificant differences in the cash flow sensitivities among firms in the lowest deciles of the distribution of leverage. In contrast, higher values of the *SA* index are associated with a higher cash flow sensitivity of debt repayment and lower cash flow sensitivities of cash saving and investment for firms in the highest deciles of the distribution of leverage. Second, the cash flow sensitivities seem to be unaffected by changes

in the *SA* index among firms in the highest deciles of the distribution of cash holdings. However, higher values of the *SA* index are associated with a lower cash flow sensitivity of investment and a higher cash flow sensitivity of debt repayment for firms in the lowest deciles of the distribution of cash holdings. Third, the cash flow sensitivity of investment is decreasing in the *SA* index only among firms in the highest deciles of the distribution of capital. For firms in the lowest deciles of the distribution of capital, higher values of the *SA* index are associated with a higher cash flow sensitivity of debt repayment, lower cash flow sensitivities of cash savings and dividend distribution, and have no significant association with the cash flow sensitivity of investment.

Collectively, these results suggest that the proxy of financial frictions captures the likelihood of firms experiencing financial constraints, significantly altering the average cash flow sensitivities of the various uses of cash flow. However, the specific economic and financial situation of each firm (as captured by standard pre-determined state variables) interacts with the likelihood of suffering financial constraints creating a rich and nuanced landscape. Uncontrolled-for variability in those and other relevant state variables across samples might explain the apparently conflicting findings of the literature that has tried to establish if and how financial constraints affect the allocation of cash flow with reduced form methodologies. It also highlights the value that more structural approaches (able to simultaneously account for the dynamics of the relevant state variables) may have in this field.

Furthermore, the findings bear direct relevance to policymakers. The analysis of heterogeneous effects of financial frictions offers valuable insights to enhance the precision of policies, such as direct transfers or subsidized credit guarantees, targeting specific firms for purposes such as promoting investment. These results contribute to anticipating the response of certain firms when provided with additional cash flow (in the case of a direct transfer) or increased borrowing capacity (in the case of a subsidized credit guarantee).

2 Introduction

Cash flow is one of the primary funding sources for firms, enabling them to allocate funds towards cash savings, investment, debt repayment, and net dividend distributions. Financial frictions limit the availability and increase the cost of funding coming from external debt and equity issuance, which makes cash flow a more relevant determinant of these uses. The seminal paper of Fazzari, Hubbard, and Petersen (1987) offers evidence suggesting that financially constrained firms feature greater sensitivity of investment to cash flow. Since then, many theoretical and empirical studies have aimed at confirming the generality of this finding and characterizing differential patterns across more and less financially constrained firms in other uses of cash flow.

In this paper, I add to this literature by analyzing the allocation of cash flow to alternative uses by Spanish non-financial corporations during the period 2003-2019. I extend the standard approach by interacting a proxy for financial frictions with pre-determined variables describing the economic and financial situation of each firm. My findings suggest that larger financial frictions are, on average, associated with a higher share of cash flow allocated to debt repayment, and with lower shares allocated to cash savings, in-

vestment, and dividend distribution. Importantly, my analysis shows that the effects of the proxy for financial frictions on the allocation of cash flow vary significantly with predetermined state variables capturing the economic and financial situation of each firm, such as leverage, cash holdings, capital stock, and availability of investment opportunities.

As in Chang, Dasgupta, Wong, and Yao (2014), I build the empirical analysis on a cash flow identity (akin to a simplified uses and sources of funds identity) that equates cash flow to the sum of all its alternative (net) uses. My baseline specifications consist of regressing the different uses of cash flow on cash flow, control variables, and firm and year fixed effects. I use ordinary least squares (OLS) regressions to estimate each equation separately. The coefficients associated with cash flow, which are deemed *cash flow sensitivities*, measure how much of a marginal unit of cash flow is allocated to the corresponding use. I find that, at the margin, an additional euro of cash flow is followed by average increases of 29 cents in cash savings, 44 cents in investment, 23 cents in debt repayment, and 4 cents in dividend distribution. By construction, the estimated cash flow sensitivities add up to one, as noticed by Chang et al. (2014).[1]

To study how financial frictions affect the allocation of marginal cash flow by firms, I extend the baseline specifications with the interaction of the cash flow variable with a proxy for the potential importance of financial frictions. The proxy I use is the *Size-Age* (*SA*) index proposed by Hadlock and Pierce (2010), which captures the idea that younger and smaller firms are more likely to face financial frictions.

Conventional wisdom sustains that firms facing larger financial frictions should allocate a higher share of their cash flow to investment, the rationale being that investment is more profitable at the margin for constrained than for unconstrained firms because of the difficulty of financing it externally (Fazzari et al., 1987). Consistent with this view, many empirical papers show that financial frictions are associated with higher cash flow sensitivities of investment (Fazzari et al., 2000; Drobetz et al., 2017). However, firms facing larger financial frictions might also find it differentially attractive to allocate cash flow towards other uses such as cash savings or debt repayment, since these might contribute to alleviating present and future financial frictions (Dasgupta and Sengupta, 2007; DeAngelo et al., 2011; Denis and McKeon, 2012). Thus, whether investment or other uses of cash flow exhibit larger or lower average sensitivity to cash flow in the presence of financial constraints remains an empirical question that can have different answers in different samples and time periods. Indeed, several studies find that firms facing larger financial frictions allocate a lower share of their cash flow towards investment, and higher shares towards cash savings and debt repayment (Cleary, 1999; Chang et al., 2014).

My findings for Spanish firms during the period 2003-2019 indicate that higher values of the *SA* index are associated, on average, with higher cash flow sensitivities of debt repayment and lower cash flow sensitivities of each of the other uses. Specifically, a one-unit increase in the *SA* index (whose standard deviation is 0.85) is associated with an

[1] This property is a direct implication of the cash flow identity. While the system of equations can be estimated using seemingly unrelated regressions (SUR) to enforce that the sum of the coefficients associated with cash flow equals one, SUR estimates are equivalent to equation-by-equation OLS estimates if the same set of explanatory variables is included in each equation (Greene, 2012).

average increase of 6 cents in the share of marginal cash flow allocated towards debt repayment and average decreases of 2 cents, 3 cents, and 1 cent in the shares allocated to cash savings, investment, and net dividends, respectively. These results support the notion that firms facing larger financial frictions in Spain during the analyzed period employ a larger portion of their cash flow to mitigate these frictions by reducing their reliance on external debt.

Next, I explore whether the impact of potential financial frictions on cash flow sensitivities further varies with variables that are typically treated as *state variables* in structural dynamic models of firms' behavior (e.g., Gao, Whited, and Zhang, 2021): leverage, cash holdings, capital, and availability of investment opportunities. The empirical analysis consists of splitting the observations in my sample based on lagged values of these variables and comparing the effects of financial frictions on the allocation of cash flow across the resulting subsamples.

The main results from this exercise are the following. First, changes in the *SA* index imply minor or insignificant differences in the cash flow sensitivities among firms in the lowest deciles of the distribution of leverage. In contrast, higher values of the *SA* index are associated with a higher cash flow sensitivity of debt repayment and lower cash flow sensitivities of cash saving and investment for firms in the highest deciles of the distribution of leverage. Second, the cash flow sensitivities seem to be unaffected by changes in the *SA* index among firms in the highest deciles of the distribution of cash holdings. However, higher values of the *SA* index are associated with a lower cash flow sensitivity of investment and a higher cash flow sensitivity of debt repayment for firms in the lowest deciles of the distribution of cash holdings. Third, the cash flow sensitivity of investment is decreasing in the *SA* index only among firms in the highest deciles of the distribution of capital. For firms in the lowest deciles of the distribution of capital, higher values of the *SA* index are associated with a higher cash flow sensitivity of debt repayment, lower cash flow sensitivities of cash savings and dividend distribution, and have no significant association with the cash flow sensitivity of investment.

Collectively, these results suggest that the proxy of financial frictions (the *SA* index of Hadlock and Pierce, 2010) captures the likelihood of firms experiencing financial constraints, significantly altering the average cash flow sensitivities of the various uses of cash flow. However, the specific economic and financial situation of each firm (as captured by standard pre-determined state variables) interacts with the likelihood of suffering financial constraints creating a rich and nuanced landscape. Uncontrolled-for variability in those and other relevant state variables across samples might explain the apparently conflicting findings of the literature that has tried to establish if and how financial constraints affect the allocation of cash flow with reduced form methodologies. It also highlights the value that more structural approaches (able to simultaneously account for the dynamics of the relevant state variables) may have in this field.

Finally, I extend the baseline specifications to conduct further analysis that sheds light on other interesting aspects of the allocation of cash flow. First, by dissecting cash flow into its persistent and transitory components, I find that firms allocate these distinct components differently, with a greater proportion of the persistent component channeled into investment, consistent with it containing more information about firms' future profitability. Second, I disaggregate investment into its three components (real investment, finan-

cial investment, and trade credit provision), and I find that almost all of the cash flow sensitivity of investment is attributable to the cash flow sensitivities of real investment and trade credit provision. Third, I add two lags of the cash flow variable as additional regressors to explore the existence of slow adjustment in the allocation of cash flow. The results suggest that while some delays exist in the allocation of cash flow to investment, such delays are generally of low magnitude. Fourth, in examining asymmetries between positive and negative cash flow, I find that the main uses to which a marginal unit of positive cash flow is allocated are cash savings and investment, while the main sources that firms use to accommodate an additional unit of negative cash flow are increases in debt and reductions in investment.

Related Literature. My paper relates directly to the literature that studies the effects of financial frictions on the allocation of cash flow. The existing empirical literature explores how financial frictions affect the cash flow sensitivities of investment (Fazzari et al., 1987; Kaplan and Zingales, 1997; Allayannis and Mozumdar, 2004), cash savings (Almeida et al., 2004; Riddick and Whited, 2009), or of all the uses in a unified empirical framework (Gatchev et al., 2010; Chang et al., 2014; Drobetz et al., 2017). My contribution consists of extending the empirical approach of Chang et al. (2014) to explore whether the effects of financial frictions on the allocation of cash flow are contingent upon the economic and financial situation of each firm (as captured by pre-determined state variables such as leverage, cash holdings, capital and the availability of investment opportunities).

Other papers in the literature have already explored heterogeneity in the effects of financial frictions on the allocation of cash flow. For instance, Drobetz et al. (2017) compare the effects of financial frictions in the allocation of cash flow before and after the 2007 crisis. Other papers have explored heterogeneity across the business cycle (e.g., Almeida et al. (2004)), or across the persistent and temporary components of cash flow (e.g., Chang et al. (2014)), for instance. Instead, I present a general overview of heterogeneity across variables that are usually treated as state factors in structural models of firms' behavior. This list of variables is seen as a comprehensive summary, capturing all the important information firms need for deciding how to allocate their cash flow. My findings point to the high relevance of firms' state variables to determining the effects of financial frictions on the allocation of cash flow. This can help reconcile the existence of conflicting findings in the literature that looks at the average effects of measures of firms' predisposition to suffer financial constraints on cash flow sensitivities.

The rest of the article is organized as follows. Section 3 introduces the analytical framework that guides the empirical analysis. Section 4 introduces the data sources and presents summary statistics of key variables. Section 5 describes the empirical strategy and the baseline allocation of cash flow. Section 6 explores how financial frictions affect the allocation of cash flow. Section 7 discusses additional results. Finally, section 8 concludes.

3 The analytical framework

In this section, I outline the construction of the cash flow variable and establish the relationship between the uses of cash flow through a cash flow identity. To provide a clear

understanding of each variable mentioned, their definitions are presented in Table A1. To maintain internal coherence in the definition of variables used throughout the empirical analysis, I begin with the uses and sources of funds identity:

$$CF_t + \Delta Debt_t + Equity\ issuance_t = \Delta Cash_t + Inv_t + Div_t, \tag{1}$$

where the sources of funds are the internally generated cash flow (CF_t), and the funding coming from net debt issuance $(\Delta Debt_t)$ and equity issuance ($Equity\ issuance_t$). The uses of funds are cash savings $(\Delta Cash_t)$, investment (Inv_t) and any net amount of cash distributed to shareholders through dividends or share repurchases (Div_t).

The variable CF_t captures the cash flow resulting from the company's production and sales, in addition to net financial revenue and net of corporate tax payments:

$$CF_t := Operating\ income_t + Net\ financial\ revenue_t - Corporate\ taxes_t.$$

$Debt_t$ is the sum of current and non-current liabilities, and $Cash_t$ is the sum of cash plus equivalent short-term financial assets that can be readily converted into cash. Since I do not observe Inv_t in my data, I compute it as the change in the value of assets excluding cash between periods $t-1$ and t. It is important to note that this measure might include the impact of non-cash items, which do not involve an actual cash flow. Consequently, to ensure that these non-cash items do not influence my measure of investment, I introduce an adjustment term in the definition of Inv_t as follows:

$$Inv_t := (Assets_t - Assets_{t-1}) - (Cash_t - Cash_{t-1}) + Adjustment_t, \tag{2}$$

where the *Adjustment* term is the result of adding and subtracting several components, including depreciation, gains on disposals, changes in the fair value of assets, changes in provisions, variations in the stocks of final goods or raw materials, and the proceeds of tasks performed by the company for assets.

I also lack direct measures of Div_t and $Equity\ issuance_t$. Only a measure of end-of-year equity ($Equity_t$) is observed every year t. The change in equity between years $t-1$ and t then reflects the joint effects of equity issuance, distribution of dividends, and retained earnings during year t. I construct a measure of net dividends for each year t ($Net\ Div_t$), which captures the difference between Div_t and $Equity\ issuance_t$, as the difference between Profit or Loss for the year $(P\&L_t)$ and the change in equity $(\Delta Equity_t)$:

$$Net\ Div_t := P\&L_t - \Delta Equity_t. \tag{3}$$

The rationale for this definition is as follows: If the profit or loss for the year is different from the change in equity in a period, then the difference must have been distributed to (if positive) or covered by (if negative) shareholders. With this definition, it is helpful to rewrite the identity (4) as a *cash flow identity* where the cash flow of year t equals the sum of its alternative (net) uses at the same year:

$$CF_t = \Delta Cash_t + Inv_t - \Delta Debt_t + Net\ div_t. \tag{4}$$

4 The data

4.1 Sample selection

My sample consists of balance sheet and income statements data of firms appearing in the *Central de Balances Integrada* (CBI) dataset at any point between 2003 and 2019. The CBI contains information on the quasi-universe of Spanish firms, providing an accurate representation of the Spanish economic structure. Following common practice, I focus on for-profit, not government-owned corporations that do not belong to the financial industry, industries heavily influenced by the state (Education, Health, and International organizations), or industries where firms are a minority with respect to self-employed households.

I also apply a variety of filters that exclude observations without valid and consistent information for the variables used in the analysis. After applying these filters, I keep firms for which I have at least 5 observations where all variables are well-defined. The final sample consists of 2,919,644 firm-year observations from 341,562 unique firms. A detailed explanation of the cleaning steps can be found in Appendix A.

4.2 Summary statistics of key variables

Table 1 reports summary statistics for the cash flow variable, the variables that were used to construct it, and other firm-level variables of interest for the rest of the analysis. The definition of each variable is presented in Table A1. To ease the interpretation of the summary statistics, all the variables are expressed as ratios to the average value of assets of each firm unless otherwise specified.

The numbers in Table 1 indicate that on average, every year, firms save cash by 0.89%, invest 3.58%, and distribute net dividends by -0.01%. This is financed using the internally generated cash flow, which amounts to 4.33%, and by increasing debt by 0.12%. The cash flow variable and its uses display significant variation in the sample.

The variable *Residual* measures the residual of the identity (4). The statistics associated with this variable show that the identity (4) holds very well in the data, although not perfectly. Small errors may exist because, as indicated in Appendix A, small discrepancies between the left- and right-hand sides of equation (4) may survive the cleaning process.

5 The average allocation of cash flow by Spanish firms

5.1 Empirical specification

My empirical analysis follows the approach of Chang et al. (2014). The baseline estimations rely on equation (4) and consist of regressing the different uses of cash ($\Delta Cash$, Inv, $-\Delta Debt$, and $Net\ Div$) on cash flow (CF), control variables (X), and firm (f) and year (λ) fixed effects. The regression equations are as follows:

$$\Delta Cash_{it} = \beta_1 CF_{it} + \gamma_1 X_{i,t-1} + f_{1,i} + \lambda_{1,t} + \epsilon_{1,it} \tag{5}$$

TABLE 1. Summary statistics of key variables.

Panel A: Components of cash flow, cash flow, and its uses.

			Percentiles				
Variable	**Mean**	**Sd**	**5**	**25**	**50**	**75**	**95**
Sales	1.1435	1.05	0.05	0.32	0.91	1.64	3.17
Inputs	0.7604	0.82	0.02	0.16	0.53	1.07	2.37
Personnel costs	0.3232	0.39	0	0.06	0.21	0.43	1.08
Net financial revenue	-0.0091	0.01	-0.04	-0.01	-4e-3	0	0.01
Corporate taxes	0.0075	0.02	-0.02	0	3e-3	0.01	0.04
CF	0.0433	0.11	-0.12	0.01	0.04	0.09	0.21
Δ*Debt*	0.0012	0.16	-0.24	-0.06	-0.01	0.06	0.27
Δ*Cash*	0.0089	0.10	-0.15	-0.02	1e-3	0.04	0.18
Inv	0.0358	0.17	-0.20	-0.03	0.01	0.09	0.32
Net Div	-0.0001	0.06	-0.02	-4e-18	0	1e-5	0.04
Residual	-9e-8	7e-5	-9e-5	-1e-5	0	1e-5	9e-5

Panel B: Other variables

			Percentiles				
Variable	**Mean**	**Sd**	**5**	**25**	**50**	**75**	**95**
P&L	0.0141	0.09	-0.13	-2e-3	0.01	0.05	0.14
Δ*Equity*	0.0143	0.10	-0.13	-0.01	0.01	0.04	0.15
Cash	0.1753	0.22	1e-3	0.02	0.09	0.25	0.64
Other assets	0.8247	0.29	0.31	0.65	0.85	1.00	1.26
Account receivables	0.2173	0.23	4e-4	0.04	0.15	0.33	0.67
Δ*Account receivables*	0.0032	0.12	-0.17	-0.03	3e-4	0.03	0.17
Current assets	0.5560	0.35	0.04	0.26	0.55	0.81	1.13
Non-current assets	0.4440	0.33	0.02	0.15	0.38	0.70	1.00
Current liabilities	0.4149	0.41	0.01	0.14	0.33	0.58	1.07
Non-current liabilities	0.1913	0.28	0	0	0.07	0.29	0.75
Account payables	0.3548	0.41	0.01	0.10	0.25	0.49	1.02
Δ*Account payables*	0.0028	0.14	-0.20	-0.04	6e-4	0.04	0.21
Debt	0.6062	0.46	0.04	0.28	0.55	0.84	1.31
Equity	0.3919	0.45	-0.17	0.16	0.39	0.67	0.99
Employees	0.0136	0.02	0	2e-3	0.01	0.02	0.05
Ln(Assets)	-0.6412	1.40	-2.72	-1.58	-0.73	0.18	1.70
Tangibility	0.3886	0.32	0.01	0.11	0.31	0.63	0.97
Sales growth	-0.0117	0.34	-0.49	-0.11	0.01	0.12	0.41
Age	16	9	5	10	15	21	32

Note: Variables are defined in Table A1. All the variables are measured as ratios to the average value of the assets of each firm *(Assets)*, except for *Ln(Assets)*, *Sales growth*, and *Age* which are the logarithm of *Assets*, the one-year log difference of *Sales*, and the difference between the reporting year and the year of incorporation, respectively.

$$Inv_{it} = \beta_2 CF_{it} + \gamma_2 X_{i,t-1} + f_{2,i} + \lambda_{2,t} + \epsilon_{2,it} \tag{6}$$

$$-\Delta Debt_{it} = \beta_3 CF_{it} + \gamma_3 X_{i,t-1} + f_{3,i} + \lambda_{3,t} + \epsilon_{3,it} \tag{7}$$

$$Net\ Div_{it} = \beta_4 CF_{it} + \gamma_4 X_{i,t-1} + f_{4,i} + \lambda_{4,t} + \epsilon_{4,it} \tag{8}$$

As control variables, I incorporate lagged values of cash holdings (*Cash*), capital (*Capital*), leverage (*Debt*), the logarithm of Assets (*Ln(Assets)*), the annual growth rate of sales (*Sales growth*), and a measure of asset tangibility (*Tangibility*). The inclusion of *Sales growth* as a control variable serves as a proxy for Tobin's Q, which cannot be directly utilized in this analysis due to the majority of firms not being publicly listed. All variables, except *Ln(Assets)* and *Sales growth*, are divided by the average assets of the firm. The allocation of cash flow is characterized by the coefficients $\{\beta_k\}_{k=1}^4$ in equations (5)-(8). It follows from the cash flow identity (4) that the coefficients in equations (5)-(8) have to satisfy the following conditions:

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1, \tag{9}$$

$$\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = \mathbf{0}, \tag{10}$$

$$f_{1,i} + f_{2,i} + f_{3,i} + f_{4,i} = 0 \quad \forall i, \text{ and} \tag{11}$$

$$\lambda_{1,t} + \lambda_{2,t} + \lambda_{3,t} + \lambda_{4,t} = 0 \quad \forall t. \tag{12}$$

Condition (9) implies that any change in cash flow has to be used to increase cash savings, invest, repay debt, or distribute dividends. In turn, condition (10) means that the total response of the uses of cash flow to a change in any control variable other than cash flow adds up to zero. To build intuition on this, imagine that an increase in a covariate different from cash flow is correlated with higher values of investment, debt repayment, and distribution of dividends. It follows that a lower value of cash savings is needed to satisfy the identity (4). A similar intuition applies to conditions (11) and (12).

Equations (5)-(8) can be estimated simultaneously using seemingly unrelated regressions (SUR), imposing explicitly the conditions (9)-(12). However, Greene (2012) shows that SUR estimates are equivalent to equation-by-equation ordinary least squares (OLS) estimates if the same set of explanatory variables is included in each equation. As a consequence, as long as the identity (4) holds in the data, the conditions (9)-(12) automatically hold without the need to impose them explicitly. Then, I estimate equations (5)-(8) separately by OLS without imposing the conditions (9)-(12).

5.2 Baseline results

Table 2 reports the baseline results of estimating equations (5)-(8). The coefficients of interest are those associated with *CF*, which measure *cash flow sensitivities*. These coefficients represent what portion of an additional unit of cash flow is dedicated, on average, to the corresponding use.

Table 2 documents positive cash flow sensitivities for every use of cash flow. More concretely, at the margin, an additional euro of cash flow is followed by average increases of 29 cents in cash savings, 44 cents in investment, 23 cents in debt repayment, and 4 cents in net dividends. Note that the sum of the cash flow sensitivities equals one, which means

that the condition (9) holds despite not imposing it explicitly in the estimation procedure. Additionally, Table 2 shows that the sum of the coefficients associated with the control variables adds up to zero, as stated in conditions (10).

Table 3 shows the variance decomposition of equations (5)-(8). This Table shows that the cash flow variable explains a sizable portion of the variation in the outcomes (ranging from 9 to 27 percent of the explained variation), although the firm fixed effects are the single most important determinant. Also, Figure 1 shows the evolution of the time fixed effects of equations (5)-(8). The main aggregate trends that are captured by the year fixed effects are two: First, an increase in cash savings during the entire sample period. Second, an increase in investment which is accompanied by an increase in indebtedness in the years leading to the global financial crisis.

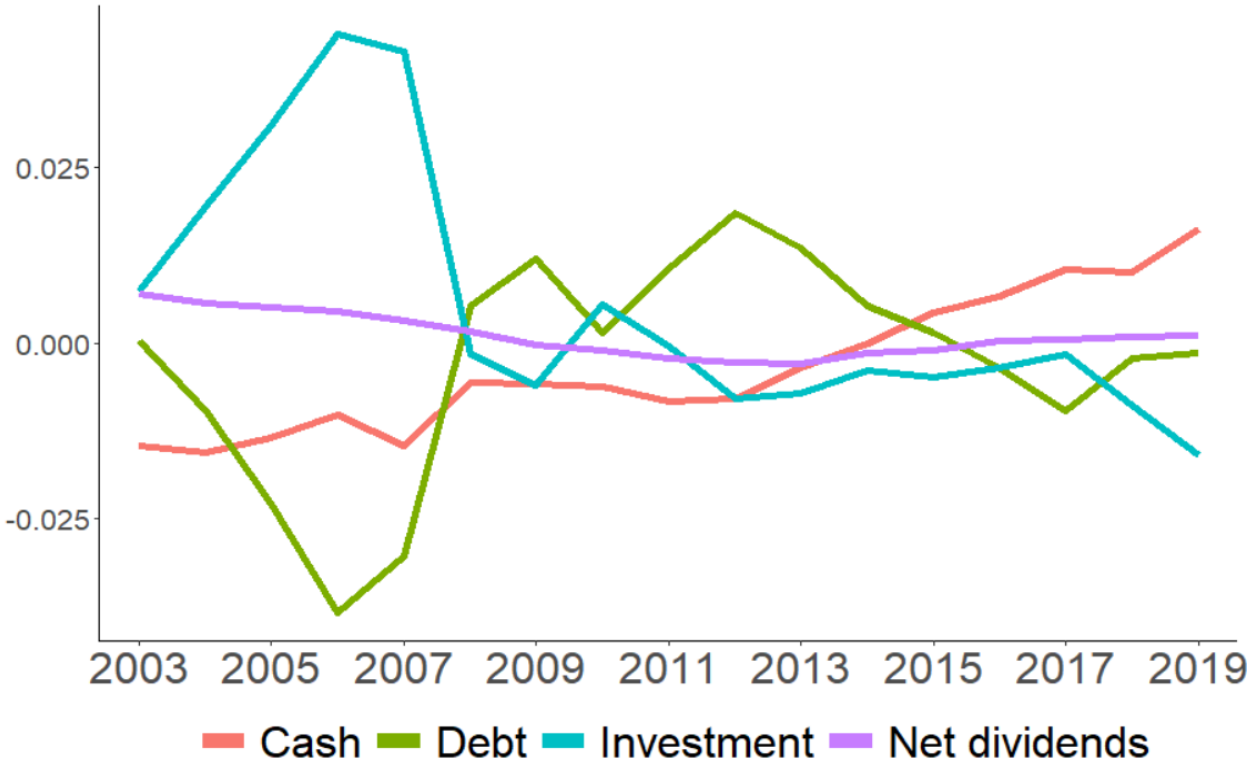
In section 7.4 I explore the asymmetries between the allocation of positive and negative realizations on cash flow.

TABLE 2. The allocation of cash flow across competing uses.

	Dependent variable:			
	$\Delta Cash$	*Inv*	-$\Delta Debt$	*Net Div*
	(1)	(2)	(3)	(4)
$CF_{i,t}$	0.29***	0.44***	0.23***	0.04***
	(1.3e-3)	(2.1e-3)	(2.3e-3)	(7e-4)
$Cash_{i,t-1}$	-0.36***	0.29***	0.06***	0.01***
	(1.3e-3)	(1.4e-3)	(1.3e-3)	(4e-4)
$Capital_{i,t-1}$	-0.06***	0.09***	-0.03***	-2.9e-3***
	(6e-4)	(1.3e-3)	(1.2e-3)	(4e-4)
$Debt_{i,t-1}$	-0.06***	-0.10***	0.20***	-0.05***
	(6e-4)	(1.0e-3)	(1.5e-3)	(4e-4)
$Ln(Assets)_{i,t-1}$	0.07***	-0.12***	0.02***	0.03***
	(6e-4)	(1.1e-3)	(1.2e-3)	(3e-4)
$Sales\ growth_{i,t-1}$	-1.2e-3***	-4.3e-3***	4.1e-3***	1.4e-3***
	(2e-4)	(3e-4)	(3e-4)	(1.4e-3)
$Tangibility_{i,t-1}$	-1.73e-9	3.89e-8**	-4.66e-8***	9.45e-9***
	(2.98e-9)	(1.52e-8)	(1.53e-8)	(2.8e-9)
Fixed-effects				
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.279	0.273	0.248	0.261

This Table reports the results of estimating equations (5) - (8). All variables are defined as in Table A1. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

FIGURE 1. Year fixed effects in the allocation of cash flow, 2003-2019.



This Figure plots the value of the year fixed effects in equations (5) - (8). The solid line connects the estimated fixed effects for each year.

TABLE 3. Decomposition of variance between the cash flow variable, the controls taken as a whole, firm fixed effects, and year fixed effects.

	Dependent variable:			
	$\Delta Cash$	*Inv*	$-\Delta Debt$	*Net Div*
Variable	(1)	(2)	(3)	(4)
CF	0.19	0.27	0.11	0.09
Controls	0.14	0.03	0.13	0.07
Firm FE	0.65	0.68	0.74	0.83
Year FE	0.02	0.02	0.02	0.01
R^2	0.29	0.29	0.26	0.27

The numbers associated with each variable indicate what portion of the explained sum of squares is attributed to the corresponding variable.

6 The effects of financial frictions

6.1 Proxying for financial frictions

To assess the impact of financial frictions on the allocation of cash flow, it is necessary to quantify the level of financial frictions experienced by firms. However, since this variable is not directly observable, researchers often resort to using observable firm characteristics to construct an index that serves as a proxy for the potential relevance of financial frictions to affect firms' behavior. These indexes are designed in a manner where higher values indicate tighter financial frictions, for instance borrowing limits that are more likely to bind, wider interest rate spreads, or financing conditions that more quickly deteriorate with funding needs.

Among the available indexes, I use the *Size-Age* (*SA*) index proposed by Hadlock and Pierce (2010):

$$SA_{it} = -0.737 * Size + 0.043 * Size^2 - 0.04 * Age, \tag{13}$$

where *Size* is the logarithm of Assets, which are measured in millions of inflation-adjusted year 2004 dollars following the authors. To develop this index, Hadlock and Pierce (2010) first used qualitative information to categorize firms' financial constraint status. Then, they estimated a set of ordered logit models in which a firm's categorized level of constraints was modeled as a function of various quantitative factors that the literature identifies as indicators of constraints. The *SA* index was the best-performing model in terms of predictive power among the ones that they assessed.

Hadlock and Pierce (2010) also cast doubts on the validity and convenience of using other popular indexes such as the ones proposed by Kaplan and Zingales (1997) and Whited and Wu (2006), and they make a compelling case that size and age are prominent predictors of the severity of financial frictions. Additionally, an index based on size and age offers the advantage of relying on variables that are arguably more pre-determined than other characteristics used in the construction of other indexes, such as leverage or dividend distribution.

However, using the *SA* index to proxy financial frictions is subject to two potential shortcomings: First, for large enough values of *Size*, the index turns increasing (rather than decreasing) in *Size*. Concretely, the point beyond which the quadratic relationship in *Size* is increasing lies approximately at 8.56. Second, Hadlock and Pierce (2010) constructed their index using COMPUSTAT data for the years 1971-2004 and it is fair to wonder whether this indicator remains valid for a sample of Spanish firms covering the years 2003-2019. I address these two shortcomings as follows.

Table 4 and Figure 2 address the first concern. Table 4 shows summary statistics of *Size* and *Age* in my sample. The first row shows that the maximum value of *Size* in my sample is 7.54, which is below the turning point at which the quadratic relationship in *Size* becomes increasing. Additionally, Figure 2 shows the density of observations by *Size* and *Age*, as well as the binscatter plot of the unconditional mean of the *SA* index in bins that contains 5% of the observations each. This Figure confirms that the *SA* index is

negatively related to *Size* and *Age*.[2]

Regarding the second potential shortcoming, I acknowledge that the ideal index of financial frictions to perform my tests should be constructed using data that is representative of Spanish firms during the period comprised in this article. Unfortunately, there is no qualitative data on financial frictions to replicate the analysis of Hadlock and Pierce (2010) using my sample. For this reason, I borrow the index from Hadlock and Pierce (2010) and I perform a robustness test to ensure the validity of the *SA* index as a measure of financial frictions. Additionally, using the index proposed by Hadlock and Pierce (2010) enhances the comparability of my results with those obtained by previous papers in the literature.

The data in Table 4 shows that in my sample firms are, on average, smaller in terms of *Size* but similar in terms of *Age* to the firms in the sample used by Hadlock and Pierce (2010) to construct the *SA* index. To alleviate potential concerns originated by these dissimilarities, I explore how the *SA* index relates to various firm-level characteristics that are considered to be related to financial frictions by the literature. Table 5 reports the results of regressions of such characteristics, which are listed by row, on the *SA* index and sector and year fixed effects. Column (1) of Table 5 shows that the results obtained using the entire sample period, while the results in columns (2) to (4) of Table 5 explore the stability of the coefficients in column (1) across different subsamples of the data.

The findings presented in Table 5 support the validity of the *SA* index as a measure of potential financial frictions. The coefficients in Column (1) demonstrate that the *SA* index is negatively correlated with cash flow, real investment, dividend distribution, and sales growth. Conversely, it is positively correlated with cash holdings and leverage. [3]

One plausible interpretation of these results is that firms with higher values of the *SA* index are more likely to be distressed firms facing limited profitability and investment opportunities, and having already exhausted their external funding sources. Consequently, these firms are more inclined to reduce their investment activities and dividend distribution.

6.2 Empirical specification and results

To estimate the effects of financial frictions on the allocation of marginal cash flow, I extend the baseline specification with the interaction of the cash flow variable and the *SA* index:

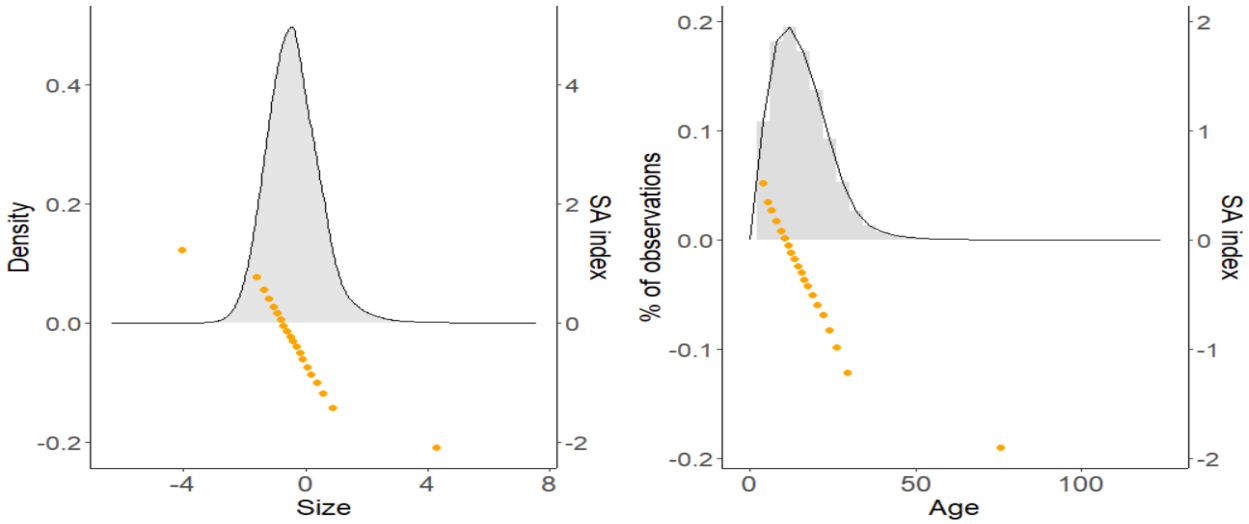
$$y_{it} = \beta_{y,1} CF_{it} + \beta_{y,2} SA_{it} + \beta_{y,3} CF_{it} \times SA_{it} + \gamma_y X_{i,t-1} + f_{y,i} + \lambda_{y,t} + \epsilon_{y,it}, \tag{14}$$

where y_{it} is any of the outcome variables in (5)-(8). The main coefficient of interest is $\beta_{y,3}$, which measures how the cash flow sensitivity of y changes with changes in the *SA* index.

[2] Figure 2 highlights the existence of outlier observations in terms of *Size* and *Age*. In unreported results, I verify that the findings in this article are robust to trimming the sample at the 99.9% percentile of *Age*, which corresponds to 74 years, and at the 0.1% and 99.9% percentiles of *Size*, which correspond to -2.48 and 2.52.

[3] In unreported results, I replicate this analysis using the index proposed by Whited and Wu (2006) and I find qualitatively similar results.

FIGURE 2. Distribution of observations and binscatter plot of the *SA* index by *Size* and *Age*.



This Figure plots the distribution of the firm-year observations by *Size* (left graph) and *Age* (right graph). Additionally, the orange dots in each graph depict the average *SA* index in a bin containing exactly 5% of the observations.

TABLE 4. Summary statistics of the financial frictions index (*SA* index) and its components.

				Percentiles				
		Mean	**Sd**	**5**	**25**	**50**	**75**	**95**
	SA	-0.31	0.85	-1.76	-0.82	-0.25	0.26	0.99
Own sample	***Size***	-0.41	0.89	-1.74	-1	-0.46	0.11	1.07
	Age	16	9	5	10	15	21	32
Hadlock & Pierce (2010)	***Size***	6.77	-	-	-	5.11	-	-
	Age	15	-	-	-	9	-	-

This Table reports summary statistics of the *SA* index, as well as of the *Size* and *Age* variables in my sample and in the sample used by Hadlock and Pierce (2010) to construct the *SA* index. *Age* is defined as specified in Table A1. The variable *Size* is computed as the natural logarithm of the firms' assets, denoted in millions of inflation-adjusted year 2004 dollars. This adjustment ensures that the measurement unit of the *Size* variable is similar to the one employed by Hadlock and Pierce (2010). The statistics associated with the sample in Hadlock and Pierce (2010) come from Columns (2) and (5) of Table 2 in their article.

TABLE 5. Relationship between the financial frictions index (*SA* index) and firm characteristics.

	Explanatory variable: $SA_{i,t}$			
	Sample period:			
	2003-2019	**2003-2007**	**2008-2012**	**2013-2019**
Dependent variable	(1)	(2)	(3)	(4)
$Sales\ growth_{i,t}$	-0.005***	-0.001**	-0.006***	-0.006***
	(2e-4)	(4e-4)	(5e-4)	(3e-4)
$Real\ investment_{i,t}$	-0.003***	3e-4	-0.002***	-0.004***
	(1e-4)	(2e-4)	(2e-4)	(2e-4)
$Cash\ flow_{i,t}$	-0.003***	0.003***	-0.008***	-0.003***
	(2e-4)	(2e-4)	(3e-4)	(2e-4)
$Dividend\ payer_{i,t}$	-0.056***	-0.056***	-0.054***	-0.056***
	(5e-4)	(9e-4)	(8e-4)	(6e-4)
$Cash_{i,t}$	0.004**	0.007***	0.002***	0.004***
	(4e-4)	(5e-4)	(5e-4)	(5e-4)
$Debt_{i,t}$	0.083***	0.037***	0.082***	0.105***
	(1.0e-3)	(1.1e-4)	(1.3e-4)	(1.4e-4)

This Table reports the OLS coefficients of SA_{it} in regressions where the dependent variables are listed by row. All variables are defined in Table A1. The regressions also include year and 4 digits industry fixed effect. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

The results of estimating specification (14) are presented in Table 6. The coefficients associated with $CF_t \times SA_{it}$ indicate that higher values of the *SA* index are, on average, associated with an increase in the cash flow sensitivity of debt repayment at the expense of a reduction in the cash flow sensitivity of the rest of the uses.[4] More concretely, a unit increase in the *SA* index (whose standard deviation is 0.85) is associated with an average increase of 6 cents in the share of a marginal unit of cash flow allocated towards debt repayment and average decreases of 2 cents, 3 cents, and 1 cent in the shares allocated to cash savings, investment, and net dividends, respectively.[5]

These results are consistent with theories that emphasize the value that firms attribute to financial flexibility, which takes the form of unused debt capacity in this case. Along these lines, one interpretation of my results is that firms characterized by higher values of the *SA* index tend to increase their allocation of marginal cash flow towards debt repayment to preserve financial flexibility to fund future investment opportunities (Dasgupta and Sengupta, 2007). Also, these results are in line with the predictions of the model in DeAngelo et al. (2011), which emphasizes that firms' ex-ante financial policies aim to preserve the ability to access the capital markets ex-post in the event of unexpected earnings

[4] In unreported results, I replicate this analysis using the index proposed by Whited and Wu (2006) and I find qualitatively similar results.

[5] Table B1 in the Appendix replicates the results presented in Table 6 using the index proposed by Whited and Wu (2006). Although there are differences in the magnitudes of the estimated coefficients, the qualitative results are similar.

shortfalls or investment opportunities.

TABLE 6. The impact of financial frictions on the allocation of cash flow.

	Dependent variable:			
	$\Delta Cash_{i,t}$	$Inv_{i,t}$	$-\Delta Debt_{i,t}$	*Net Div*
	(1)	(2)	(3)	(4)
$CF_{i,t}$	0.24***	0.33***	0.38***	0.05***
	(0.001)	(0.002)	(0.002)	(0.001)
$SA_{i,t}$	-0.23***	-0.54***	0.75***	0.02***
	(0.001)	(0.002)	(0.002)	(4e-4)
$CF_{i,t} \times SA_{i,t}$	-0.02***	-0.03***	0.06***	-0.01***
	(0.001)	(0.002)	(0.003)	(0.001)
Fixed-effects				
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.332	0.389	0.506	0.255

This Table reports the results of estimating specification (14) for each use of cash flow. All variables are defined in Table A1. The specifications include the same controls as in equations (5) - (8) except for *Ln(Assets)*, which has been converted to a discrete variable indicating that *Ln(Assets)* is in the bottom (top) half of its distribution when it equals 0 (1). For brevity, the results related to the controls are omitted from the Table. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

6.3 Interaction with state variables

The previous subsection shows that higher values of the *SA* index are, on average, associated with an increase in the allocation of marginal cash flow toward debt repayment. This subsection explores if such insight can be generalized to the population of firms or if, alternatively, the effects of changes in the *SA* index on the allocation of marginal cash flow are contingent upon the specific economic and financial situation of each firm.

I explore heterogeneity across a list of variables that are typically treated as state variables in structural dynamic models of firms' behavior (e.g., Gao, Whited, and Zhang, 2021): leverage (*Debt*), cash holdings (*Cash*), capital stock (*Capital*), and availability of investment opportunities (*Sales growth*).

Using each of the above variables at a time, I divide the sample into deciles according to the distribution of the corresponding variable. Then, I estimate specification (14) for each subsample separately and I compare the coefficient associated to $CF_{it} \times SA_{it}$ across subsamples.[6]

The results are presented in Figure 3. The Figure is divided into four panels, one for each state variable that is used to split the sample. In turn, each panel contains four

[6]In these specifications, the variable that is used to split the sample is not included in the list of controls.

columns, one for each use of cash flow. By comparing the estimated coefficients across subsamples (within a panel) one obtains insights on how the effects of changes in the *SA* index on the allocation of marginal cash flow differ across firms with different values of the chosen state variable.

I highlight the following results: First, looking at the heterogeneity across *Debt* (Panel A) we observe that changes in the *SA* index have either small or insignificant effects on the cash flow sensitivities across the first 7 deciles of the distribution of *Debt*. However, higher values of the *SA* index are correlated with a significantly smaller (larger) cash flow sensitivity of investment (debt repayment) in the last three deciles of the distribution of *Debt*.

Second, looking at Panel B we observe that the impact of changes in the *SA* index on the cash flow sensitivities of investment and debt repayment is larger among firms in the lowest deciles of the distribution of *Cash*. In contrast, among firms in the highest deciles of the distribution of *Cash*, the cash flow sensitivities of investment and debt repayment are not significantly affected by changes in the *SA* index.

Third, looking at Panel C we observe that the cash flow sensitivity of investment does not react to changes in the *SA* index for firms in the first 8 deciles of the distribution of *Capital*. For these firms, changes in the *SA* index are correlated with a decrease (increase) in the cash flow sensitivities of cash savings and net dividends (debt repayment). However, for firms in the last two deciles of the distribution of *Capital*, the cash flow sensitivity of investment also decreases with the *SA* index.

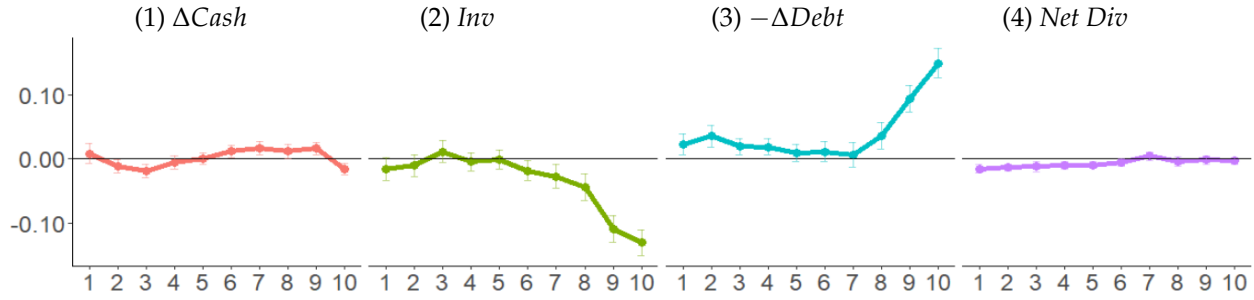
Finally, it is harder to observe a clear pattern when looking at the heterogeneity across *Sales growth* (Panel D). One aspect to highlight is that, among firms in the last deciles of the distribution of *Sales growth*, higher values of the *SA* index are correlated with decreases in the cash flow sensitivity of investment and cash savings, and by increases in the cash flow sensitivity of debt repayment.

The discussion of the results shed light on a potential way to reconcile the apparently conflicting findings of the literature. One prominent example of the debates that have been sustained in the literature is the **contradiction** between the results of Almeida et al. (2004, 2021) and Riddick and Whited (2009). On one hand, Almeida et al. (2004, 2021) show that more financially constrained firms display a larger cash flow sensitivity of cash. They claim that this is evidence of an optimal cash savings policy to balance the profitability of current and future investments. On the other hand, Riddick and Whited (2009) shows that more financially constrained firms display a lower (and even negative) cash flow sensitivity of cash. They argue that this is observed because firms lower cash reserves to invest after receiving positive cash-flow shocks, and vice versa.

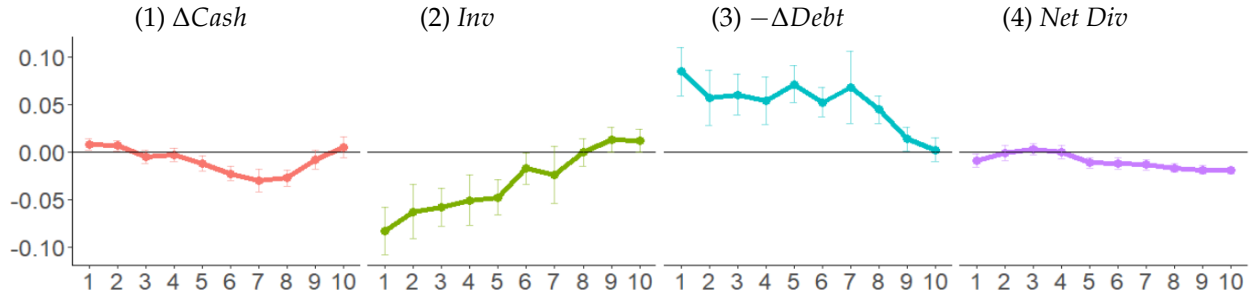
Figure 3 shows that firms will behave in one or another of the ways predicted by Almeida et al. (2004, 2021) and Riddick and Whited (2009) depending on their specific economic and financial situation. For instance, the results in Panel B show that firms in the lowest deciles of the distribution of *Cash* display a larger cash flow sensitivity of cash, which is consistent with the predictions of Almeida et al. (2004, 2021). However, firms in the highest deciles of the distribution of *Cash* display a lower (or null) cash flow sensitivity of cash, which is consistent with the predictions of Riddick and Whited (2009).

FIGURE 3. The impact of financial frictions on the allocation of cash flow. Heterogeneous effects across variables describing the economic and financial situation of firms.

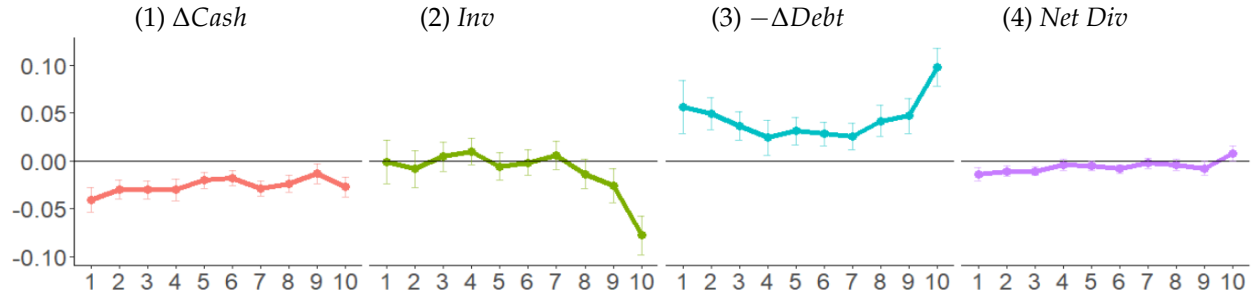
Panel A: Sample split using deciles of *Debt*.



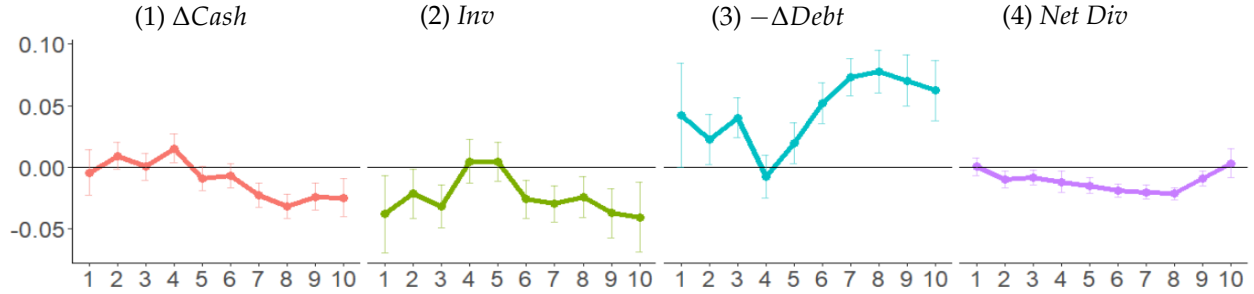
Panel B: Sample split using deciles of *Cash*.



Panel C: Sample split using deciles of *Capital*.



Panel D: Sample split using deciles of *Sales growth*.



This Figure explores the existence of heterogeneity in the effects of financial frictions on the allocation of cash flow across a list of variables: leverage (*Debt*), cash holdings (*Cash*), capital stock (*Capital*), and availability of investment opportunities (*Sales growth*). Using each of these variables at a time, I divide the sample into deciles according to the distribution of the corresponding variable. Then, I estimate specification (14) for each subsample separately and report the coefficients associated with $CF_{it} \times SA_{it}$. In these specifications, the variable that is used to split the sample is not included in the list of controls. The Figure is divided into four panels, one for each state variable that is used to split the sample. In turn, each panel contains four columns, one for each use of cash flow. All variables are defined as in Table A1. The bars centered around each coefficient represent 95% confidence intervals.

7 Additional analysis

7.1 Persistent and transitory components of cash flow

This subsection further characterizes the allocation of cash flow by extending the specifications in equations (5) - (8). In this analysis, the cash flow variable is replaced with its persistent and transitory components, which are obtained by applying the Hodrick-Prescott filter to the unscaled cash flow variable and by dividing the resulting cycle and trend components by average assets. This exercise seeks to explore whether firms allocate these different components of cash flow in distinct ways. Concretely, the specifications that are estimated are the following:

$$y_{it} = \beta_{y,1} Transitory_{it} + \beta_{y,2} Persistent_{it} + \gamma_y X_{i,t-1} + f_{y,i} + \lambda_{y,t} + \epsilon_{y,it}, \tag{15}$$

where the coefficients $\beta_{y,1}$ and $\beta_{y,2}$ measure what portion of an additional unit of the transitory and persistent components are allocated to the use y_{it}. $X_{i,t-1}$ includes the same controls used in equations (5) - (8).

The results are shown in Table 7. First, Table 7 shows that a constraint equivalent to (9) holds for both the transitory and persistent components of cash flow. Second, notice that the estimates of $\beta_{y,1}$ in the first row of Table 7 are close to the cash flow sensitivities presented in Table 2. Third, when comparing the coefficients that characterize the marginal allocation of the transitory and persistent components we observe that a larger proportion of an additional unit of the persistent component is devoted to investment, compared to how an additional unit of the transitory component is allocated. This finding is in line with existing evidence (for instance, Chang et al. (2014) and Byun et al. (2019) find similar results) and it is consistent with the conjecture that the persistent component of cash flow contains more information about the future profitability of the firms than the transitory component. In unreported results, I replicated this analysis using a balanced panel of observations, and the results remained qualitatively similar.

7.2 Disagregation of investment

The variable *Inv* that is used in the baseline characterization of the allocation of cash flow can be disgregated into three separate components: real investment (*Real investment*), financial investment (*Financial investment*), and provision of trade credit to customers (Δ *Account receivables*). In this subsection, I study how much of the cash flow sensitivity of investment (*Inv*) can be attributed to each of its components.

I reproduce the baseline estimation presented in Table 2 but decompose the *Inv* variable into its three components. The results are shown in Table 8, where we observe that in response to a unit increase in cash flow, real investment increases by 23 cents, financial investment by 1 cent, and trade credit provision by 19 cents. The sum of these three coefficients is equal to the cash flow sensitivity of investment (*Inv*) in Table 2.

These results underscore that almost all of the cash flow sensitivity of investment is attributable to the cash flow sensitivities of real investment and trade credit provision. The forthcoming evidence in the following subsection further illuminates the potential

TABLE 7. The allocation of the transitory and persistent components of cash flow.

	Dependent variable:			
	ΔCash	*Inv*	-ΔDebt	*Net Div*
	(1)	(2)	(3)	(4)
$Transitory_{i,t}$	0.27^{***}	0.43^{***}	0.27^{***}	0.04^{***}
	(0.002)	(0.003)	(0.003)	(8e-4)
$Persistent_{i,t}$	0.31^{***}	0.45^{***}	0.18^{***}	0.06^{***}
	(0.002)	(0.003)	(0.004)	(0.001)
Fixed-effects				
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.278	0.274	0.247	0.261

This Table reports the results of estimating the specification (15) for each use of cash flow. The *Transitory* and *Persistent* variables are obtained by applying the Hodrick-Prescott filter to the unscaled cash flow variable and by dividing the resulting cycle and trend components by average assets. The specifications include the same controls as in equations (5) - (8) but, for brevity, the results related to the controls are omitted from the Table. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

reasons behind firms allocating such a substantial share of their cash flow to trade credit provision. One plausible interpretation is that firms utilize trade credit provision as a temporary allocation of their cash flow, which is subsequently redirected towards other uses in the future.

7.3 Allowing for richer dynamics in the allocation of cash flow

Firms may decide to stage their investment after a change in cash flows. For instance, a firm that receives a large inflow of unexpected income may react by saving it first and using it for investment in subsequent periods. To examine the presence of such delays in investment decisions, I expand upon the specifications outlined in equations (5)-(8) by incorporating two lags of the cash flow variable.

The outcomes of this analysis are detailed in Table 9. When focusing on the coefficients of the *Real investment* regression, it becomes clear that while there is a significant relationship between contemporaneous *Real investment* and the lagged components of cash flow, this relationship is relatively modest in magnitude compared to the contemporaneous relationship between *Real investment* and *CF*. As a result, it is fair to say that while there exists some evidence of delays in the allocation of cash flow towards real investment, the bulk of the cash flow sensitivity of real investment is contemporaneous.

The results in column (4) uncover another interesting pattern. The firms exhibit a large contemporaneous cash flow sensitivity of trade credit provision, allocating 22 cents out of an additional euro of cash flow to this purpose. However, what makes this pattern particularly interesting is that firms seem to reallocate part of these funds in subsequent periods, particularly towards debt repayment and real investment. The analysis presented

TABLE 8. The allocation of cash flow across the three components of investment (*Inv*): real investment, financial investment, and provision of trade credit to customers.

	Dependent variable:		
	Real investment	*Financial investment*	Δ *Account receivables*
	(1)	(2)	(3)
CF_t	0.23***	0.01***	0.19***
	(1.6e-4)	(4e-4)	(1.5e-4)
$Cash_{t-1}$	0.07***	0.04***	0.19***
	(9e-4)	(4e-4)	(1.0e-4)
$Capital_{t-1}$	-0.10***	0.02***	0.16***
	(1.0e-4)	(4e-4)	(9e-4)
$Debt_{t-1}$	-0.06***	1e-4	-0.04***
	(7e-4)	(2e-4)	(6e-4)
$Ln(Assets)_{t-1}$	0.03***	-0.02***	0.13***
	(7e-4)	(3e-4)	(8e-4)
$Sales\ growth_{t-1}$	0.01***	8e-4***	-0.01***
	(2e-4)	(9.67e-5)	(2e-4)
$Tangibility_{t-1}$	3.42e-8**	3.82e-9	9.08e-10
	(1.13e-8)	(2.88e-9)	(2.78e-9)
Fixed-effects			
Firm	Yes	Yes	Yes
Year	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644
R^2	0.252	0.11	40.165

This Table reports the results of estimating a specification similar to equation (6) where the outcome variables are the three components of investment (*Inv*): real investment, financial investment, and provision of trade credit to customers. All variables are defined as in Table A1. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

in Chang et al. (2014) also reveals a similar pattern, although their findings suggest that firms employ cash holdings, rather than trade credit provision, as a temporary allocation of cash flow that is subsequently utilized to boost investment in the future. [7]

TABLE 9. Dynamic effects in the allocation of cash flow across competing uses.

	Dependent variable:					
	$\Delta Cash$	*Real investment*	*Financial investment*	Δ *Account receivables*	-$\Delta Debt$	*Net Div*
	(1)	(2)	(3)	(4)	(5)	(6)
CF_t	0.29***	0.23***	0.01***	0.22***	0.22***	0.04***
	(0.001)	(0.002)	(4e-4)	(0.002)	(0.002)	(7e-4)
CF_{t-1}	5e-5	0.03***	2.8e-3***	-0.11***	0.07***	0.01***
	(0.001)	(0.001)	(4e-4)	(0.001)	(0.002)	(6e-4)
CF_{t-2}	0.02***	0.01***	1e-4	-0.03***	4.6e-3***	-3.8e-3***
	(0.001)	(0.001)	(4e-4)	(0.001)	(0.002)	(6e-4)
Fixed-effects						
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.279	0.252	0.114	0.171	0.249	0.261

This Table reports the results of estimating an extension of the specifications in equations (5) - (8) where I include two lags of the cash flow variable. All variables are defined as in Table A1. The specifications include the same controls as in equations (5) - (8) but, for brevity, the results related to the controls are omitted from the Table. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

7.4 Asymmetries across positive and negative cash flow

I explore the existence of asymmetric reactions to positive and negative changes in cash flow by extending the specifications in equations (5)-(8) with a dummy variable $NEG_{i,t}$ that equals 1 when the cash flow is negative, and with the interaction of this dummy and the cash flow variable.

To begin with, Table B2 in the Appendix presents summary statistics about the $NEG_{i,t}$ variable, and about negative $(CF_{i,t} \times NEG_{i,t})$ and positive $(CF_{i,t} \times (1 - NEG_{i,t}))$ realizations of cash flow separately. We observe that the averages of negative and positive cash flow are similar in absolute value, and the dispersion of negative and positive cash flow

[7] In unreported results, I explore an extension of the analysis presented in Table 9 where I interacte the cash flow variable and its lags with the *SA* index. Among other interesting results, I find that the relationship between *Real investment* and the lagged components of cash flow is higher the higher the *SA* index. This is consistent with the discussion of the results in Section 6.2, where I argued that one potential interpretation of my results is that firms characterized by higher values of the *SA* index tend to increase their allocation of marginal cash flow towards debt repayment to preserve financial flexibility to fund future investment opportunities.

is also similar. However, the first line of Table B2 highlights that the positive realizations of cash flow are much more frequent than the negative ones.

Next, Table 10 presents the results of extending the baseline allocation by adding the $NEG_{i,t}$ variable and its interaction with the cash flow variable. The coefficients in the first line are cash flow sensitivities when the marginal unit of cash flow is positive, and the coefficients in the third line measure how the cash flow sensitivities differ when the cash flow is negative instead of positive.

The results indicate that the cash flow sensitivity of *Net Div* is 4 cents regardless of the sign of the cash flow, which means that *Net Div* is adjusted symmetrically for positive and negative cash flow realizations. Turning the attention to the cash flow sensitivity of cash, the results indicate that an additional unit of positive cash flow is followed by an average increase of 45 cents in cash savings, while an additional unit of negative cash flow is followed by an average decrease of 10 (45 - 35) cents in cash savings. Similarly, an additional unit of positive cash flow is followed by an average increase of 49 cents in investment, while an additional unit of negative cash flow is followed by an average decrease of 39 (49 - 10) cents in investment. Finally, we observe that the cash flow sensitivity of debt repayment is only 2 cents when the cash flow is positive, while it is 47 (2 + 45) cents when the cash flow is negative.

In summary, the main uses to which a marginal unit of positive cash flow is allocated are cash savings (45 cents) and investment (49 cents), while the main sources that firms use to accommodate an additional unit of negative cash flow are increases in debt (47 cents) and reductions in investment (39 cents).

8 Conclusion

This paper examines the impact of financial frictions on the allocation of cash flow by Spanish non-financial corporations. My findings suggest that larger financial frictions, as proxied by the *Size-Age* (*SA*) index proposed by Hadlock and Pierce (2010), are associated, on average, with a higher proportion of cash flow allocated to debt repayment and lower proportions allocated to cash savings, investment, and dividend distribution. Importantly, the analysis highlights that the effect of the proxy for financial frictions on the allocation of cash flow varies significantly with variables capturing the economic and financial situation of each firm.

The empirical analysis is built on a cash flow identity (which is derived from a uses and sources of funds identity) that equates cash flow to the sum of all its alternative uses. The empirical analysis uses administrative data from a comprehensive sample of Spanish non-financial corporations over the period from 2003 to 2019. By regressing the different uses of cash flow on cash flow itself, control variables, and firm and year fixed effects, I obtain empirical estimates of cash flow sensitivities, which quantify the marginal allocation of cash flow to each use.

The baseline characterization of the allocation of cash flow indicates that an additional euro of cash flow is associated, on average, with increases in cash savings, investment, debt repayment, and dividend distribution. Next, I extend the analysis to explore how changes in the proxy for financial frictions affect the allocation of cash flow. I find that

TABLE 10. Allowing for asymmetry in the allocation of positive and negative cash flow across competing uses.

	Dependent variable:			
	$\Delta Cash$	*Inv*	$-\Delta Debt$	*Net Div*
	(1)	(2)	(3)	(4)
$CF_{i,t}$	0.45***	0.49***	0.02***	0.04***
	(0.002)	(0.003)	(0.003)	(9e-4)
$NEG_{i,t}$	-4e-4	1.7e-3***	-2.0e-3**	1e-3***
	(2e-4)	(4e-4)	(5e-4)	(2e-4)
$CF_{i,t} \times NEG_{i,t}$	-0.35***	-0.10***	0.45***	-2e-4
	(0.003)	(0.005)	(0.005)	(0.002)
Fixed-effects				
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.289	0.274	0.255	0.261

This Table reports the results of estimating an extension of the specifications in equations (5) - (8) where I interact the cash flow variable with the indicator $NEG_{i,t}$. All variables are defined as in Table A1, except for the $NEG_{i,t}$ dummy, which is equal to 1 when the cash flow variable is negative. The specifications include the same controls as in equations (5) - (8) but, for brevity, the results related to the controls are omitted from the Table. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

higher values of the proxy for financial frictions are associated, on average, with a higher cash flow sensitivity of debt repayment and lower cash flow sensitivities of cash savings, investment, and dividend distribution. Furthermore, I document that the effect of the proxy for financial frictions on the allocation of cash flow varies significantly with pre-determined variables that characterize the economic and financial situation of each firm, such as leverage, cash holdings, capital, and the availability of investment opportunities.

The results discussed in this article are relevant for three main reasons. First, they contribute to the understanding of the relationship between financial frictions and the allocation of cash flow. The results suggest that the effects of financial frictions on cash flow allocation are contingent upon the specific economic and financial situation of each firm (as captured by standard pre-determined state variables). Uncontrolled-for variability in those and other relevant state variables across samples might explain the apparently conflicting findings of the literature that has tried to establish if and how financial constraints affect the allocation of cash flow with reduced form methodologies. Second, they also highlight the value that more structural approaches (able to simultaneously account for the dynamics of the relevant state variables) may have in this field. A structural model can provide precise predictions on the sign and magnitude of the cash flow sensitivities contingent on the state of the firm. Such predictions can be validated with the empirical methodology proposed in this paper. Finally, the findings bear direct relevance to policymakers. The analysis of heterogeneous effects of financial frictions offers valuable insights to enhance the precision of policies, such as direct transfers or subsidized credit guarantees, targeting specific firms for purposes such as promoting investment. These results contribute to anticipating the response of certain firms when provided with additional cash flow (in the case of a direct transfer) or increased borrowing capacity (in the case of a subsidized credit guarantee).

References

ALLAYANNIS, G. AND A. MOZUMDAR (2004): "The impact of negative cash flow and influential observations on investment–cash flow sensitivity estimates," *Journal of banking & finance*, 28, 901–930.

ALMEIDA, H., M. CAMPELLO, AND M. S. WEISBACH (2004): "The cash flow sensitivity of cash," *The journal of finance*, 59, 1777–1804.

——— (2021): "The cash flow sensitivity of cash: Replication, extension, and robustness," *Fisher College of Business Working Paper*, 002.

ALMUNIA, M., D. LOPEZ RODRIGUEZ, AND E. MORAL-BENITO (2018): "Evaluating the macro-representativeness of a firm-level database: an application for the Spanish economy," *Banco de Espana Ocassional Paper*.

BYUN, S. K., V. POLKOVNICHENKO, AND M. J. REBELLO (2019): "Composition of cash flow shocks, firm investment, and cash holdings," *Firm Investment, and Cash Holdings (November 15, 2019)*.

CHANG, X., S. DASGUPTA, G. WONG, AND J. YAO (2014): "Cash-flow sensitivities and the allocation of internal cash flow," *The Review of Financial Studies*, 27, 3628–3657.

CLEARY, S. (1999): "The relationship between firm investment and financial status," *The journal of finance*, 54, 673–692.

DASGUPTA, S. AND K. SENGUPTA (2007): "Corporate liquidity, investment and financial constraints: Implications from a multi-period model," *Journal of Financial Intermediation*, 16, 151–174.

DEANGELO, H., L. DEANGELO, AND T. M. WHITED (2011): "Capital structure dynamics and transitory debt," *Journal of financial economics*, 99, 235–261.

DENIS, D. J. AND S. B. MCKEON (2012): "Debt financing and financial flexibility evidence from proactive leverage increases," *The Review of Financial Studies*, 25, 1897–1929.

DROBETZ, W., R. HALLER, I. MEIER, AND V. TARHAN (2017): "The impact of liquidity crises on cash flow sensitivities," *The Quarterly Review of Economics and Finance*, 66, 225–239.

FAZZARI, S., R. G. HUBBARD, AND B. C. PETERSEN (1987): "Financing constraints and corporate investment," *Brookings Papers on Economic Activity, Mass., USA*, 1998, 141–206.

FAZZARI, S. M., R. G. HUBBARD, AND B. C. PETERSEN (2000): "Investment-cash flow sensitivities are useful: A comment on Kaplan and Zingales," *The Quarterly Journal of Economics*, 115, 695–705.

GAO, X., T. M. WHITED, AND N. ZHANG (2021): "Corporate money demand," *The Review of Financial Studies*, 34, 1834–1866.

GATCHEV, V. A., T. PULVINO, AND V. TARHAN (2010): "The interdependent and intertemporal nature of financial decisions: An application to cash flow sensitivities," *The Journal of Finance*, 65, 725–763.

GREENE, W. H. (2012): *Econometric analysis*, Upper Saddle River: Prentice-Hall.

HADLOCK, C. J. AND J. R. PIERCE (2010): "New evidence on measuring financial constraints: Moving beyond the KZ index," *The review of financial studies*, 23, 1909–1940.

KAPLAN, S. N. AND L. ZINGALES (1997): "Do investment-cash flow sensitivities provide useful measures of financing constraints?" *The quarterly journal of economics*, 112, 169–215.

RIDDICK, L. A. AND T. M. WHITED (2009): "The corporate propensity to save," *The Journal of Finance*, 64, 1729–1766.

WHITED, T. M. AND G. WU (2006): "Financial constraints risk," *The review of financial studies*, 19, 531–559.

Appendix A: Data Appendix

This appendix outlines the steps undertaken in the cleaning process of the *Central de Balances Integrada* (CBI) dataset. These steps were executed to yield the sample of observations that form the basis of analysis in this study. The definitions of the variables mentioned below can be found in Table A1.

The CBI contains the balance sheet and profit and loss account of Spanish corporations since 1995. Following Almunia et al. (2018), I exclude data from 1995 to 1999 because in these years the quality of the data was relatively poor and its coverage was limited. To prevent the events triggered by the pandemic from affecting my estimates, the data included in this study only extends until 2019.

Step 1. I delete observations that do not pass the quality tests applied by the provider, which are reflected in the variable *calidad* that takes a value of 1 if the observation passes the tests and 0 otherwise. The forms received from the companies are subject to a company-by-company filtering process to guarantee the quality and consistency of the information incorporated into the database. This implies that the raw data received by the managers of the CBI is not integrated into the database until it has passed numerous tests, both logical and arithmetical, which are aimed at guaranteeing internal and external consistency. The details of this filtering process can be found in the supplementary material that appends the annual publication of the main results in the CBI dataset by the Bank of Spain. This supplementary material is only available in Spanish (see, for example, *suplemento metodológico 2020*).

Step 2. Further to the quality controls applied by the data provider, I implement three filters to focus on private, for-profit, non-financial corporations. First, I use the first letter of the tax identifier (*cif*) to exclude entities that are arguably not-for-profit enterprises. Table A2 lists what type and how many observations were deleted in this step. Second, I use the *gsec09* variable, which represents the section codes of the National Classification of Economic Activities (CNAE-2009), to delete observations for firms belonging to the financial industry (K), the public administration (O), industries heavily influenced by the government (P, Q, U), or industries where firms are a minority compared to self-employed households (T). Table A3 provides details on what industries and how many observations were deleted in this step. Third, I use the *grup* variable, which is an identifier of government or non-government ownership of firms, to delete any remaining entities controlled by the government. Panel A of Table A4 provides details on how many observations were deleted in this step.

Step 3. I remove all observations corresponding to firms that report dubious information on their *Year of incorporation*. Specifically, I exclude firms that report a *Year of incorporation* that is either negative or greater than any of the years in which the firm reports data.

Step 4. I apply a set of filters to remove observations that involve apparent reporting inconsistencies or are likely to belong to inactive or insolvent firms. First, I delete observations with negative or zero *output*, *sales*, *assets*, or *capital*. Second, I delete observations where *cash* is negative or larger than *assets*. Third, I eliminate observations that exhibit negative values for either *personnel costs* or *inputs*. Fourth, I delete observations where *long-term interest-bearing debt* is negative or larger than *non-current liabilities*. Similarly, I

TABLE A1. Definition of variables.

Panel A: Variables used to measure cash flow and its uses

Name	Definition
Sales	*Net turnover* (c200001) plus *Other operating income* (c200006).
Inputs	*Net purchases* (c200010) plus *Other operating costs.* (c200012).
Personnel Costs	*Personnel Costs* (c200025).
Operating income	*Sales - Inputs - Personnel Costs*.
Net Financial Revenue	*Net Financial Revenue* (c290042).
Corporate income taxes	*Corporate income taxes* (c200069).
Provisions	*Provisions* (c200177).
Assets	*Assets* (c200135).
Cash	*Cash and equivalents* (c200129) plus *Short-term financial investment* (c200128).
Adjustment	*Depreciation* (c200043) - *Gains on disposal* (c290059) - *Changes in fair value* (c290068) + Annual change in *Provisions* (c200177) - *Variation in stock of final goods* (c200003) - *Variation in stock of raw materials* (c200011) - *Tasks performed for asset* (c200005).
P&L	*Profit (loss) for the year* (c290070).
Equity	*Equity* (c200145).
Debt	*Non-current Liabilities* (c200158) plus *Current Liabilities* (c200180).
Residual	Difference between the left- and right-hand sides of identity (4).

Panel B: Other variables

Name	Variables definition
Year	Year associated with the information reported by the firm (any).
Year of incorporation	Year in which the firm was incorporated (anyconst).
Age	Year minus Year of incorporation.
Calidad	Indicator if the firm complies or not with quality standards (calidad).
Cif	Tax identification number associated with the firm (cif).
Gsec09	CNAE 2009 section code. It has a length of one alphanumeric position (gsec09).
grup	Identifier of government or non-government ownership of firms (grup).
Employees	Average number of employees (units) (c200084).
Output	Value of output (c200075).
Tangibility	*Tangible assets and Property* (c200098).
Capital	*Intangible assets* (c200089) plus *Tangibility* (c200098).
Sales growth	Annual change in log of *Sales*.
Current liabilities	*Current liabilities* (c200180).
Non-current liabilities	*Non-current liabilities* (c200158).
Long-term interest-bearing debt	*Long-term external funds* (c200151).
Short-term interest-bearing debt	*Short-term interest-bearing external funds* (c209166).
Account payables	*Short-term non-interest-bearing external funds* (c209179).
Account receivables	*Trade and other receivables* (c200121).
Financial investment	*Long-term financial investment* (c200103).
Real investment	*Inv* - Annual change in *Financial investment* - Annual change in *Account receivables*

This Table reports the variables in the CBI dataset that were used to construct the variables used in the empirical analysis. The names under the "Name" column refer to the names used in this paper, while the names and the codes under the "Definition" column refer to the names and the codes of the variables in the CBI dataset. The codes are reported in parentheses. In the empirical analysis, all the variables are measured as ratios to the average value of assets of each firm (*Assets*), except for *Ln(Assets)*, *Sales growth*, and *Age* which are measured as the logarithm of *Assets*, the one-year log difference of *Sales*, and the difference between the reporting year and the year of incorporation, respectively.

delete observations where *short-term interest-bearing debt* or *account payables* is negative or larger than *current liabilities*. Fifth, I delete observations for which the sum of *equity*, *debt*, and *provisions* differs substantially from the value of assets. Concretely, I delete observations for which the ratio of the sum of equity, liabilities, and provisions to assets is larger (lower) than the percentile 99.9 (0.1) in the sample prior to this deletion. Additionally, I delete observations for which the ratio of the sum of the uses of cash flow to cash flow is larger (lower) than 1.01 (0.99). [8] Finally, I delete observations where the ratio of the right- to the left-hand side of the cash flow identity (4) is bigger (smaller) than 1.01 (0.99).

Step 5. I delete observations where any of the following variables take values smaller (larger) than the 0.01 (0.99) percentile of the corresponding variable in sample prior to this deletion: *Sales*, *Inputs*, *Personnel costs*, *Net financial revenue*, *Corporate income taxes*, *CF*, *Inv*, Δ *Debt*, Δ *Cash*, *Net Div*. This last step aims to eliminate any remaining outliers that could potentially result from measurement errors in the reported data.

Step 6. Finally, I exclude firms for which the clean sample contains less than 5 observations. Panel C of Table A4 summarizes how each of the cleaning steps mentioned above affects the number of observations in my dataset.

[8] The quality filter applied by the data provider admits arithmetic errors that are not substantial in magnitude. More details on this can be found in the supplementary material (*suplemento metodológico*). In the same spirit, I do not require the accounting identities to hold exactly in the data.

TABLE A2. Description of preserved and deleted observations based on tax identifiers.

	Code	Firm-year Observations	Description	Description in Spanish
Preserved entities	A	1,236,381	Corporation	*Sociedades anónimas*
	B	12,855,214	Limited liability company	*Sociedades de responsabilidad limitada*
	C	2,395	Business partnership	*Sociedades colectivas*
	D	1,242	Limited partnership	*Sociedades comanditarias*
	J	4,976	Civil society	*Sociedades civiles*
	U	1,758	Joint venture	*Uniones Temporales de Empresas*
	N	569	Foreign entity	*Entidades extranjeras*
	W	1,336	Branch entity	*Establecimientos permanentes de entidades no residentes en territorio español*
Deleted entities	E	966	Joint ownership, inheritance in abeyance, or other entity	*Comunidades de bienes, herencias yacentes y demás entidades carentes de personalidad jurídica no incluidas expresamente en otras claves*
	F	46,490	Cooperative society	*Sociedades cooperativas*
	G	14,070	Association	*Asociaciones*
	H	88	Residents' association under a horizontal property regime	*Comunidades de propietarios en régimen de propiedad horizontal*
	P	8	Local corporation	*Corporaciones Locales*
	Q	1,079	Public institution	*Organismos públicos*
	R	257	Religious institution	*Congregaciones e instituciones religiosas*
	S	13	State or Autonomous Community Institution	*Órganos de la Administración del Estado y de las Comunidades Autónomas*
	V	11,178	Other type undefined in another code	*Otros tipos no definidos en el resto de claves*
	Invalid	961	Observations having a cif with an invalid structure	

TABLE A3. Description of preserved and deleted industries.

	Code	Firm-year observations	Description
Preserved industries	A	351,080	Agriculture, livestock, forestry and fisheries
	B	32,676	Extractive industries
	C	1,454,500	Manufacturing industry
	D	217,694	Supply of electric energy, gas, steam and air conditioning
	E	37,238	Water supply, sanitation activities, waste management and decontamination
	F	2,390,059	Construction
	G	3,138,008	Wholesale and retail; repair of motor vehicles and motorcycles
	H	497,186	Transportation and storage
	I	821,142	Hospitality
	J	398,044	Information and communications
	L	1,443,679	Real estate activities
	M	1,465,585	Professional, scientific and technical activities
	N	562,090	Administrative activities and auxiliary services
	R	244,530	Artistic, recreational and entertainment activities
	S	256,332	Other services
Deleted industries	K	48,857	Financial and insurance activities
	O	0	Public administration and defense; compulsory social security
	P	206,896	Education
	Q	302,770	Health and social services activities
	T	0	Household activities as employers of domestic personnel; household activities as producers of goods and services for their own use
	U	0	Activities of extraterritorial organizations and organizations
	-	235,505	Missing data

TABLE A4. Selecting the final sample.

Panel A: Details on the number of observations deleted using the *grup* variable.

	Code	Firm-year observations	Description
Preserved firms	1	13,296,895	Private
Deleted firms	0	0	Not reported
	2	884	Government-owned (unclassified)
	3	1,922	Central government-owned
	4	1,992	Regional government-owned
	5	8,150	Local government-owned

Panel B: Details on the number of observations deleted in Step 4.

	Remaining firm-year Observations
4.1 Negative or zero *output*, *sales*, *assets*, or *capital*	9,792,111
4.2 Negative *cash* or *cash* larger than *assets*	9,595,479
4.3 Negative *personnel costs* or *inputs*	9,585,526
4.4 Discrepancies in liability structure	9,357,713
4.5 Discrepancies in main balance sheet components or cash flow identity	9,007,777

Panel C: Summary of cleaning steps.

Sequentially applied filters	**Remaining firm-year observations**
0. Initial sample	17,767,139
1. Preserving observations passing the quality filter (*calidad*)	14,178,981
2. Preserving for-profit firms	13,296,895
3. Deleting firms with dubious data on year of incorporation	12,378,693
4. Deleting observations with apparent reporting inconsistencies or likely belonging to inactive firms	9,007,777
5. Deleting outliers	7,952,533
6. Preserving firms with at least 5 observations in the clean sample	2,919,644

This Table contains details about the cleaning steps that are applied to the CBI dataset. Please, see the text of Appendix A for additional explanations.

Appendix B: Additional results

TABLE B1. The impact of financial frictions on the allocation of cash flow. Financial constraints using the index proposed by Whited and Wu (2006).

	Dependent variable:			
	$\Delta Cash_{i,t}$	$Inv_{i,t}$	-$\Delta Debt_{i,t}$	*Net Div*
	(1)	(2)	(3)	(4)
$CF_{i,t}$	0.29***	0.42***	0.26***	0.03***
	(2.5e-3)	(2.3e-3)	(3.5e-3)	(9e-4)
$WW_{i,t}$	-0.06***	-0.12***	0.28***	-0.10***
	(3.2e-3)	(4e-3)	(7.3e-3)	(2.5e-3)
$CF_{i,t} \times WW_{i,t}$	-0.18***	-0.16***	0.47***	-0.13***
	(0.05)	(0.01)	(0.05)	(0.01)
Fixed-effects				
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.270	0.264	0.258	0.263

This Table reports the results of estimating specification (14) for each use of cash flow. *WW* is the index of financial frictions proposed by Whited and Wu (2006), which is computed as follows: -0.091 *CF/Assets* - 0.062 *{Net Div > 0}* + 0.021 * *Non-current liabilities/Assets* - 0.044 $Ln(Assets)_{1997}$ + 0.102 *isg* - 0.035 *sg*. $Ln(Assets)_{1997}$ is the log of *Assets* expressed in millions of 1997 dollars, *isg* is the annual change in the log of 3-digits industry sales, and *sg* is the annual change in the log of firm-level sales. All other variables are defined in Table A1. The specifications include the same controls as in equations (5) - (8) except for *Ln(Assets)*, which has been converted to a discrete variable indicating that *Ln(Assets)* is in the bottom (top) half of its distribution when it equals 0 (1). For brevity, the results related to the controls are omitted from the Table. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

TABLE B2. Summary statistics of negative and positive cash flow.

			Percentiles				
Variable	**Mean**	**Sd**	**5**	**25**	**50**	**75**	**95**
$NEG_{i,t}$	0.2141	0.41	0	0	0	0	1
$CF_{i,t} \times (1 - NEG_{i,t})$	0.0793	0.07	0.01	0.03	0.06	0.11	0.23
$CF_{i,t} \times NEG_{i,t}$	-0.0885	0.11	-0.3	-0.12	-0.05	-0.02	0

The results in table B4 show the effects of financial friction on the allocation of positive and negative cash flow across competing uses.

The coefficients in row (VI) of Table B4 indicate how the allocation of positive cash flow is affected by the *SA* index. The coefficients associated with $CF_t \times SA_{it}$ indicate that higher values of the *SA* index are, on average, associated with an increase in the cash flow sensitivities of investment and debt repayment at the expense of a reduction in

TABLE B3. The effects of financial frictions on the dynamic effects in the allocation of cash flow across competing uses.

	Dependent variable:					
	$\Delta Cash$	*Real investment*	*Financial investment*	Δ *Account receivables*	-$\Delta Debt$	*Net Div*
	(1)	(2)	(3)	(4)	(5)	(6)
CF_t	0.23***	0.15***	1.8e-3***	0.13***	0.43***	0.05***
	(0.001)	(0.002)	(4e-5)	(0.001)	(0.002)	(7e-4)
CF_{t-1}	0.01***	0.04***	4.1e-3***	-0.10***	0.03***	0.01***
	(0.001)	(0.001)	(4e-4)	(0.001)	(0.001)	(6e-4)
CF_{t-2}	0.02***	0.01***	2e-4	-0.03***	0.01***	-3.9e-3***
	(0.001)	(0.001)	(4e-4)	(0.001)	(0.001)	(6e-4)
$CF_t \times SA_{i,t}$	-8.4e-3***	-0.06***	-2.6e-3***	-0.04***	0.11***	-1e-3
	(0.001)	(0.002)	(5e-4)	(0.002)	(0.002)	(7e-4)
$CF_{t-1} \times SA_{i,t}$	-0.03***	0.01***	-2.1e-3***	0.04***	-7.2e-3***	-0.01***
	(0.001)	(0.001)	(5e-4)	(0.001)	(0.002)	(7e-4)
$CF_{t-2} \times SA_{i,t}$	-0.02***	0.01***	-1.6e-3***	0.03***	-6.7e-3***	-7.9e-3***
	(0.001)	(0.001)	(4e-4)	(0.001)	(0.001)	(6e-4)
Fixed-effects						
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.335	0.347	0.121	0.307	0.634	0.271

This Table reports the results of estimating an extension of the specifications in equations (5) - (8) where I include two lags of the cash flow variable and the interaction of the cash flow variable and its lags with the *SA* index. All variables are defined as in Table A1. The specifications include the same controls as in equations (5) - (8) but, for brevity, the results related to the controls are omitted from the Table. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

the cash flow sensitivity of the rest of the uses. This is slightly different to the average results reported in Table 6, where only the allocation of cash flow to debt repayment was positively related to the *SA* index.

The coefficients in row (VII) of Table B4 measure the differential impact of changes in the *SA* index on the allocation of cash flow when the cash flow is negative. To measure the effects of financial frictions in the allocation of negative cash flow, it is necessary to add up the coefficients in rows (VI) and (VII). The results indicate that higher values of the *SA* index are, on average, associated with an increase in the cash flow sensitivities of cash savings and debt repayment, and with a reduction in the cash flow sensitivity of investment. What it means is that more financially constrained firms rely more on reductions in cash savings and increases in debt to accommodate negative cash flow realizations than less financially constrained firms. In contrast, more financially constrained firms rely less on reductions in investment to accommodate negative cash flow realizations.

All in all, these results extend the findings in Table 6 by showing that the effects of financial frictions on the allocation of cash flow are contingent on the sign of the cash flow. In particular, one aspect that these results highlight is that the lower cash flow sensitivity of investment by more financially constrained firms is driven by the lower cash flow sensitivity of investment when the cash flow is negative, which means that the investment of more financially constrained firms is less sensitive to negative cash flow realizations than that of less financially constrained firms.

TABLE B4. The impact of financial frictions on the allocation of cash flow.

	Dependent variable:			
	$\Delta Cash_{i,t}$	$Inv_{i,t}$	$-\Delta Debt_{i,t}$	*Net Div*
	(1)	(2)	(3)	(4)
(I) $CF_{i,t}$	0.37***	0.28***	0.30***	0.05***
	(0.002)	(0.003)	(0.003)	(0.001)
(II) $SA_{i,t}$	-0.22***	-0.55***	0.74***	0.02***
	(0.001)	(0.002)	(0.002)	(4e-4)
(III) $NEG_{i,t}$	-3.9e-3***	2.4e-3***	1.2e-3**	2e-4
	(3e-4)	(5e-4)	(5e-4)	(2e-4)
(IV) $NEG_{i,t} \times SA_{i,t}$	-8.6e-3***	-7.1e-3***	0.02***	-1.3e-3***
	(3e-4)	(6e-4)	(6e-4)	(2e-4)
(V) $CF_{i,t} \times NEG_{i,t}$	-0.30***	0.15***	0.17***	-0.01***
	(0.003)	(0.005)	(0.005)	(0.002)
(VI) $CF_{i,t} \times SA_{i,t}$	-0.04***	0.01**	0.07***	-0.03***
	(0.002)	(0.003)	(0.003)	(0.001)
(VII) $CF_{i,t} \times NEG_{i,t} \times SA_{i,t}$	0.06***	-0.14***	0.04***	0.03***
	(0.003)	(0.007)	(0.007)	(0.002)
Impact of financial frictions on the allocation of				
positive cash flow (VI)	-0.04	0.01	0.07	-0.03
negative cash flow (VI + VII)	0.02	-0.13	0.11	0.00
Fixed-effects				
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.332	0.389	0.506	0.255

This Table reports the results of estimating an extension of specification (14) for each use of cash flow that includes the indicator *NEG* and its interactions with *SA* and *CF*. All variables are defined in Table A1. The specifications include the same controls as in equations (5) - (8) except for *Ln(Assets)*, which has been converted to a discrete variable indicating that *Ln(Assets)* is in the bottom (top) half of its distribution when it equals 0 (1). For brevity, the results related to the controls are omitted from the Table. ***, ** and * denote significance at the 1%, 5% and 10% levels, respectively. The standard errors are presented in parentheses.

The Allocation of Cash Flow by Spanish Firms: New Evidence on the Impact of Financial Frictions

Martin Farias

Workshop on New Insights from Financial Statements

October 17, 2024

Introduction

- Cash flow is one of the main sources of financing for firms
  - → Competitive uses of cash flow: Cash savings, investment, debt repayment, and dividend distribution

- Financial frictions imply costly and limited external financing
  - ⇛ The allocation of cash flow becomes a key determinant of corporate decisions

- Fazzari *et al* (1998) find that firms with financial constraints show greater sensitivity of investment to cash flow
  - → An extensive literature has tried to confirm this finding and characterize similar patterns concerning other uses of cash flow

This Paper

Questions

1. **How does the allocation of cash flow change when firms face greater financial frictions?**
2. **Is the relationship between financial frictions and cash flow allocation by firms heterogeneous?**

- I use data from Spanish firms from 2003 to 2019

  → I characterize the allocation of marginal cash flow among alternative uses

  → I estimate the average effect of financial frictions on the allocation of cash flow

  → I explore how the effects of financial frictions on cash flow allocation vary based on relevant firm characteristics

Main Results

1. **The Allocation of Cash Flow**
   From a marginal unit of cash flow, 29 cents go to cash savings, 44 to investment, 23 to debt repayment, and 4 to dividends

2. **Average Effects of Financial Frictions**
   On average, greater financial frictions are associated with an increase in the sensitivity of debt repayment to cash flow and with decreases in the sensitivities of the other uses

3. **Heterogeneous Effects of Financial Frictions**
   I document substantial heterogeneity in the effects of financial frictions on cash flow allocation across relevant firm characteristics

   → The results are useful for improving the design of policies aimed at promoting investment or deleveraging by firms (e.g., the direct aid lines implemented by the Spanish government in response to the COVID-19 crisis)

Data Sources

Data Sources

- I use data from the *Central de Balances Integrada* (provided by the Bank of Spain)
  - Balance sheets and income statements of Spanish firms for 2003-2012
- I focus on private for-profit firms. I exclude certain sectors (e.g., Financial, Education, Health)
- I apply a set of filters to exclude observations with unreliable information
  - These filters verify some basic accounting identities
- I require firms to have at least 5 observations to be part of the sample
  - This is to ensure sufficient information to measure firm-level fixed effects
- The final sample consists of 2,919,644 observations from 341,562 firms

The Allocation of Marginal Cash Flow

The Allocation of Marginal Cash Flow

- The empirical analysis is based on a ***cash flow identity***:

$$\text{CF}_t = \Delta\text{Cash}_t + \text{Investment}_t - \Delta\text{Debt}_t + \text{Dividends}_t.$$

- The basic specifications derived from this identity are:

$$\begin{aligned}
\Delta\text{Cash}_{it} &= \beta_1 \text{CF}_{it} + \gamma_1 X_{i,t-1} + f_{1,i} + \lambda_{1,t} + \epsilon_{1,it} \\
\text{Investment}_{it} &= \beta_2 \text{CF}_{it} + \gamma_2 X_{i,t-1} + f_{2,i} + \lambda_{2,t} + \epsilon_{2,it} \\
-\Delta\text{Debt}_{it} &= \beta_3 \text{CF}_{it} + \gamma_3 X_{i,t-1} + f_{3,i} + \lambda_{3,t} + \epsilon_{3,it} \\
\text{Dividends}_{it} &= \beta_4 \text{CF}_{it} + \gamma_4 X_{i,t-1} + f_{4,i} + \lambda_{4,t} + \epsilon_{4,it}
\end{aligned}$$

  - Control variables $X_{i,t-1}$: Ln(Assets), asset tangibility, leverage, sales growth
  - Estimated separately by OLS (equivalent to SUR estimation)
  - The $\beta's$ are ***cash flow sensitivities***, and must satisfy

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$$

The Allocation of Marginal Cash Flow

- From a marginal unit of cash flow, 29 cents go to cash savings, 44 to investment, 23 to debt repayment, and 4 to dividends
  - Cash flow sensitivities sum to 1
  - Estimates are in line with existing evidence (e.g., Chang *et al*, 2014)

	Dependent Variable:			
	$\Delta Cash$	*Investment*	$-\Delta Debt$	*Dividends*
	(1)	(2)	(3)	(4)
$CF_{i,t}$	0.29^{***}	0.44^{***}	0.23^{***}	0.04^{***}
	(1.3e-3)	(2.1e-3)	(2.3e-3)	(7e-4)
Controls	Yes	Yes	Yes	Yes
Fixed effects				
Firm	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.279	0.273	0.248	0.261

The Impact of Financial Frictions on the Allocation of Cash Flow

The Impact of Financial Frictions on the Allocation of Cash Flow: Measuring Financial Frictions

- In the absence of direct measures of financial frictions, the literature typically relies on *proxies*
- I use the *Size-Age (SA)* index created by Hadlock and Pierce (2010):

$$SA_{it} = -0.737 x Ln(Assets) + 0.043 x Ln(Assets)^2 - 0.04 x Age$$

  - Interpretation: Higher values of $SA_{it} \implies$ stricter financial frictions
    - $\rightarrow$ Captures the idea that younger and smaller firms are more likely to face financial frictions
  - I verify that the SA index correlates with firm characteristics typically associated with financial constraints (e.g., leverage, cash holdings) Table

The Impact of Financial Frictions on the Allocation of Cash Flow: Average Effects

- An increase of 1 unit in the SA index is associated with
  - an increase of 6 cents in the sensitivity of debt repayment to cash flow
  - decreases of 2, 3, and 1 cents in the sensitivities of cash savings, investment, and dividends to cash flow, respectively
- The results are consistent with theories that emphasize the value firms place on financial flexibility

	$\Delta Cash$	*Inverstment*	-$\Delta Debt$	*Dividends*
$CF_{i,t}$	0.24^{***}	0.33^{***}	0.38^{***}	0.05^{***}
	(0.001)	(0.002)	(0.002)	(0.001)
$SA_{i,t}$	-0.23^{***}	-0.54^{***}	0.75^{***}	0.02^{***}
	(0.001)	(0.002)	(0.002)	(4e-4)
$CF_{i,t} \times SA_{i,t}$	-0.02^{***}	-0.03^{***}	0.06^{***}	-0.01^{***}
	(0.001)	(0.002)	(0.003)	(0.001)
Controls & Fixed effects	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.332	0.389	0.506	0.255

The Impact of Financial Frictions on the Allocation of Cash Flow: Heterogeneous Effects

- On average, higher values of the SA index are associated with an increase in the sensitivity of debt repayment to cash flow
  - → **Can this result be generalized to all firms?**

- I measure heterogeneity through variables that characterize the firm's condition
  - Leverage, cash holdings, capital stock, and sales growth
  - Taking each variable one at a time

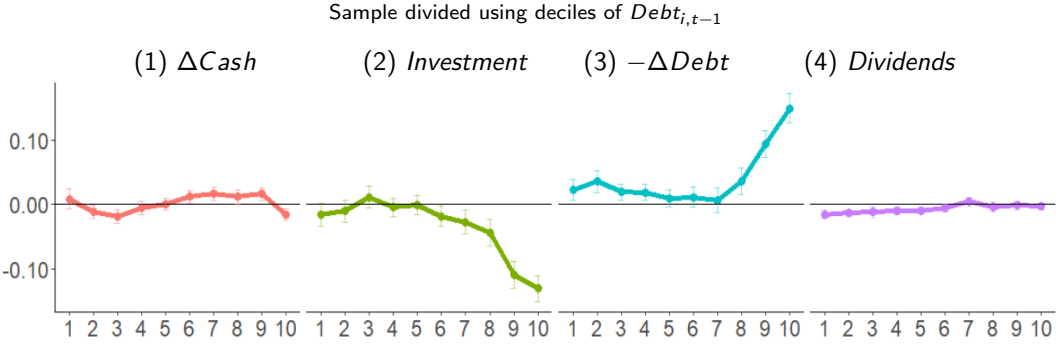
    → I divide the sample into deciles (using lags of the variable)

    → I estimate the average effect of financial frictions on cash flow allocation in each decile

    → I compare the coefficient associated with $CF_{i,t} \times SA_{i,t}$ across deciles

The Impact of Financial Frictions on the Allocation of Cash Flow: Heterogeneous Effects Across Leverage

- Changes in the SA index have small effects on cash flow allocation for firms in the lower deciles of leverage
- For firms in the higher deciles of leverage, changes in the SA index are associated with
  - → an increase in the sensitivity of debt repayment to cash flow
  - → a decrease in the sensitivity of investment to cash flow



The Impact of Financial Frictions on the Allocation of Cash Flow: Discussion about Policy Relevance

- The results are useful for improving the design of policies aimed at promoting investment or deleveraging of firms
  - For example, cash transfers used to guarantee the solvency of firms implemented by the Spanish government in response to the COVID-19 crisis
    - → Among them, there was one program aimed at repaying debts incurred since March 2020 (payments to suppliers, utilities, wages, rents, or financial debt reduction)
- Understanding the allocation of cash flow (and the effects of financial frictions on it) is crucial to anticipate how different firms will use these transfers differently
  - For example, firms with high levels of leverage are more likely to use resources to repay debt

⇛ **This can help improve the targeting of these policies, increasing their effectiveness**

Additional Analysis

Additional Analysis

- Decomposition of cash flow into persistent and transitory components
  - All uses except debt repayment are more sensitive to the persistent component than to the transitory one

Results

- Disaggregation of Investment into real investment, financial investment, and trade credit provision
  - All previous findings apply to real investment
  - Trade credit provision is a relevant use of cash flow

Results

- Test of lags in the response to cash flow changes
  - Real investment and debt repayment increase with cash flow lags, at the expense of less cash savings and trade credit provision

Results

- Test of asymmetric cash flow sensitivities
  - Firms rely heavily on external financing when facing negative cash flows

Results

Conclusions and remarks to statistical producers

Conclusions and remarks to statistical producers

- I use data from Spanish non-financial firms to answer two questions
  1. **How does cash flow allocation change when firms face greater financial frictions?**
  2. **Is the relationship between financial frictions and cash flow allocation heterogeneous across firms?**

- On average, greater financial frictions are associated with $\uparrow$ sensitivity of debt repayment to cash flow and $\downarrow$ sensitivities of the rest of the uses to cash flow

- I document the existence of heterogeneity in the effects of financial frictions on cash flow allocation across variables that characterize the firm's condition

- I discuss how the results are useful for improving the targeting of policies aimed at promoting investment or deleveraging of firms

Conclusions and remarks to statistical producers

- The main weakness of the analysis is the imperfect measurement of financial frictions
- Better measures of financial frictions could be derived from the information contained in other datasets such as
  - Survey on the access to finance of enterprises (SAFE)
  - Banco de España Survey on Business Activity (EBAE)
- Currently, it is not possible to merge the *Central de Balances Integrada* with SAFE or EBAE
  - **Being able to merge these datasets while guaranteeing confindentiality would make the *Central de Balances Integrada* a much more appealing dataset and would greatly benefit the research community**

Another Related Project

"Decoding Distress: The Behavior of Firms Preceding Bankruptcy"

1. **What characteristics of firms predict bankruptcy filing?**
2. **What is the typical dynamics of firms in the years leading up to bankruptcy filing?**

- I use data from Spanish firms from 2003 to 2019 (*Integrated Central Balance Sheet*)
- I estimate a model to predict the decision to enter bankruptcy
  - → I identify the most important predictors: net equity, profitability, liquidity, investment, aggregate credit, among others
- I provide a graphical description of the typical dynamics of firms in the 5 years before entering bankruptcy
  - → Firms enter bankruptcy only after experiencing severe financial difficulties (large reductions in profitability and net equity)
    - ⇛ Consistent with the view that Spanish firms use bankruptcy as a last resort
    - ⇛ **The results invite reflection on the effectiveness of the Spanish bankruptcy law in restructuring viable firms in time**

Appendix

Relationship between the Financial Frictions Index (SA Index) and Firm Characteristics

	Explanatory Variable: $SA_{i,t}$			
	Period:			
	2003-2019	**2003-2007**	**2008-2012**	**2013-2019**
Dependent Variable	(1)	(2)	(3)	(4)
Sales growth$_{i,t}$	-0.005***	-0.001**	-0.006***	-0.006***
	(2e-4)	(4e-4)	(5e-4)	(3e-4)
Real Investment$_{i,t}$	-0.003***	3e-4	-0.002***	-0.004***
	(1e-4)	(2e-4)	(2e-4)	(2e-4)
Cash flow$_{i,t}$	-0.003***	0.003***	-0.008***	-0.003***
	(2e-4)	(2e-4)	(3e-4)	(2e-4)
Positive Dividends$_{i,t}$	-0.056***	-0.056***	-0.054***	-0.056***
	(5e-4)	(9e-4)	(8e-4)	(6e-4)
Cash$_{i,t}$	0.004**	0.007***	0.002***	0.004***
	(4e-4)	(5e-4)	(5e-4)	(5e-4)
Debt$_{i,t}$	0.083***	0.037***	0.082***	0.105***
	(1.0e-3)	(1.1e-4)	(1.3e-4)	(1.4e-4)

Back

The Allocation of the Transitory and Persistent Components of Cash Flow

	Dependent Variable:			
	$\Delta Cash$	*Investment*	$-\Delta Debt$	*Dividends*
$TranYestory_{i,t}$	0.27^{***}	0.43^{***}	0.27^{***}	0.04^{***}
	(0.002)	(0.003)	(0.003)	(8e-4)
$PerYesstent_{i,t}$	0.31^{***}	0.45^{***}	0.18^{***}	0.06^{***}
	(0.002)	(0.003)	(0.004)	(0.001)
Controls & Fixed effects	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.278	0.274	0.247	0.261

Back

The Allocation of the Three Components of Investment: Real Investment, Financial Investment, and Provision of Trade Credit

	Dependent Variable:		
	Real Investment	*Financial Investment*	Δ *Account Receivables*
CF_t	0.23^{***}	0.01^{***}	0.19^{***}
	(1.6e-4)	(4e-4)	(1.5e-4)
Controls & Fixed effects	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644
R^2	0.252	0.11	40.165

Back

Dynamic Effects in the Allocation of Cash Flow Across Competing Uses

	Dependent Variable:					
	Δ*Efectivo*	*Real Investment*	*Financial Investment*	Δ *Account Receivables*	-Δ*Debt*	*Dividends*
	(1)	(2)	(3)	(4)	(5)	(6)
CF_t	0.29***	0.23***	0.01***	0.22***	0.22***	0.04***
	(0.001)	(0.002)	(4e-4)	(0.002)	(0.002)	(7e-4)
CF_{t-1}	5e-5	0.03***	2.8e-3***	-0.11***	0.07***	0.01***
	(0.001)	(0.001)	(4e-4)	(0.001)	(0.002)	(6e-4)
CF_{t-2}	0.02***	0.01***	1e-4	-0.03***	4.6e-3***	-3.8e-3***
	(0.001)	(0.001)	(4e-4)	(0.001)	(0.002)	(6e-4)
Controls & Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.279	0.252	0.114	0.171	0.249	0.261

Back

The Allocation of Marginal Cash Flow: Asymmetries Between Positive and Negative Flows

	Dependent Variable:			
	$\Delta Cash$	*Investment*	-$\Delta Debt$	*Dividends*
	(1)	(2)	(3)	(4)
$CF_{i,t}$	0.45***	0.49***	0.02***	0.04***
	(0.002)	(0.003)	(0.003)	(9e-4)
$NEG_{i,t}$	-4e-4	1.7e-3***	-2.0e-3**	1e-3***
	(2e-4)	(4e-4)	(5e-4)	(2e-4)
$CF_{i,t} \times NEG_{i,t}$	-0.35***	-0.10***	0.45***	-2e-4
	(0.003)	(0.005)	(0.005)	(0.002)
Controls & Fixed effects	Yes	Yes	Yes	Yes
Observations	2,919,644	2,919,644	2,919,644	2,919,644
R^2	0.289	0.274	0.255	0.261

Back

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

A study on complementary indicators to the ICR for identifying distressed firms[1]

Jiyoung Choi, Sejong Lee, and Hye Won Shin,
Bank of Korea

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

A study on Complementary Indicators to the ICR for Identifying Distressed Firms

Sejong Lee[1], Hye Won Shin[2], Jiyoung Choi[3]

Abstract

This study aims to examine the practical issues associated with the ICR widely used for identifying distressed firms and, based on prior research, to find an appropriate alternative indicator. The ICR calculated as operating profit divided by interest expenses, is commonly used to determine a distressed firm. However, the ICR is undefined when interest expenses are zero. And when interest expenses are very low, even small changes in profit levels can significantly alter the assessment of the firm's operating status. Thus, this study identifies complementary indicators to the ICR. To achieve this, the study examines the desirable characteristics of indicators for identifying distressed firms, reviews various alternative indicators used in existing research, and evaluates their applicability in light of these desirable characteristics. The results suggest that the return on operating assets (ROOA) or the return on net operating assets (RNOA) could serve as robust complementary indicators to the ICR for identifying distressed firms.

Keywords: Distressed firms, Zombie firms, ICR (Interest Coverage Ratio), ROOA (Return on Operating Asset), RNOA (Return on Net Operating Asset)

JEL classification: L25, M20, M40

Contents

A study on Complementary Indicators to the ICR for Identifying Distressed Firms..... 1

1. Introduction 2

2. Overview of Distressed Firm Indicators 3

    2.1 Review of the determinants of distressed firms 3

    2.2 Review of the conditions for Alternative Indicators 7

3. Review of the characteristics of Alternative Indicators 8

4. Conclusion and Implications 16

References 17

1 Associate Professor, Business School University of Seoul, E-mail: lee0819@uos.ac.kr

2 Senior Economist, The Bank of Korea, E-mail: hw.shin@bok.or.kr

3 Economist, The Bank of Korea. Korea, E-mail: cjy486@bok.or.kr

Disclaimer: The views expressed in this article do not necessarily reflect those of Bank of Korea.

1. Introduction

A company's Interest Coverage Ratio (ICR) is defined as operating income divided by interest expense. If the ICR is greater than 1, it indicates that the company can cover its interest payments with operating income and maintain normal operations. Conversely, a low ICR suggests that sustaining normal business activities may be challenging. In principle, the ICR is used as one of the indicators to assess a company's profitability. Due to its relatively simple calculation and high comprehensibility, it is widely used in previous studies and reports to identify distressed firms.

The Bank of Korea (June 2015) defines marginal firms as those with an ICR of 1 or less for three consecutive years. Additionally, the Bank of Korea (June 2024) classifies firms with an ICR below 1 as vulnerable firms in terms of ICR. The Bank compiles and publishes the changes in the proportion of these vulnerable firms each year.

However, specifically, relying solely on the ICR to calculate the proportion of distressed firms or to analyze a company's vulnerability poses the following issues:

First, a company's interest expense can be zero not only when the company has no borrowings and therefore incurs no interest expense, but also when all borrowing interest is capitalized and interest expense is included as part of depreciation. In such cases, the denominator of the ICR becomes zero and the indicator itself would be undefined, which raises concerns about its universal applicability. This issue is less pronounced when analyzing relatively small samples, such as listed companies, but it becomes a major problem when surveying large samples, such as nationwide business survey.

For example, according to a press release by the Bank of Korea (2023), of 910,206 companies surveyed in the Financial Statement Analysis for 2022, 441,958 companies, or 48.6%, reported zero interest expense. This means that nearly half of all companies would be excluded from an ICR-based analysis of distressed firms in Korea. Thus, if policymakers use the ICR to assess changes in distressed firms without considering its limited applicability, the results could be distorted.

Furthermore, even if a company's interest expense is not zero, but close to zero, the stability and continuity of the ICR indicator can be compromised. The ICR, as a continuous indicator, should ideally reflect profitability. However, the operating income (the numerator) can have both positive and negative values, so for a company with a small interest expense, the magnitude of the ICR can change dramatically as soon as the sign of the company's operating income changes from negative in one year to positive in the next. This means that a company that is judged to be in a very severe vulnerable situation according to the ICR can change its judgment to a good company in the next year with only a small improvement in operating profit.

Therefore, it is necessary to find an alternative or complementary indicator that can address the limitations of the ICR in identifying distressed firms. In other words, the complementary indicator should be able to determine whether a company is a distressed company by covering all 48.6% of companies

that do not have a defined ICR as of 2022, and should have stability that does not change significantly due to small changes in corporate profits. In addition, since the ICR has been widely used to determine distressed firms, it should be consistent with the purpose of the existing ICR and reflect the actual level of business activity of the company.

Therefore, this study examines the conditions that should be met by the distressed firm judgment indicators as a complementary indicator to the ICR from a practical perspective, and examines the indicators that can be used to judge distressed firms (hereinafter referred to as Distressed Firm Indicators) based on various previous studies.[4] As a result, we proposed ROOA(Return On Operating Assets) or RNOA(Return on Net Operating Assets), which is defined as a complementary indicator that is consistent with the purpose of the ICR and can compensate for the shortcomings of the ICR in determining distressed firms, by modifying the previous research.

This study is organized as follows. In Chapter 2, we reviewed the indicators that can be used to determine distressed firms from the literature and summarized the desirable characteristics of the indicators. Chapter 3 analyzes the characteristics of each candidate indicator in light of the desirable characteristics of a complementary indicator. Chapter 4 concludes by proposing an indicator that can compensate for the shortcomings of the ICR based on the analysis in Chapter 3.

2. Overview of Distressed Firm Indicators

2.1 Review of the determinants of distressed firms

2.1.1. ICR

The ICR is one of the indicators of a company's profitability and measures the ability of a company's operating income to cover interest expenses. Also known as ICR, it is calculated by dividing a company's operating income by its interest expense. The calculation method for the ICR for the year is as follows.

$$\text{Coverage}_{i,t} = \frac{\text{Operating Income}_{i,t}}{\text{Interexp}_{i,t}}$$

$\text{Coverage}_{i,t}$: ICR at year t for firm i $\text{Operating Income}_{i,t}$: Operatinng Income at year t for firm i

$\text{Interexp}_{i,t}$: Interest expense at year t for firm i

If the ICR is less than 1, it means that the interest expense is greater than the operating income, and the company is classified as a distressed company that cannot cover interest from its operating activities. In

4 Different studies have used the terms zombie, marginal, distressed firms, but for the purposes of this study, we will refer to them as distressed firms.

addition, a larger ICR indicates that a firm's operating income significantly exceeds its interest expense and can be used as a continuous indicator of a firm's profitability. Adalet McGowan et al. (2018) classify firms with ICRs below 1 for three consecutive years as distressed firms, citing high comparability as an advantage of using ICRs as the main criterion. Alvarez et al. (2023) classified a firm as a distressed firm if it has been established for more than five years, has an ICR of 1 or less, and has experienced capital erosion for more than three consecutive years.

2.1.2 Caballero et al. (2008): CHK

It is an indicator of distressed firms introduced in Caballero et al. (2008) and defined based on the idea that distressed firms tend to be prolonged through low policy interest rates. In this study, we refer to it as the CHK indicator after the initials of the authors' last names Caballero, Hoshi, and Kashyap. It combines a firm's debt structure with the market interest rate to calculate the minimum required interest rate, and if a firm's actual interest rate is lower than the minimum required interest rate, it is classified as a distressed firm because it is prolonging its life with low-interest loans from financial institutions. The specific calculation of a firm's minimum required interest payment R* is as follows.[5]

$$R*_{i,t} = rs_{t-1}BS_{i,t-1} + \left(\frac{1}{5}\sum_{j=1}^{5} rl_{t-j}\right)BL_{i,t-1} + rcb_{\text{min over last 5 years, t}} \times Bonds_{i,t-1}$$

$R*_{i,t}$: minimum required interest rate at year t for firm i

rs_{t-1}: short − term market interest rate at year $t-1$, treasury bond rate 1 year

$BS_{i,t-1}$: short − term borrowings at year $t-1$ for firm i

rl_{t-j}: long − term market interest rate at year $t-j$, treasury bond rate 5 year

$BL_{i,t-1}$: long − term borrowings at year $t-1$ for firm i

$rcb_{\text{min over last 5 years, t}}$: minimum corporate bond rate over last 5 years t years ago, CP interest rate

$Bonds_{i,t-1}$: bonds outstanding at $t-1$ year for i firm . corporate bonds among liabilites

In addition, the firm's minimum required interest rate (Rminrate) and the firm's paid interest rate (Rrate) are calculated as follows using the firm's borrowing structure and the firm's minimum required interest payment R*.

$$Rminrate_{i,t} = R*_{i,t}/(BS_{i,t-1} + BL_{i,t-1} + Bonds_{i,t-1})$$

[5] In fact, Caballero et al. (2008) used the prime rate, which is the preferential interest rate for banks, to calculate the minimum required interest payment, but we used the treasury bond rate due to the lack of general statistics on the prime rate in Korea. Therefore, the measure of CHK in this study is likely to be more conservative than that in Caballero et al. (2008).

$$\mathrm{Rrate}_{i,t} = \text{Interest expense}_{i,t} / (\mathrm{BS}_{i,t-1} + \mathrm{BL}_{i,t-1} + \mathrm{Bonds}_{i,t-1})$$

The CHK indicator shows that when a firm's paid interest rate (Rrate) is less than the minimum required interest rate (Rminrate), the firm is assumed to be prolonging its existence by borrowing from financial institutions at low interest rates. In this case, the firm is categorized as a distressed firm.

2.1.3 Fukuda and Nakamura (2011): FN

This indicator is introduced in Fukuda and Nakamura (2011) and is referred to as the FN indicator after the first letter of the authors' last names. Fukuda and Nakamura (2011) basically use the CHK indicator to determine distressed firms, but they point out the following two aspects as shortcomings of the existing CHK indicator.

First, there is a possibility that the CHK indicator may classify a firm as a distressed firm based on the level of financing rate alone, regardless of the firm's soundness, and second, there is a possibility that a firm whose debt continues to increase due to management problems may not be classified as a distressed firm if the financing rate is above a certain level. Therefore, this study follows the methodology of Fukuda and Nakamura (2011) and uses the CHK indicator, but additionally considers the following two criteria.

First, if a firm's operating income was greater than the minimum required interest payment, it was excluded as a distressed firm, even if it was a distressed firm under the CHK criteria. Second, if a firm's total borrowing liabilities were more than half of its assets and its borrowing liabilities were increasing, it was further included as a distressed firm even if it was not a distressed firm under CHK.[6]

2.1.4 Acharya et al. (2020): ACEE

This indicator was introduced by Acharya et al. (2020) and is referred to as the ACEE indicator after the authors' last names. ACEE uses both the ICR and the debt-to-equity ratio (leverage) to determine distressed firms, and also includes the low borrowing rate considered in the CHK and FN indicators. According to ACEE's criteria, a company is classified as distressed if it meets both of the following criteria

First, based on the ICR and debt-to-equity ratio, the ICR is less than the industry-year median and the leverage ratio is greater than the industry-year median. Second, the magnitude of interest expense/borrowed debt (i.e., interest rate) is smaller than the magnitude of the interest rate for blue-chip companies.[7]

---

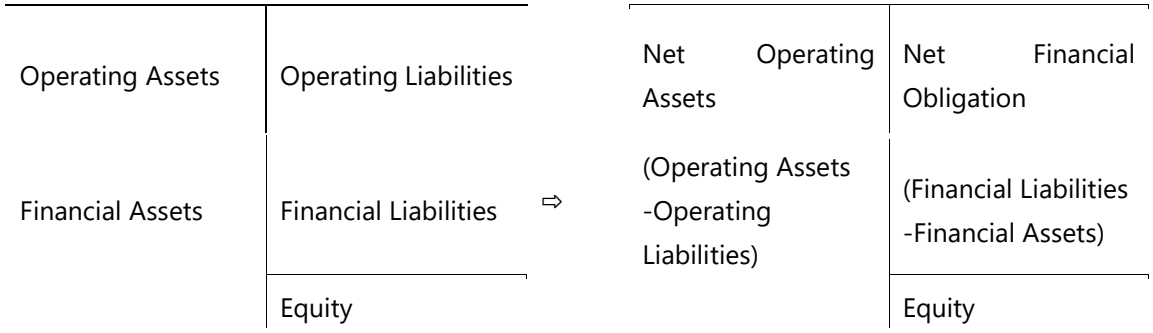
[6] In this study, we measure a firm's total borrowing liabilities as "short-term borrowings + current long-term liabilities + corporate bonds + long-term borrowings + finance lease liabilities", referring to Fukuda and Nakamura (2011).

[7] In addition, Acharya et al. (2020) determined that the second criterion is satisfied if the interest rate is lower than the interest

2.1.5 RNOA and ROOA

Return on Net Operating Assets (RNOA) is a metric originally used by Penman and Zhang (2002), which is defined as operating income divided by Net Operating Assets. Penman and Zhang (2002) tracked changes in RNOA based on the quality of earnings, where Net Operating Assets is computed by subtracting operating liabilities from operating assets, and can be measured by reclassifying the statement of financial position as follows

Reclassification of the statement of financial position Figure 1



As shown in the figure 1 above, a typical statement of financial position shows assets on the left-hand side of the balance sheet and liabilities and equity on the right-hand side of the balance sheet. Assets can be further categorized into Operating Assets, which are used in a company's primary operating activities, and Financial Assets, which represent the company's holdings of financial assets, while liabilities can similarly be categorized into Operating Liabilities, which are liabilities incurred in the company's operating activities, and Financial Liabilities, which represent the company's obligations. When both Operating Assets and Operating Liabilities are moved to the debit side, Operating Liabilities are represented as a subtraction from Operating Assets, and the indicator of Operating Assets minus Operating Liabilities is defined as Net Operating Assets (NOA). On the other hand, if Financial Assets are moved to the right-hand side of the balance sheet, Net Financial Obligation (NFO) is defined as Financial Liabilities minus Financial Assets, and the reclassification of the balance sheet, which was previously expressed as Assets = Liabilities + Equity, is expressed as Net Operating Assets (NOA) = Net Financial Obligation (NFO) + Equity.

Penman and Zhang (2002) measure RNOA using NOA as the denominator and Core Earnings, which is the operating income, as the numerator. This measure of RNOA looks similar to Return on Assets (ROA), which is a common profitability indicator, but unlike ROA, it uses operating income rather than net

rate of companies with a corporate bond rating of AAA or higher, but since there are few companies in Korea with a corporate bond rating of AAA or higher except financial companies, the second criterion is satisfied if the interest rate is lower than the median of companies with a corporate bond rating of AA- or higher for the application of the indicator.

income in the numerator.[8] In addition, total assets, the denominator of ROA, can change even if a company does not conduct any operating activities if it holds borrowed funds in the form of financial assets, whereas NOA, the denominator of RNOA, does not change with changes in financial structure unless the size of the company's operating assets changes.

In this study, RNOA is defined as "operating income - interest expense" divided by NOA, which is a slight modification of Penman and Zhang's (2002) RNOA. By slightly modifying the existing RNOA, interest expense distributed to creditors can be excluded from the return calculation, and the modified RNOA metric has the advantage of being consistent with the ICR that has been widely used in the past. In addition, we further defined Return on Operating Assets (ROOA) using operating assets rather than net operating assets as the denominator to examine its usefulness as an alternative measure to ICR.[9] The RNOA (ROOA) metric is virtually the same as the ICR, and has the advantage that it can be applied even when interest expense is zero, and the metric has continuity.

2.2 Review of the conditions for Alternative Indicator

In light of the above indicators, there are two types of alternative indicators that can be used to determine distressed firms in the literature. First, the CHK, FN, and ACEE indicators are based on the level of financing interest rates of individual firms, assuming that firms on the brink of bankruptcy become distressed if they are supported by social interventions such as low-interest support from policy funds. Second, another approach emphasizes profitability indicators, particularly the ICR, as low profitability is often a key sign of distress. Considering the practical applicability of complementary indicator, we can summarize the following characteristics of an alternative indicator for determining distressed firms as an alternative to the ICR.

First, and most importantly, the scope of applicable companies needs to be broader than the existing approach using the ICR for universality, i.e., it is necessary to devise an alternative indicator to determine whether a company is a distressed company by including all companies with zero interest expense.

The second is understandability, which requires indicators to be intuitive and simple in concept so that they can be easily understood by non-expert users of the indicators as well as researchers.

The third is affordability, which means that the timely construction of the indicator requires quick organization and aggregation of data. Therefore, it is necessary to design an indicator that can be completed in a short time without requiring overly complex additional data.

The fourth is that there should be no discontinuity in the indicator that may occur when using the ICR as an indicator for determining distressed firms, such as excessive fluctuations in the judgment of the

---

[8] Of course, there are also ways to calculate ROA that use operating income as the numerator rather than net income.

[9] Referring to Penman (2010), operating assets and operating liabilities were measured as follows

- Operating assets: Total assets - Financial assets (Financial assets: cash and cash equivalents + short-term financial assets + long-term financial assets, except that cash equal to 0.5% of sales is classified as operating assets)
- Operating liabilities: Total liabilities - Financial liabilities (Financial liabilities: Short-term financial liabilities + long-term financial liabilities + finance lease liabilities + preferred share capital)

degree of vulnerability of the company. It is desirable for an indicator to not only determine whether a firm is distressed but also continuously reflect the degree of the firm's vulnerability through the size of the indicator.

Fifth, it is necessary to maintain a certain level of consistency with the ICR, which has been widely used as an indicator for determining distressed firms. Therefore, the judgment of distressed firms based on the alternative indicator and the judgment of distressed firms based on the ICR should not be significantly different, so that the existing position regarding the judgment of distressed firms can be maintained.

Sixth, the indicator used to determine distressed firms should reflect the actual level of a firm's activities. In other words, there must be a systematic correlation between the distressed firm indicator and declines in areas such as investment, employment, and productivity, which can occur when a firm is vulnerable.

Based on these six conditions, we reviewed the content and characteristics of the indicators that could be considered as alternative complements to the ICR.

3. Review of the characteristics of Alternative Indicators

3.1 Universality

This is done by verifying that the indicator can be defined for a larger sample than the traditional ICR. The universality of the indicator is measured by the number of companies for which the indicator can be measured among all available observations, and the sample for the analysis was composed of companies in Fnguide's corporate data that met the following conditions

- KOSPI-KOSDAQ companies excluding financials from 2011 to 2022
- December year-end entities only
- No missing assets, liabilities, or equity in the database
- The accounting equation is complete (Assets = Liabilities + Equity)
- Companies with data for two consecutive years

The total number of available firm observations satisfying the above conditions was measured to be 20,813. After reviewing the universality of each indicator, we found that 19,908 firms, or 95.7% of the total, were able to define an ICR. The return on operating assets (ROOA) was applicable to all 20,813 companies, which is 100%, and for the return on net operating assets (RNOA), the number of applicable companies was determined to be 20,510 companies, which is 98.5%, as there are cases where the denominator, net operating assets, has a negative value. For the CHK, FN, and ACEE indicators, 17,941

(86.2%), 17,941 (86.2%), and 18,665 (89.6%) firms, respectively, were measurable, indicating ROOA and RNOA are more common than ICR.

Universality Table 1

Year	Obs	ICR	Return on operating assets (ROOA)	Net Sales Return on Assets (RNOA)	CHK	FN	ACEE
2012	100.0	94.9	100.0	98.5	90.4	90.4	90.1
2013	100.0	94.4	100.0	98.4	89.2	89.2	89.0
2014	100.0	94.0	100.0	98.7	87.9	87.9	88.0
2015	100.0	94.0	100.0	98.4	87.7	87.7	87.4
2016	100.0	92.7	100.0	98.1	85.3	85.3	85.3
2017	100.0	92.8	100.0	99.0	84.7	84.7	84.4
2018	100.0	93.0	100.0	98.4	85.1	85.1	85.3
2019	100.0	99.1	100.0	98.8	84.2	84.2	85.2
2020	100.0	99.4	100.0	98.8	83.9	83.9	97.4
2021	100.0	99.1	100.0	98.6	85.2	85.2	97.4
2022	100.0	99.1	100.0	98.3	84.7	84.7	97.0
Total	100.0	95.7	100.0	98.5	86.2	86.2	89.6

Note: The data is expressed in percentage units.

3.2. Understandability and affordability of metric creation

Compared to the ICR, the understandability of this indicator is not likely to be significantly different for ROOA and RNOA. However, indicators such as CHK, FN, and ACEE are concepts that are discussed in academic papers and are very complicated to create. On the other hand, the affordability of the indicator can be measured by the amount of time and effort invested in creating the indicator, which can be measured by the length of the code for each indicator.[10] In the case of ICR, it can be written in a simple 27 lines of code, and ROOA and RNOA can also be written in less than 100 lines of code, but indicators such as CHK, FN, and ACEE require more than 200 lines of code depending on the indicator. Therefore, from the perspective of the understandability of the indicator and the affordability of the indicator, ROOA or RNOA are appropriate as an alternative indicator.

3.3 Continuity

It is highly unlikely that a distressed firm will suddenly become a sound firm in the following year without any fundamental change, or vice versa. However, if the ICR is used as an indicator for determining

[10] Measured by the length of code in SAS used to analyze the data.

distressed firms, it is possible for a firm with a very small number of interest expenses to experience such a change, which requires an indicator that can compensate for this shortcoming.

Therefore, to determine the continuity of each indicator, we extracted the sample used to determine the universality of the indicator and measured the proportion of the sample that was in the bottom (top) 10% of vulnerability in a given year and changed to the top (bottom) 10% in the following year. However, for the CHK, FN, and ACEE indicators, continuity cannot be determined because they are defined as 0 or 1 values.

Continuity Table 2

Year	ICR	(Extreme Change)	ROOA	(Extreme Change)	RNOA	(Extreme Change)
2013	100.0	0.9	100.0	0.3	100.0	0.4
2014	100.0	0.6	100.0	0.4	100.0	0.6
2015	100.0	0.9	100.0	0.6	100.0	0.8
2016	100.0	0.9	100.0	0.7	100.0	0.8
2017	100.0	0.8	100.0	0.5	100.0	0.8
2018	100.0	0.8	100.0	0.4	100.0	0.3
2019	100.0	0.9	100.0	0.6	100.0	0.8
2020	100.0	1.0	100.0	0.7	100.0	0.5
2021	100.0	0.8	100.0	0.4	100.0	0.2
2022	100.0	0.8	100.0	0.2	100.0	0.2
Total	100.0	0.8	100.0	0.5	100.0	0.5

Note: The data is expressed in percentage units.

As shown in Table 2, the percentage of extreme changes in the ICR indicator is about 0.8%, but the percentage of extreme changes in ROOA and RNOA is about 0.5%, indicating that the percentage of extreme changes in the ROOA and RNOA is smaller.

3.4 Consistency with ICR

Since the ICR has been widely used as an indicator for determining distressed firms, it is also important to check its consistency with the ICR to ensure timely consistency in the determination. Therefore, when defining a distressed firm as one with an ICR of less than 1, we examined how other indicators classify firms as distressed. For this analysis, we used only observations of firms for which the ICR is measurable in the sample used to determine the universality of the indicator, and classified distressed firms as distressed firms if the ROOA and RNOA are less than zero, while the CHK, FN, and ACEE indicators, which are defined as 0 and 1, are classified as distressed firms if they have a value of 1.

Consistency with ICR Table 3

ICR	Obs	Judgment	ROOA	RNOA	CHK	FN	ACEE
Non-Distressed firms	13,450 (100.00%)	Non-Distressed	100.00%	98.81%	76.86%	89.08%	90.23%
		Distressed	0.00%	0.00%	12.39%	0.17%	3.49%
		Unclassifiable	0.00%	1.19%	10.75%	10.75%	6.28%
Distressed firms	6,458 (100.00%)	Non-Distressed	0.03%	0.03%	85.06%	82.73%	86.22%
		Distressed	99.97%	98.41%	4.85%	7.17%	7.45%
		Unclassifiable	0.00%	1.56%	10.10%	10.10%	6.33%

As shown in Table 3, when comparing the classification of distressed firms based on the ICR with other indicators, it was found that in cases where a firm was not considered distressed based on the ICR, the other indicators also judged the firm as non-distressed with a high proportion. However, when a firm is identified as distressed based on the ICR, more than 98% of the time, the ROOA and RNOA also classify the firm as distressed, demonstrating a high degree of consistency with the ICR-based assessment. On the other hand, indicators such as CHK, FN, and ACEE identify less than 10% of firms as distressed. Therefore, based on this consistency analysis with the ICR, the CHK, FN, and ACEE indicators are not considered applicable.

3.5 Relevance to business activities

Due to the characteristics of distressed firms that are difficult to generate normal profits, in order for the distressed firm judgment indicators to be reliable, it is believed that there should be a significant relationship between the indicators of distressed firms and the level of firm business activity. The independent variables are ICR, return on operating assets (ROOA), return on net operating assets (RNOA), CHK, FN, and ACEE, which are the six indicators. Referring to Acharya et at (2019), the dependent variables are six: employment growth rate (employgrowth), productivity(productivity), investment in tangible assets (CAPX), cash asset growth (dcash), debt growth (ddebt), and return on assets (ROA). The dependent variable was set as an indicator at year t+1, in order to observe the relationship between the distressed firm indicators and the following year's firm business activities. The control variables are each firm's size (lnassets), leverage (Leverage), ratio of tangible asset (Tangibility), firm age (lnage), and Market to Book ratio (MTB), and firm fixed effects and year fixed effects were added. The regression equation and the measures of dependent and control variables are as follows.

$$\begin{aligned}\text{Activity}_{i,t+1} = \beta_0 + \beta_1 \text{distress}_{i,t} + \beta_2 \text{lnassets}_{i,t} + \beta_3 \text{Levergae}_{i,t} + \beta_4 \text{Tangibility}_{i,t} \\ + \beta_5 \text{lnage}_{i,t} + \beta_6 \text{MTB}_{i,t} + \text{firmfixed} + \text{yearfixed} + \varepsilon_{i,t}\end{aligned}$$

$distress_{I,t}$: Distressed firm indicator for firm i at year t.

$Activity_{i,t+1}$: Business activitiy indicator for firm i at year $t+1$.

$employgrowth_{i,t+1}$: $\ln(\text{number of employees at } t+1) - \ln(\text{number of employees at t})$ for firm i

$productivity_{i,t+1}$: $\ln(\text{sales at } t+1) - \left(\frac{2}{3}\right) \times \ln(\text{number of employees at } t+1)$

$$- \left(\frac{1}{3}\right) \times \ln(\text{tangible asset at } t+1) \text{ for firm i}$$

$CAPX_{i,t+1}$: $\frac{(\text{tangible asset at } t+1 - \text{ tangible asset at t} + \text{depreciation at } t+1)}{\text{tangible asset at t}}$ for firm i if negative, adjust to 0

$dcash_{i,t+1}$: $\frac{\text{cash at } t+1 - \text{cash at t}}{\text{total asset at t}}$ for firm i . $ddebt_{i,t+1}$: $\frac{\text{debt at } t+1 - \text{debt at t}}{\text{total asset at t}}$ for firm i

$ROA_{i,t+1}$: $\frac{\text{net income at } t+1}{\text{asset at t}}$ for firm i.

$lnassets_{i,t}$: $\ln(\text{asset})$ at t for firm i, $Leverage_{i,t}$: $\frac{\text{debt at t}}{\text{asset at t}}$ for firm i.

$Tangibility_{i,t}$: $\frac{\text{tangible asset at t}}{\text{total asset at t}}$ for firm i, $lnage_{i,t}$: $\ln(\text{year } t - \text{year of establishement})$ for firm i,

$MTB_{i,t}$: $\frac{\text{market capitalization}}{\text{equity}}$ for firm i

The characteristics of distressed firms are expected to be low employment growth, low productivity, low investment in tangible assets, low cash holdings, and low profitability. Therefore, for the continuous variables such as ICR, ROOA, and RNOA, we expect a positive relationship with employment growth rate, productivity, investment in tangible assets, cash flow, and ROA, as the higher the value of these indicators, the better the firm, and the lower the value, the more vulnerable the firm. It is difficult to conclude the relationship between debt growth and the distress indicators as blue-chip firms may repay their debt, but they may also invest through additional financing. On the other hand, for the binary variables CHK, FN, and ACEE, a value of 1 indicates that the firm is classified as a distressed firm, so a negative relationship is expected with employment growth rate, productivity, investment in tangible assets, cash asset growth, and ROA. In addition, it is difficult to say that the vulnerability of a company is the only reason for the increase in debt, so it is difficult to conclude the relationship between debt growth and the distress indicators. In order to analyze the relevance of the determinants of distressed firms to business activities, a sample of 14,702 firms was selected from the Fnguide that met the following criteria

- KOSPI-KOSDAQ companies excluding financials from 2011 to 2022
- December year-end entities only
- No missing assets, liabilities, or equity in the database
- The accounting equation is complete (Assets = Liabilities + Equity)
- Companies with data for two consecutive years
- All six distressed firm determinants used as independent variables are definable
- Firms with no missing dependent and control variables
- Winsorization at the top and bottom 1% to reduce the impact of extremes

Table 4 shows the descriptive statistics of each variable. In the case of Coverage, the mean is very large compared to the median, indicating that the distribution of this ratio is skewed toward the small end. In addition, the mean values of ROOA and RNOA are around 0.025 and 0.042, respectively, suggesting that the profit margin of operating assets earned by the company is 2.5%, and the profit margin of net operating assets excluding operating liabilities is estimated to be 4.2%. In addition, the mean values of CHK, FN, and ACEE indicators are 0.117, 0.027, and 0.055, respectively, indicating that the largest number of firms, 11.7%, are classified as distressed firms according to CHK, while the share of distressed firms is only about 2.7% according to FN.

Descriptive statistics Table 4

Variable	Name	Obs	Mean	Std	p25	p50	p75
Independent	Coverage	14,702	31.66299	117.7512	0.076134	3.386579	13.65355
	ROOA	14,702	0.024931	0.123968	-0.01691	0.026401	0.077282
	RNOA	14,702	0.041883	0.19652	-0.02256	0.03574	0.107498
	CHK	14,702	0.117467	0.321987	0	0	0
	FN	14,702	0.026595	0.160902	0	0	0
	ACEE	14,702	0.05455	0.227108	0	0	0
Dependent	employgrowth	14,702	-0.00168	0.209894	-0.05506	0.006309	0.071417
	productivity	14,702	9.444969	0.983673	8.806624	9.316189	9.930551
	CAPX	14,702	0.205454	0.553763	0	0.033951	0.162192
	dcash	14,702	0.012324	0.078814	-0.01906	0.003928	0.033645
	ddebt	14,702	0.032608	0.166417	-0.03997	0.010692	0.075299
	ROA	14,702	0.001947	0.111211	-0.02285	0.02058	0.05705
Control	lnassets	14,702	19.34726	1.494858	18.29844	19.06587	20.12398
	Leverage	14,702	0.458306	0.195472	0.307475	0.461298	0.602147
	Tangibility	14,702	0.323537	0.187402	0.180481	0.316713	0.451209
	lnage	14,702	3.200093	0.709745	2.833213	3.258097	3.7612
	MTB	14,702	1.690409	1.868921	0.658098	1.106711	1.960307

Table 5 shows the correlation analysis of each variable. The correlations among the six indicators are inconsistent, suggesting that the determination of distressed firms may vary depending on which distress indicator is used.

Correlation analysis Table 5

	Coverage	**ROOA**	**RNOA**	**CHK**	**FN**	**ACEE**	**employgrowth**	**productivity**
ROOA	0.3539*							
RNOA	0.3506*	0.9414*						
CHK	0.2732*	0.1553*	0.1383*					
FN	-0.0638*	-0.1338*	-0.1251*	0.3126*				
ACEE	-0.0688*	-0.0854*	-0.0907*	0.0649*	0.1688*			
employgrowth	0.0530*	0.2036*	0.1924*	0.0170*	-0.0455*	-0.0352*		
productivity	0.004	0.1420*	0.1453*	-0.001	-0.0388*	0.0199*	-0.0447*	
capx	0.0162*	-0.013	0.006	-0.004	-0.011	-0.0313*	0.0947*	0.0249*
dcash	0.0294*	0.016	0.0198*	0.0172*	-0.008	-0.014	0.0420*	0.010
ddebt	0.016	0.0602*	0.0606*	0.005	-0.0262*	-0.0301*	0.1659*	-0.004
ROA	0.2060*	0.5733*	0.5203*	0.1172*	-0.0784*	-0.0316*	0.1882*	0.1840*
lnassets	-0.013	0.1969*	0.1821*	-0.014	-0.0209*	0.1044*	0.0186*	0.4204*
Leverage	-0.3229*	-0.3039*	-0.2531*	-0.2346*	0.0404*	0.1567*	-0.0870*	0.1321*
Tangibility	-0.0886*	0.005	-0.0462*	-0.0307*	0.0382*	0.0954*	-0.012	-0.2897*
lnage	-0.0298*	-0.0589*	-0.0624*	-0.011	0.012	0.0300*	-0.0468*	0.1316*
MTB	-0.007	-0.1607*	-0.1282*	-0.0325*	0.0222*	-0.0294*	0.0293*	-0.2088*

	capx	**dcash**	**ddebt**	**ROA**	**lnassets**	**Leverage**	**Tangibility**	**lnage**
ROOA								
RNOA								
CHK								
FN								
ACEE								
employgrowth								
productivity								
capx								
dcash	0.0468*							
ddebt	0.3269*	0.2284*						
ROA	-0.0364*	0.0653*	-0.003					
lnassets	-0.1089*	-0.0733*	-0.0437*	0.2210*				
Leverage	-0.0632*	-0.0248*	-0.1513*	-0.1966*	0.2135*			
Tangibility	-0.2942*	-0.0223*	-0.0256*	0.0592*	0.1383*	0.2036*		
lnage	-0.0954*	-0.0467*	-0.0573*	0.0350*	0.2358*	0.0630*	0.0886*	
MTB	0.1206*	0.0999*	0.0450*	-0.1663*	-0.2574*	0.0584*	-0.1301*	-0.1669*

Note: * denotes statistical significance at the 5% level.

As shown in Table 6, the analysis of the relationship between ICR and business activity variables shows that there is a significant positive relationship with employment growth, cash asset growth and profitability, and no significant relationship with other dependent variables. This suggests that the relationship between ICR and business activity is not always consistent.

Distress Indicators and their relation to business activities Table 6

Independent variables	Dependent variables					
	employgrowth	**productivity**	**CAPX**	**dcash**	**ddebt**	**ROA**
Coverage	0.000***	0.000	0.000	0.000**	-0.000	0.000***
	(0.008)	(0.232)	(0.416)	(0.010)	(0.889)	(0.000)
ROOA	0.347***	0.524***	0.215**	0.040***	0.010	0.247***
	(0.000)	(0.000)	(0.019)	(0.001)	(0.669)	(0.000)
RNOA	0.191***	0.273***	0.129**	0.020***	0.003	0.133***
	(0.000)	(0.000)	(0.024)	(0.005)	(0.826)	(0.000)
CHK	-0.014**	0.027*	-0.004	0.007***	-0.013**	0.007***
	(0.036)	(0.062)	(0.798)	(0.007)	(0.012)	(0.007)
FN	-0.046***	-0.033	-0.004	0.000	-0.013	-0.011*
	(0.002)	(0.306)	(0.885)	(1.000)	(0.223)	(0.074)
ACEE	-0.018*	-0.042**	0.014	-0.002	-0.002	-0.006*
	(0.056)	(0.035)	(0.423)	(0.547)	(0.805)	(0.080)

Note: Asterisks, *, ** and ***, denote statistical significance at the 10%, 5%, and 1% levels.

In the case of relationship analysis between ROOA and firm business activity variables, the results show a significant positive relationship with all business activities and no significant relationship with debt growth. These results suggest that ROOA is closely related to business activities, as higher ROOA means higher future employment, investment, productivity, cash asset, and profitability.

As with ROOA, the RNOA analysis shows a significant positive relationship with business activities and no significant relationship with debt growth. It can be seen that RNOA is also closely related to business activities.

Since CHK indicator is classified as a distressed firm when it has a value of 1, a negative relationship is expected with each business activity variable except debt growth. As expected, a significant negative relationship is observed with employment growth but no significant negative relationship is observed with other business activity indicators, and a significant opposite sign is observed with productivity and profitability (ROA). This result suggests that using the CHK indicator as an indicator of distressed firms may not reflect the actual business activity.

In regards to FN indicator, when the FN indicator has a value of 1, it is classified as a distressed firm. Thus, a negative relationship is expected with each business activity variable except for debt growth. As expected, a significant negative relationship was observed with employment growth and ROA, but no significant relationship was observed with other business activity indicators.

The ACEE indicator is also an indicator that is categorized as a distressed firm when it has a value of 1. As expected, there is a significant negative relationship with employment growth, productivity, and ROA, but no significant relationship was observed with other dependent variables. The results of the analysis are more in line with expectations than the CHK and FN indicators in the previous analysis. However, it is still limited in terms of significance to be used as an indicator of distressed firms.

3.6. Comprehensive judgment of each metric

After analyzing the complementary indicators to the ICR based on universality, understandability, affordability, continuity, consistency with the ICR and relevance to the firm business activities, which have been proposed as desirable criteria for identifying distressed firms, we believe that ROOA and RNOA can be used as a strong alternative.

4. Conclusion and Implications

In this study, we examine the indicators for determining distressed firms that can compensate for the shortcomings of the ICR, which is widely used to determine distressed firms but has limitations in terms of universality and continuity. For this purpose, we examined six indicators for determining distressed firms based on prior research, and constructed each indicator for KOSPI and KOSDAQ-listed companies. We defined the attributes of desirable distressed firm indicators as universality, understandability, affordability, continuity, consistency with ICR, and relevance to firm business activities, and judged each indicator based on the above attributes. As a result, we found that ROOA and RNOA are two indicators that have less problems with the universality and continuity of the ICR, but are more consistent with the ICR and have significant relevance to various firm business activities. Therefore, we believe that both ROOA and RNOA can be applied as a substitute for the ICR for distressed firms.

While both ROOA and RNOA are generally appropriate alternatives for ICRs, each has its own limitations.

First, in the case of ROOA, the idealized analysis results, but there is a lack of theoretical basis, as there are no prior studies or textbooks using similar indicators. On the other hand, RONA has the advantage of having a theoretical basis due to the existence of studies and textbooks using similar indicators, but it has the problem of universality that it cannot be applied when net operating assets have a negative value.

Therefore, it is very difficult to define a perfect substitute for the ICR and there is a possibility of inconsistency with existing studies if the standard of distressed firms is suddenly changed. Therefore, it is recommended to consider using indicators such as RNOA and ROOA as verification or supplementary indicators in addition to the existing ICR, rather than changing the standard of distressed firms completely.

A limitation of this study is the restriction of the analysis to listed companies. While this study acknowledges that the reported limitations of the ICR are significant in terms of the universality and attempts to identify alternative indicators to replace it, due to data access limitations, the study is limited to listed companies only and thus does not directly demonstrate the most important issue of the universality of the ICR.

References

Acharya, V.V., Crosignani, M., Eisert, T., and Eufinger, C.(2020). Zombie credit and (dis-)inflation: evidence from Europe. *NBER Working Paper* 27158.

Acharya, V.V., Eisert, T., Eufinger, C., and Hirsch, C.(2019). Whatever It Takes: The Real Effects of Unconventional Monetary Policy, *The Review of Financial Studies* 32, 3366-3411.

Adalet McGowan, M., Andrews, D., and Millot, V.(2018). The walking dead? Zombie firms and productivity performance in OECD countries. *Economic Policy* 33, 685-736

Álvarez, L., García-Posada, M., and Mayordomo, S.(2023). Distressed firms, zombie firms and zombie lending: A taxonomy. *Journal of Banking & Finance* 149, 106762

Bank of Korea (June 2015). *Financial Stability Report*. Bank of Korea.

_____ (2023). *2022 Annual Financial Statement Analysis* Results. Bank of Korea press release (October 25).

_____ (June 2024). *Financial Stability Report*. Bank of Korea.

Caballero, R. J., Hoshi, T., and Kashyap, A. K.(2008). Zombie Lending and Depressed Restructuring in Japan. *The American Economic Review* 98, 1943-1977

Fukuda, S.-i., and Nakamura, J.-i.(2011). Why Did 'Zombie' Firms Recover in Japan? *The World Economy* 34, 1124-1137

Penman, S. H., and Zhang, X.-J.(2002). Accounting Conservatism, the Quality of Earnings, and Stock Returns. *The Accounting Review* 77, 237-264

Penman, S. H.(2010), Financial Statement Analysis and Security Valuation (Fourth Edition),

BANK OF KOREA

A study on Complementary Indicators to the ICR for Identifying Distressed Firms

Hyewon Shin(hw.shin@bok.or.kr)

* The contents of this presentation are the views of the author and do not necessarily reflect the official views of the Bank of Korea.

◆ Table of Contents

Ⅰ. Introduction

Ⅱ. Overview of Distressed Firm Indicators

Ⅲ. Characteristics of Alternative Indicators

Ⅳ. Conclusions and implications

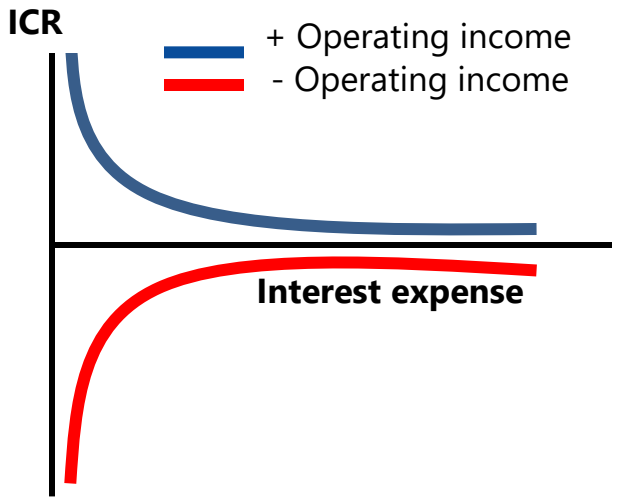
Ⅰ. Introduction : ICR and its limitation

What is ICR (Interest Coverage Ratio)?

- **Definition** : Operating income / interest expense
- **Meaning** : If the ICR is above 1, the company can cover interest expenses and maintain normal operations. Widely used to assess a company's profitability and to identify distressed firms.
- **Limitations** : When interest expense is zero, the ICR becomes undefined. According to the Bank of Korea (2023), 48.6% of 910,206 companies surveyed in 2022 had zero interest expenses, making ICR unusable for them.

Unreliability of ICR

- **When interest expense is small**: A small change in operating income can lead to dramatic changes in the ICR.
- **Problem**: Companies may switch from being classified as distressed to sound (or vice versa) based on minor profitability changes, compromising the stability of the ICR.



Ⅰ. Introduction : Need for an ICR Complementary Proposal

Why We Need an Alternative to ICR?

- **Limitations** : Excluding 48.6% of companies with zero interest expenses from the analysis can lead to distorted policy decisions.
- **Need for a Complementary Indicator**: A stable indicator applicable to all companies, even those with minor changes in profitability

Proposed Alternative Indicator

- **Objective** : To find an indicator that aligns with the ICR's purpose but overcomes its limitations.
- **Proposed Indicators**: ROOA(Return on Operating Assets), RNOA(Return on Net Operating Income)

Ⅱ. Overview of Distressed Firm Indicators

1. Literature Review

❖ First, an individual company is a distressed firm based on the level of its financing interest rate, assuming that a company becomes a distressed firm if it is prolonged by social intervention such as low-interest support from policy funds.

- **Caballero et al. (CHK, 2008)** : If actual interest rate($Rrate_{i,t}$) >minimum required interest rate($Rminrate_{i,t}$), the firm is considered zombie firm because it is being sustained by low-interest loans from financial institutions.

- **Fukuda and Nakamura (FN, 2011)** : Building on the work of Caballero et al. (2008) by excluding zombie firms if a firm's operating income > minimum required interest expenses, and including zombie firm if total borrowing liabilities are more than half of assets and borrowing liabilities are increasing

- **Acharya et al. (ACEE, 2020)** : ⓐ zombie firm if a firm's ICR < the industry-year median and its debt-to-equity ratio > the industry-year median ⓑ the firm's financing rate < financing rate of blue-chip firms like CHK and FN

Ⅱ. Overview of Distressed Firm Indicators

1. Literature Review

❖ Second, low profitability is an indicator of distressed firms, and the methodology focuses on profitability indicators, especially ICR, to identify distressed firms.

- **The Bank of Korea** : distressed firms with an ICR <1 in its Financial Stability Report (2023)
- **Adalet McGowan et al. (2018)** : identified firms with ICR of 1 or less for three consecutive years as zombie firms, citing high comparability as an advantage of focusing on ICR.
- **Alvarez et al. (2023)** : classify a firm as distressed if it has been in existence for more than five years, has an ICR of less than or equal to one, and has three or more consecutive years of capital erosion.

Ⅱ. Overview of Distressed Firm Indicators

Summary

Method	Pros	Cons
Based on Financing Interest Rate (CHK, FN, ACEE)	- Reflects real-world functioning of distressed firms. - Considers market interest rates.	- Requires detailed analysis of each company's data. - Vulnerability is a discrete variable (0 or 1), making continuous assessment difficult.
Based on ICR (ICR)	- Easy to understand and compare across companies. - Practical for real-world application.	- Undefined for companies with zero interest expense. - High volatility if operating income fluctuates, especially for firms with very low interest expenses.

Ⅱ. Overview of Distressed Firm Indicators

2. Measurement of Distressed Firm indicator

❖ **Interest Coverage Ratio (ICR)** : $Coverage_{i,t} = \frac{Operating\ Income_{i,t}}{interexp_{i,t}}$

❖ **Caballero et al.(2008, CHK)** : $Rrate_{i,t} < Rminrate_{i,t}$

$Rminrate_{i,t} = R^{*}_{i,t} / (BS_{i,t-1} + BL_{i,t-1} + Bonds_{i,t-1})$

$Rrate_{i,t} = Interestexpense_{i,t} / (BS_{i,t-1} + BL_{i,t-1} + Bonds_{i,t-1})$

❖ **Fukuda and Nakamura (FN, 2011)** : Based on CHK

(-) Non-distressed : Operating income > minimum required interest expenses

(+) Distressed : Total borrowing liabilities > half of assets / borrowing liabilities are increasing

❖ **Acharya et al.(ACEE, 2020)**

ICR and leverage ratio < industry-year median.

interest expense/borrowing liability (i.e., interest rate) < interest rate of a blue-chip company

interest rate < interest rate of a company(a corporate bond rating of AAA or higher) like CHK

Ⅱ. Overview of Distressed Firm Indicators

2. Measurement of Distressed Firm indicator

❖ **Penman and Zhang(2002) : RNOA, ROOA**

$$RNOA = (Operating\ income\ - interest\ expense)\ /net\ operating\ assets$$

- For this study, Penman and Zhang's (2002) RNOA has been slightly modified to exclude interest expenses paid to creditors from the return calculation. This modification aligns the RNOA with the widely used ICR.

$$ROOA = (Operating\ income\ - interest\ expense)\ /operating\ assets$$

- We additionally define Return on Operating Assets (ROOA) using operating assets rather than net operating assets as the denominator to examine its usefulness as an alternative measure to ICR.

Net operating assets can be measured by reclassifying the statement of financial position as follows

Financial Statements

Operating Assets Financial Assets	Operating Liabilities Financial Liabilities
	Equity

Reclassified Financial Statements

Net Operating Assets (Operating Assets – Operating Liabilities)	Net Financial Obligation (Financial Liabilities – Financial Assets)
	Equity

Operating assets: Total assets - Financial assets
Operating liabilities: Total liabilities - Financial liabilities

Ⅱ. Overview of Distressed Firm Indicators

3. What the indicators for determining a distressed firm should look like?

- We summarize the characteristics of a distressed firm determinant that could serve as an alternative to ICR . We reviewed the characteristics of each judgment indicator considering the above conditions.

Characteristic	Description
1. Universality	Includes all firms, even those with no interest expenses, unlike the interest coverage ratio.
2. Understandability	Simple and easy to understand for both experts and non-experts.
3. Affordability	Quick to compile and release without excessive complexity.
4. Continuity	Shows a clear relationship between indicator changes and firm vulnerability
5. Consistency with ICR	not differ significantly from the traditional ICR
6. Relevant to Business	reflect the company's actual business activities.

Ⅲ. Characteristics of Alternative Indicators

1. Universality

- **Data Source:** FnGuide
- **Scope:** KOSPI-KOSDAQ companies (excluding financials) from 2011 to 2022
- **Conditions:**
  December year-end entities
  Complete accounting equation (Assets = Liabilities + Equity)
  No missing data for assets, liabilities, or equity
  Companies with data for two consecutive years

Indicator	Obs	Percentage
ICR	19,908	95.7%
ROOA	20,813	100%
RNOA	20,510	98.5%
CHK	17,941	86.2%
FN	17,941	86.2%
ACEE	18,655	89.6%
Total	20,813	100%

Ⅲ. Characteristics of Alternative Indicators

2-1. Understandability

- **ICR**: Widely used in financial analysis.
- **ROOA** : Introduced primarily in academic literature and some textbooks.
- **RNOA :** Derived from ROOA
- **CHK, FN, ACEE**: Available only in academic papers

	ICR	ROOA	RNOA	CHK	FN	ACEE
Understandability	High	Medium	Medium	Low	Low	Low

2-2. Affordability

- **ICR**: Created with 27 lines of simple code.
- **ROOA & RNOA**: Require less than 100 lines of code.
- **CHK, FN, ACEE**: Require 200+ lines of code, depending on the indicator.

	ICR	ROOA	RNOA	CHK	FN	ACEE
Length of Code	27	60	92	168	223	129
Affordability	High	Medium	Medium	Low	Low	Low

Ⅲ. Characteristics of Alternative Indicators

3. Continuity

To assess the reliability of indicators, measure the percentage of companies that shifted from the bottom (top) 10% to the top (bottom) 10% the following year. We tracked definable companies over two consecutive years.

- **Indicators with Issues**: CHK, FN, and ACEE cannot determine continuity due to their binary definitions (0 or 1).

- **Extreme Change** : Interest Coverage Ratio (ICR): 0.8%, ROOA and RNOA: 0.5%

Indicators defined as 0 or 1 cannot determine continuity; thus, ROOA and RNOA are considered excellent for continuity.

Ⅲ. Characteristics of Alternative Indicators

4. Consistency with ICR

- **Conditions:** Classify a company as distressed if ICR <1, using only measurable ICR observations. ROOA and RNOA are distressed if <0, while CHK, FN, and ACEE (0,1) are distressed if 1.

- **Results:** Any company that can measure ICR can also measure ROOA. However, for other indicators, there are many that are impossible to measure. If a company is judged as a distressed firm based on the ICR, the indicators of ROOA and RNOA are judged as distressed by more than 98%, while the share of other indicators is less than 10%.

ICR	Obs	Judgment	ROOA	RNOA	CHK	FN	ACEE
Non-distressed	13,450 (100%)	Non-distressed	100.0%	98.81%	76.86%	89.08%	90.23%
		Distressed	0%	0%	12.39%	0.17%	3.49%
		Unclassifiable	0%	1.19%	10.75%	10.75%	6.28%
Distressed (ICR <1)	6,458 (100%)	Non-distressed	0.03%	0.03%	85.06%	82.73%	86.22%
		Distressed	99.97%	98.41%	4.85%	7.17%	7.45%
		Unclassifiable	0%	1.56%	10.10%	10.10%	6.33%

Ⅲ. Characteristics of Alternative Indicators

5. Relevance to business activities

■ Multiple linear regression analysis

$$Activity_{i,t+1} = \beta_0 + \beta_1 distress_{i,t} + \beta_2 lnassets_{i,t} + \beta_3 Leverage_{i,t} + \beta_4 Tangibility_{i,t} + \beta_5 lnage_{i,t} + \beta_6 MTB_{i,t} + firmfixed + yearfixed + \varepsilon_{i,t}$$

- Independent variable : ICR, ROOA, RNOA, CHK, FN, and ACEE
- Dependent variable: employment growth, productivity, investment in tangible asset, cash assets growth, debt growth, and ROA.
- Control variable : firm size, leverage, tangible asset ratio, firm age, and MTB ratio.

Ⅲ. Characteristics of Alternative Indicators

5. Relevance to business activities

- **Results :** For the ICR, the expected sign appears only in some analyses. However, ROOA and RNOA consistently show the expected sign, except in debt growth. For discrete variables CHK, FN, and ACEE, the expected sign is also seen in some analyses.

Distress Indicators and their relation to business activities

Independent variables	Dependent variables					
	employgrowth	**productivity**	**CAPX**	**dcash**	**ddebt**	**ROA**
Coverage	0.000***	0.000	0.000	0.000**	-0.000	0.000***
	(0.008)	(0.232)	(0.416)	(0.010)	(0.889)	(0.000)
ROOA	0.347***	0.524***	0.215**	0.040***	0.010	0.247***
	(0.000)	(0.000)	(0.019)	(0.001)	(0.669)	(0.000)
RNOA	0.191***	0.273***	0.129**	0.020***	0.003	0.133***
	(0.000)	(0.000)	(0.024)	(0.005)	(0.826)	(0.000)
CHK	-0.014**	0.027*	-0.004	0.007***	-0.013**	0.007***
	(0.036)	(0.062)	(0.798)	(0.007)	(0.012)	(0.007)
FN	-0.046***	-0.033	-0.004	0.000	-0.013	-0.011*
	(0.002)	(0.306)	(0.885)	(1.000)	(0.223)	(0.074)
ACEE	-0.018*	-0.042**	0.014	-0.002	-0.002	-0.006*
	(0.056)	(0.035)	(0.423)	(0.547)	(0.805)	(0.080)

Note: Asterisks, *, ** and ***, denote statistical significance at the 10%, 5%, and 1% levels.

Ⅳ. Conclusion and Implications

Comprehensive judgement of indicators

Indicator	Universality	Understandability	Understandability	Continuity	Consistency with ICR	Relevance to Business Activities
Coverage	Medium	High	High	High	High	Medium
ROOA	High	Medium	Medium	High	High	High
RNOA	Medium	Medium	Medium	High	High	High
CHK	Low	Low	Low	Low	Low	Low
FN	Low	Low	Low	Low	Low	Low
ACEE	Low	Low	Low	Low	Low	Medium

- As an complementary indicator for ICR, ROOA appears to be the most desirable indicator, but there is no existing research or textbook that uses this metric.
- On the other hand, RNOA has a theoretical basis, and while it has the problem of being inapplicable when net operating assets are negative, it seems to be a good overall indicator.
- No indicator is perfect, so rather than changing the indicator for distressed firms right away, we're going to consider using indicators such as RNOA or ROOA alongside ICR.

Thank you.

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

The Banco de España's Central Balance Sheet Data Office Database: a regional perspective[1]

Alejandro Fernández Cerezo, Borja Fernández-Rosillo San Isidro and Natividad Pérez Martín,
Bank of Spain

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

The Banco de España´s Central Balance Sheet Data Office Database: A regional perspective

Alejandro Fernández Cerezo
Borja Fernández-Rosillo San Isidro
Natividad Pérez Martín

Abstract

The availability of a firm-level database that is representative of the productive sector of an economy on an aggregate scale is increasingly important to analyse the heterogeneity of different economic variables at different levels of aggregation (for instance, by region, firm size or sector). This paper seeks, first, to evaluate the representativeness of the Banco de España's Integrated Central Balance Sheet Database (Integrated CBSO database or CBI by its Spanish initials) for conducting regional analysis with firm-level data and, second, to analyse the differences in firm size distribution between the Spanish regions.

Keywords: firm data, firm size distribution, financial reporting.

JEL classification: C81, D21, L11, R11.

Contents

Abstract 1

1. Introduction 3

2. The Integrated Central Balance Sheet Database: coverage and representativeness at regional level 5

3. Employment: structure and change over time at sectoral and regional level 9

4. Regional heterogeneity in firm size distribution 13

5. Conclusions 21

References 22

1. Introduction

The availability of a firm-level database that is representative of the productive sector of an economy on an aggregate scale helps analyse the corporate sector's cyclical position and its macroeconomic implications. This is useful when designing public policies that take into account firm heterogeneity in different aspects. A regional analysis using firm-level databases provides an additional layer of detail for a deeper and more accurate understanding of economic developments, particularly in the case of the Spanish economy. First, these sources of firm-level data can be used to explore regional differences across various corporate economic and financial variables, revealing patterns and trends that are not discernible at national level, which can have significant implications for economic policies and regional development strategies. Second, macroeconomic shocks affecting the country as a whole, such as those related to the pandemic (Prades and Tello, 2020) and the recent energy crisis (Fernández Cerezo, Pacce and Sánchez, 2024), may have asymmetric impacts across regions. Also, in countries with a high degree of decentralisation, such as Spain, the regional aspect is especially important, given the existing differences in civil and commercial law and in economic sectors' level of regulation across regions (Mora-Sanguinetti and Soler, 2022).

The Integrated Central Balance Sheet Database (CBI, by its Spanish initials) contains the balance sheets and income statements of a broad range of Spanish non-financial corporations (NFCs) and is therefore a useful source of information for understanding the cyclical position and economic and financial performance of Spain's business sector (Banco de España, 2022a). In particular, it can be used to analyse various aspects, such as firm size, sector of activity and region where the registered office is located.

To provide the general public with easy access to CBI regional data, the Banco de España's Central Balance Sheet Data Office (CBSO) has prepared a report called "BExplore Regional Central Balance Sheet Data Office" which provides information on variables regarding the activity, employment, profitability and financial structure of Spanish NFCs, broken down by sector, size and region or province (see Figure 1 for a summary of its content).[1] To this end, information from the annual accounts deposit provided by the Spanish Mercantile Registries and other detailed information provided by CBSO reporting firms was used.[2]

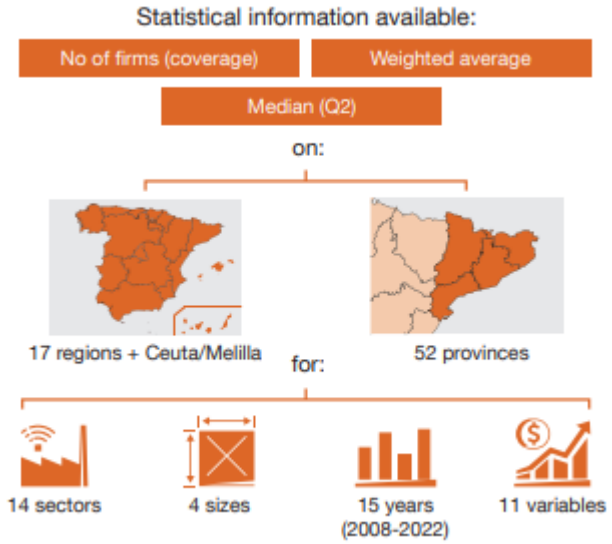
This paper seeks, first, to evaluate the coverage and representativeness, at regional level, of the CBI sample of NFCs and, second, to analyse the cross-regional differences in firm size distribution in Spain. Thus, this paper's main contribution consists of documenting the representativeness of the CBI at regional level. To this end, this paper extends the analysis in Almunia López-Rodríguez and Moral-Benito (2018) to regional level. Among other things, the aforementioned paper showed the

[1] Although the cross-checking of information has been limited to a minimum number of firms and quality control is exhaustive for the purpose of eliminating outliers, the use of the median value (Q2) as an analysis statistic is recommended to compare firms by sector, region and size in the BExplore Regional Central Balance Sheet Data Office tool.

[2] Since 2019 the research community may access CBI firm-level data through the Banco de España's data laboratory (BELab).

CBI's capacity to replicate, at national level, the growth rate of employment in the Spanish private sector and firm size distribution.

Figure 1. Information available in the report



SOURCE: Banco de España

The results show that the CBI coverage rates (in terms of number of firms) are high in all the regions and that the CBI adequately captures the differences in sectoral structure, employment dynamics and firm size distribution observed in other official sources. As regards firm size, analysis of the CBI's firm-level data shows that there are significant cross-regional differences, even within each sector of activity. This is important, for instance, to understand the factors behind the small size of Spanish enterprises. In this connection, exploiting the CBI's regional dimension allows analysts and academics to address research and economic policy issues of interest for the Spanish economy.

The rest of this paper is structured as follows. Section 2 presents the CBI database and its contents, analysing coverage with respect to the population of firms in each region. Section 3 analyses the capacity of the database to replicate the sectoral distribution of employment across the different regions and changes therein over time. Section 4 analyses cross-regional differences in firm size distribution. Section 5 concludes the paper by exploring some possible future avenues for analysis.

2. The Integrated Central Balance Sheet Database: coverage and representativeness at regional level

The CBSO is a service that analyses Spanish non-financial corporations' and corporate groups' economic and financial data. It mainly draws on two sources of yearly information. First, the information obtained directly from firms which voluntarily furnish their data in a specific questionnaire devised by the CBSO (the CBSO Annual Survey or CBA by its Spanish initials).[3] Second, that arising from the cooperation with the Mercantile Registries,[4] which send to the CBSO the annual accounts that firms are obliged to file with them in accordance with financial reporting requirements (the CBB database).[5] Since 1990 firms are required to file their annual accounts with the Mercantile Registry office of the province where their registered office is located.[6] These offices compile and process all the information that is publicly available at the provincial Mercantile Registries and which they ultimately send to the Association of Registrars' Statistical Processes Centre (CPE by its Spanish initials).[7] This is administrative information that is highly reliable, since firms are required by law to provide accurate information about their financial position. In addition, larger firms must undergo an independent review of their annual accounts by a statutory auditor. The Mercantile Registry periodically furnishes the Banco de España with the digitalised raw data (i.e. without filtering out possible accounting inconsistencies or changes in the time series) from the financial statements filed by the firms.

The CBI is constructed through a combination of the two data sources, providing a representative sample of the population of individual NFCs in the Spanish economy, obtained from nearly one million financial statements that are processed yearly. The combined processing of the sources of information on individual firms in the CBI expands the analytical capacity of the database by combining the data available for large firms, which are well represented in the CBA, with those for small and medium-sized enterprises, which are included more comprehensively in the CBB database.

Firms' economic and financial data are processed by the CBSO, harmonising the accounting items over time and cleansing the data provided by the Mercantile

[3] The cooperation with the CBSO consists of annually filling in a questionnaire based on the accounting plan, the content and details of which depend on the firm's size. In 2023 (the latest available year) around 10,000 firms cooperated voluntarily with the CBSO. https://www.bde.es/wbe/en/areas-actuacion/central-balances/servicios-para-empresas/quien-puede-colaborar/

[4] See the agreement which defines the cooperation conditions between the Banco de España and the Spanish Mercantile and Property Registrars' Association to support the filing of accounts: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2023-11942

[5] With regard to the filing of accounts, in June 1991 the Banco de España entered into an agreement with the Ministry of Justice for the Madrid Mercantile Registry. The Barcelona, Girona and Tarragona Registries followed suit over the succeeding years. Currently all the provincial Mercantile Registries collaborate with the Banco de España.

[6] There are Mercantile Registries in all the provincial capitals, as well as in Ceuta, Melilla, Ibiza, Mahón, Santiago de Compostela, Puerto de Arrecife, Puerto del Rosario, Santa Cruz de la Palma, San Sebastián de la Gomera and Valverde.

[7] https://www.registradores.org/registroonline/tramites/mercantil/estadisticas.seam

Registry to obtain a comparable, consistent and quality database. These data cleansing tasks include, for instance, excluding firms whose accounting year closing date has changed, those providing incomplete information, thus preventing assessment of the units used to prepare the annual accounts (e.g. euro, thousand or million euro)[8] and firms reporting excessive changes in certain items without justifying them. Business combinations between firms, intra-group transactions and the distribution of dividends are appropriately processed to correctly reflect the flows from such operations (for a more detailed analysis, see Banco de España, 2022b).

Table 1 compares, for the 2009-2021 period average,[9] the number of firms comprising the population of NFCs in Spain according to the Central Business Register[10] (column 1), the number of firms whose data are received and processed by the CBSO (column 3) and the number of firms that are suitable for analysis following the data cleansing process mentioned above (column 5). Of the more than 1.25 million firms in the Central Business Register population, the CBSO receives information on almost one million firms, but only around 840,000 are deemed suitable for analysis.[11]

It is noteworthy that the regional distribution of the number of firms fit for analysis in the CBI adequately replicates that of the Central Business Register population, even in the smallest regions. For instance, the number of firms in Rioja, Cantabria and Navarre account for 0.6%, 0.9% and 1.2% of the total population, respectively, (column 2 of Table 1) vs. 0.7%, 0.9% and 1.3% in the CBI (column 6). Among the larger regions, Madrid has a somewhat smaller share in the CBI than in the Central Business Register (18.3% vs. 20.2%), while the opposite is true in Catalonia (19.7% in the CBI vs. 18.8% in the Central Business Register).

[8] For example, firms that fail to report items in the balance sheet or the income statement that are necessary to ensure they are balanced are excluded.

[9] 2008 and 2022 are excluded from the analysis owing to the CBI's lower coverage in those years. Since 2008 was the first year with financial reporting under the new 2007 General Accounting Plan (PGC by its Spanish initials), firms were allowed to provide information relating only to the then current year (2008) and not to 2007. Since the CBSO requires data for two years (current and previous) to cleanse the economic and financial information of each firm, all the firms for which there were no data for 2007 were considered unfit to be included in the CBI database. Also, the information available for 2022 is incomplete, as it is a preliminary sample to be increased in subsequent updates, as a result of the inclusion of new firms. As regards the years prior to 2008, the CBSO has data on 250,000 to 550,000 firms per year from 2000 to 2007. Given the lower coverage for those years, the related historical information will gradually be included in the BExplore Regional Central Balance Sheet Data Office tool using statistical representativeness studies.

[10] Official data on the number of firms in Spain come from the Central Business Register, which is prepared by the National Statistics Institute (INE), which in turn feeds Eurostat's structural business statistics. Included are firms whose legal form falls within the institutional aggregate of NFCs (the CBSO's field of study), mainly public limited companies, private limited companies, general partnerships, limited partnerships and certain civil-law partnerships. Individuals, jointly held property, associations and non-profit private institutions are excluded.

[11] In other words, 16% of firms are rejected on average each year in the period analysed because they do not meet the quality standards set by the CBSO.

Table 1. Number of firms in the population (Central Business Register) and in the Central Balance Sheet Data Office. Average for the period 2009-2021, by region (a)

	Population of firms, according to the Central Business Register		Firms received in CBB (not adjusted for quality) and CBA		Firms suitable for analysis in the CBI		
	Number of firms	% of national total	Number of firms	% of total firms received	Number of firms	% of total suitable firms	Coverage rate
Region	1	2	3	4	5	6	7 (5 / 1)
Andalusia	176,041	14.0	133,981	13.4	110,964	13.3	63.0
Aragon	33,341	2.7	29,733	3.0	25,544	3.1	76.6
Asturias	19,964	1.6	17,303	1.7	14,967	1.8	75.0
Balearic Islands	36,188	2.9	32,020	3.2	26,478	3.2	73.2
Canary Islands	52,143	4.1	32,654	3.3	27,540	3.3	52.8
Cantabria	11,280	0.9	8,628	0.9	7,421	0.9	65.8
Castile-Leon	53,297	4.2	43,978	4.4	37,489	4.5	70.3
Castile-La Mancha	46,706	3.7	38,398	3.8	32,799	3.9	70.2
Catalonia	235,921	18.8	196,928	19.7	164,825	19.7	69.9
Valencia	139,601	11.1	114,250	11.4	97,495	11.6	69.8
Extremadura	18,308	1.5	14,649	1.5	12,336	1.5	67.4
Galicia	68,473	5.4	55,226	5.5	47,801	5.7	69.8
Madrid	254,415	20.2	187,661	18.8	152,894	18.3	60.1
Murcia	35,859	2.9	28,410	2.8	23,946	2.9	66.8
Navarre	15,615	1.2	12,724	1.3	10,681	1.3	68.4
Basque Country	50,946	4.0	44,648	4.5	38,141	4.6	74.9
Rioja	7,571	0.6	6,679	0.7	5,726	0.7	75.6
Total	1,257,924	100.0	997,870	100.0	837,046	100.0	66.5

SOURCES: Banco de España and INE

(a) Including firms whose legal form lies within the scope of analysis of the Central Balance Sheet Data Office. Excluding individuals, jointly held property and associations. Including firms with zero or more employees.

Another important aspect for demonstrating the CBI's contents is to analyse the coverage rate, defined as the ratio of the number of NFCs fit for analysis in the CBI to the number included in the Central Business Register (column 7). The coverage rate at national level is 66.5% on average for the period 2009-2021, exceeding 70% in Aragon, the Balearic Islands, Rioja, the Basque Country and Asturias, and is somewhat lower in the Canary Islands and Madrid (52.8% and 60.1%, respectively). These cross-regional differences in coverage are mainly due to provincial differences in the reception, processing and digitalisation of the annual accounts filed with the Mercantile Registries. As a result, the information available in the Association of Registrars' CPE (which the CBSO draws from) changes by region. In addition to the percentage of firms that continue to file in print (approximately 8% of the total for Spain, with some regional differences), some annual accounts are filed late, whether in breach of the law or because the firm was unable to approve its accounts at the Annual General Meeting (a shortcoming that is usually remedied during a subsequent filing).[12]

Table 2 analyses the changes over time in the CBI's coverage rates. An upward trend is observed over the period analysed across most regions. This is the result of, first, an ongoing increase in the number of firms filing their annual accounts with the Mercantile Registries, as well as the greater use of the XBRL electronic format (reducing the filing in print) and, second, an improvement in the CPE's

[12] Other factors causing regional coverage differences are: enterprises accounted for as population but which do not file their accounts with the Registries, such as civil-law partnerships, firms that are registered but which are dormant and firms that do not meet the information quality criteria and, consequently, are not included in the CBSO publications.

information receipt processes, granting the CBSO greater access to the Mercantile Registry databases, thanks to the cooperation agreements entered into by the two institutions.

Table 2. Coverage rate (number of NFCs included in the CBI relative to the Central Business Register), by region (a)

Region	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Average (2009-2021)
Andalusia	51.6	52.1	54.9	58.3	60.9	64.1	65.1	64.3	61.8	59.2	65.6	62.2	61.9	63.0
Aragon	67.9	72.4	73.4	73.3	73.6	73.5	74.9	74.9	71.3	72.4	78.1	78.1	78.8	76.6
Asturias	69.3	67.7	70.8	71.8	71.8	73.7	74.3	73.0	71.5	70.6	76.3	74.1	75.0	75.0
Balearic Islands	65.0	67.7	73.1	72.5	70.9	75.0	75.1	74.5	61.5	66.7	73.8	72.9	70.7	73.2
Canary Islands	50.9	50.0	51.4	50.5	51.9	53.2	55.1	52.6	48.1	40.8	47.4	56.0	44.4	52.8
Cantabria	57.1	52.6	52.7	55.4	59.6	66.2	73.5	71.5	63.7	63.3	69.1	68.4	73.1	65.8
Castile-Leon	64.7	64.7	65.1	67.2	67.4	68.9	70.4	70.8	63.9	66.0	70.1	70.8	71.6	70.3
Castile-La Mancha	61.6	61.6	64.6	65.9	67.1	69.6	71.6	70.6	66.9	66.9	72.4	73.0	66.8	70.2
Catalonia	61.3	61.6	63.4	65.8	70.2	71.1	69.7	68.8	64.6	63.5	67.6	67.6	68.4	69.9
Valencia	62.2	61.8	63.6	65.1	66.5	69.3	69.3	70.1	67.6	66.8	71.6	70.7	72.5	69.8
Extremadura	62.8	62.8	64.7	60.6	60.9	67.3	68.4	69.3	60.1	60.5	62.5	66.7	69.4	67.4
Galicia	63.3	63.1	66.7	66.7	67.1	67.1	68.5	70.2	63.7	64.9	72.1	72.8	73.7	69.8
Madrid	50.7	51.8	53.5	55.0	56.5	58.6	58.3	58.1	58.4	60.6	63.1	62.0	61.5	60.1
Murcia	57.3	58.1	59.9	61.8	62.9	63.8	65.6	66.3	61.4	61.8	68.9	69.9	70.6	66.8
Navarre	57.6	57.7	59.6	61.2	66.3	67.5	69.6	67.7	66.0	67.9	72.3	69.3	67.8	68.4
Basque Country	58.3	56.3	62.3	67.0	72.1	76.2	73.9	76.8	76.1	77.0	79.0	78.7	75.8	74.9
Rioja	65.8	64.7	69.0	71.4	72.3	73.5	73.2	74.7	72.9	77.0	80.0	79.6	77.7	75.6
Total	57.9	58.2	60.5	62.2	64.2	66.3	66.5	66.3	63.1	63.0	67.6	67.2	66.8	66.5

SOURCES: Banco de España and INE

(a) Including firms whose legal form lies within the scope of analysis of the Central Balance Sheet Data Office. Excluding individuals, jointly held property and associations. Including firms with zero or more employees.

A very important aspect when analysing the CBI's firm-level data from a regional viewpoint is that each firm is located in the province where it is headquartered, regardless of whether it has production sites in other areas. This is because there is only one balance sheet, which cannot therefore be fractioned on the basis of the percentage of sales or employment by province or region. This aspect is vitally important for any comparison of the CBI results at regional level with other sources of aggregate information.

It is worth asking whether allocating the firms' annual accounts in full to the region where they are headquartered skews the CBI's regional analysis. In order to approximate the geographical location of their activities, the survey for CBA reporting firms includes a question about the regional distribution of personnel costs.[13] In particular, a sub-sample of around 2,800 large firms report the percentage of personnel costs incurred in each region. Based on this information it can be deduced that firms being located in several regions could have a limited impact on

[13] This information is not available for CBSO users for confidentiality reasons.

the CBI's regional representativeness, since 87% of the firms in this sub-sample incur over 90% of their personnel costs in the same region where their headquarters are located. This means that the activity of the majority of firms that provide detailed information of their labour costs is concentrated in a single region. Also, only 5% of them have more than 50% of their personnel costs in regions other than where they are headquartered.

In sum, the CBI sample available is reasonably representative of the population of NFCs at regional level.

3. Employment: structure and change over time at sectoral and regional level

This section compares the sectoral distribution of the CBI by region with that from official sources, in order to assess the CBI's ability to capture cross-regional differences in productive systems. We are thus able to examine whether the composition of the CBI is skewed towards a certain set of firms belonging to one sector. This is of particular interest, for example, for any research aiming to take into account the asymmetric effects of economic shocks on sectors and regions.

The analysis focuses on the non-financial and non-agricultural market economy, i.e. we disregard firms operating in the following sectors: i) public administration and defence; compulsory social security, education, health and social work; ii) financial and insurance activities; iii) agriculture, forestry and fishing; and iv) activities of households as employers and producers and activities of membership organisations.[14]

To assess the representativeness of the CBI when capturing regional differences in sectoral specialisation, we use as reference the sectoral distribution of the employees registered under the General Social Security Regime.[15] It should be noted that, to calculate the number of employees in each sector and region, Social Security uses a slightly different method from that used in financial statements. The former provides the average number of workers each month, unadjusted for temporary and part-time employees. By contrast, the latter report the average

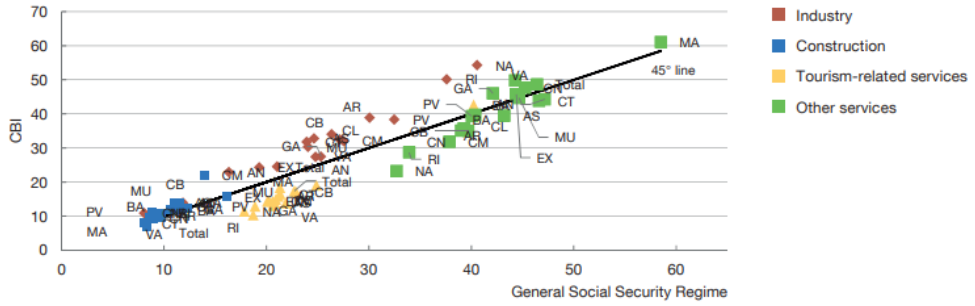
[14] Observations with no employees are removed because these firms mostly represent dormant firms or firms incorporated for tax purposes. The self-employed do not fall within the CBI's scope.

[15] The General Social Security Regime includes firms' employees under that regime. Therefore, those registered under the Special Social Security Regime for the Self-Employed are excluded, as the CBI does not include natural persons. https://w6.seg-social.es/PXWeb/pxweb/es/Afiliados%20en%20alta%20laboral/Afiliados%20en%20alta%20laboral__Afiliados%20Medios/04mb.%20Por%20CC.AA,%20%20seccion%20de%20actividad%20y%20regimen%20(RG%20y%20RETA).px/

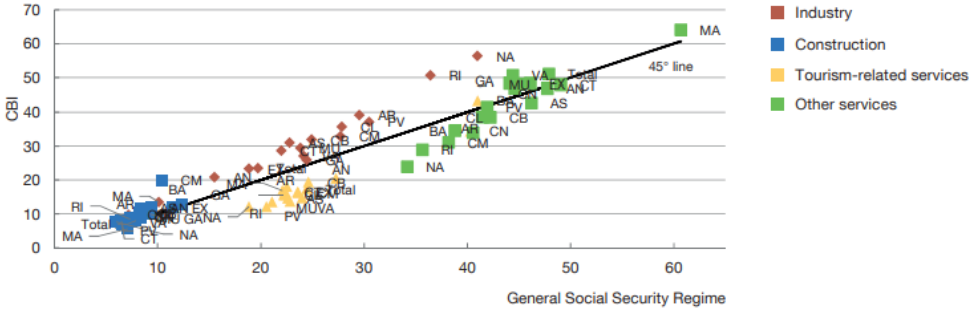
number of full-time equivalents throughout the year, considering temporary employees as a fraction of full-time employees.[16]

Chart 1. Sectoral distribution of employment, by region (% of total employment in each region) (a)

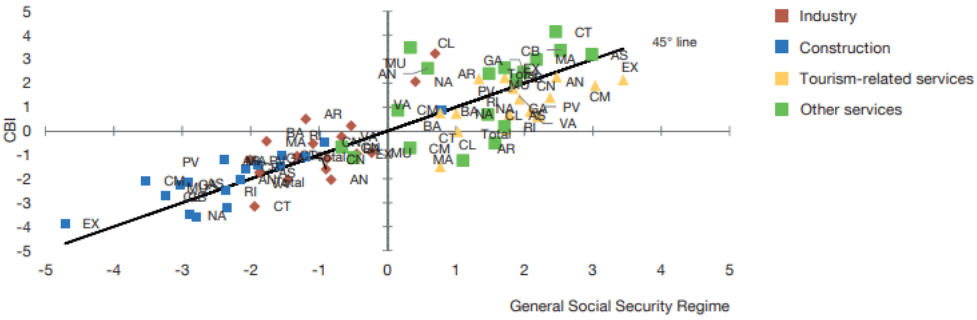
1.a Average for the period 2009-2014 (%)



1.b Average for the period 2015-2021 (%)



1.c Change between the averages for the periods 2009-2014 and 2015-2021



SOURCES: Banco de España

(a) "Industry" includes the firms under NACE Rev. 2 sections B (mining and quarrying), C (manufacturing), D (electricity, gas, steam and air-conditioning supply) and E (water supply, sewerage, waste management and remediation). "Construction" includes firms in sector F (construction). "Tourism-related services" includes firms

[16] The CBI employment figure reflects the hires and terminations of employees reported throughout the year, weighted by the months they worked at the firm. To convert part-time, temporary and permanent seasonal employment into full-time equivalent employment, the hours worked by all employees with these types of contracts are added together and then divided by 1,826 hours (the assumed number of hours that a full-time employee works per year).

in sectors H (transportation), I (accommodation and food service activities) and R (arts, entertainment and recreation). "Other services" includes firms in sectors G (wholesale and retail trade), J (information and communication), M (professional, scientific and technical activities) and N (administrative and support service activities). AN: Andalusia, AR: Aragon, AS: Asturias, BA: Balearic Islands, CN: Canary Islands, CB: Cantabria, CL: Castile-Leon, CM: Castile-La Mancha, CT: Catalonia, VA: Valencia, EX: Extremadura, GA: Galicia, MA: Madrid, MU: Murcia, NA: Navarre, PV: Basque Country, RI: Rioja.

Since the sample period (2009-2021) includes a recession and a recovery, it has been divided in two: Chart 1.a depicts the average share of employment according to the CBI and Social Security in four major sectors[17] and by region for the period 2009-2014, while Chart 1.b does the same for the period 2015-2021. The results of this comparison show that the CBI adequately reflects the differences in the regional distribution of employment observed in the social security registrations data. With regard to the services sector, the Balearic Islands and the Canary Islands have the highest share of employment in tourism-related services according to the CBI (around 42% and 32%, respectively). This is mirrored in the social security registrations data (around 40% and 37%, respectively). Meanwhile, Madrid is the region with the highest share of employment in other services (61% in the CBI and 59% in the social security data). Turning to the secondary sector, Navarre and Rioja are the two regions with the highest share of employment in industry in the two databases. However, industrial employment is overrepresented in the CBI in all regions. Castile-La Mancha and Extremadura are the regions with the highest share of construction employment in both the CBI and the social security registrations data.

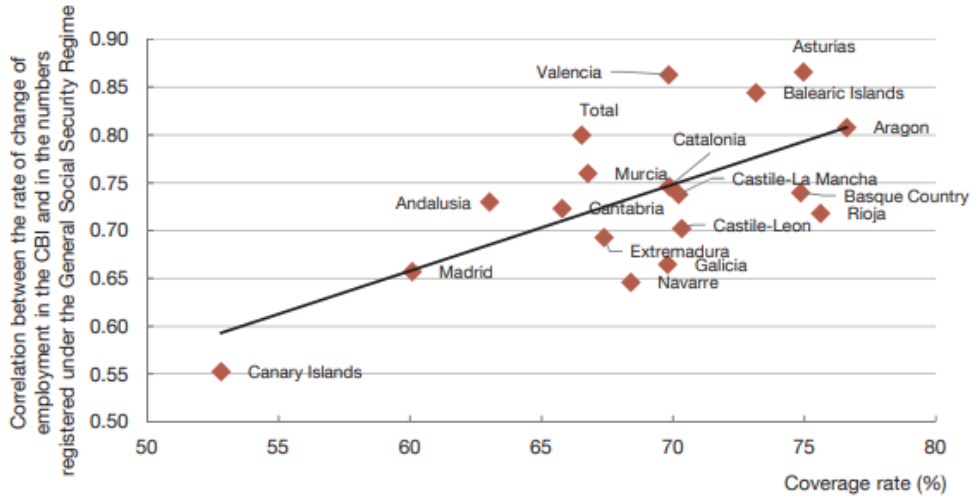
The CBI also adequately captures the change in the sectoral composition of employment that took place in the two periods under review. At national level, the shift to a services economy stepped up, with the percentage of industrial and construction employment falling by 1 pp and 1.4 pp, respectively, and that of services employment increasing by 2.4 pp, according to the CBI sample (see Chart 1.c). These figures are similar to those obtained from the social security data, which show a fall in employment in industry and construction of 1.2 pp and 1.8 pp, respectively, and a 3 pp rise in the share of services employment. By region, the CBI also performs notably when reflecting this sectoral shift. For example, according to the social security registrations data, the share of industrial employment decreased in all regions apart from Navarre, Castile-Leon and Castile-La Mancha. This is also reflected in the CBI data. Extremadura saw the largest decline in the share of construction employment in both the CBI and the social security registrations data (3.9 pp and 4.7 pp, respectively). The increase in the share of tourism-related services employment in the Canary Islands (2.3 pp) is notable. The official data yield a similar figure (2.5 pp). Catalonia, Asturias and Cantabria are the three regions where the share of

[17] "Industry" includes the firms under NACE Rev. 2 sections B (mining and quarrying), C (manufacturing), D (electricity, gas, steam and air-conditioning supply) and E (water supply, sewerage, waste management and remediation). "Construction" includes firms in section F (construction). "Tourism-related services" includes those firms in sections H (transportation and storage), I (accommodation and food service activities) and R (arts, entertainment and recreation). "Other services" includes firms in sections G (wholesale and retail trade), J (information and communication), M (professional, scientific and technical activities) and N (administrative and support service activities).

employment in other services has gained the most both in the CBI and in the social security data.

Lastly, we assessed the CBI's ability to replicate the dynamics of aggregate employment at regional level throughout the period 2009-2021. By focusing on dynamics rather than levels, this exercise is particularly relevant for any research with a macroeconomic perspective. The CBI is highly accurate at reproducing the dynamics of overall employment in Spain in the period 2000-2013 (Almunia, López-Rodríguez and Moral-Benito, 2018). The CBI also proves to be highly accurate when performing this analysis for the reference period of this paper (2009-2021): at national level, the correlation between the annual rate of change in total employment in the CBI and that of the General Social Security Regime data is 0.82.[18] When performing the same exercise for total employment in each region, the correlation between the CBI and the social security time series is 0.72 on average for the 17 regions, and tends to be higher in those regions where the CBI has a higher coverage rate (such as Asturias, Aragon and the Balearic Islands) and lower where its coverage rate is lower (e.g. the Canary Islands) (see Chart 2).

Chart 2. CBI coverage rates and correlation of employment dynamics between the CBI and registration under the General Social Security Regime (a)



SOURCES: Banco de España, INE and Ministerio de Inclusión, Seguridad Social y Migraciones

(a) Drawing on annual employment data for the period 2009-2021. Including non-financial and non-agricultural market economy firms, excluding firms without any employees. The coverage rate is defined as the ratio between the number of NFCs suitable for analysis in the CBI and the number of firms in the Central Business Register

[18] In Almunia, López-Rodríguez and Moral-Benito (2018), the correlation between total employment in the CBI firm-level sample and employment in the National Accounts is 0.91 for the period 2000-2013.

In sum, broadly speaking, the available CBI sample adequately replicates the sectoral distribution of employment in the different regions and the changes therein over time in aggregate terms.

4. Regional heterogeneity in firm size distribution

One of the Spanish business sector's main characteristics is the predominance of smaller enterprises, which is key to understanding Spain's low aggregate productivity (Moral-Benito, 2018).[19] It is also at smaller enterprises where a wider negative productivity gap vis-à-vis other European countries is observed.[20]

This section analyses the existing cross-regional differences in firm size distribution. First, we show that the CBI adequately replicates the distribution of firm size and employment at national level. Second, we document the distribution of the number of firms and employment in the CBI by size group and region, with particular focus on the enterprises with fewer than 50 employees. Lastly, we illustrate the changes over recent years in the distribution of employment in the CBI, by firm size and region.

Prior to analysing the regional differences in firm size, we should examine whether the CBI sample adequately reproduces the distribution of the number of firms and employment by firm size group reported in the official statistics at national level. Specifically, in terms of the number of firms, we use as reference the distribution of active firms by number of employees and legal form[21] reported by the Central Business Register,[22] while for employment we use the Statistics for Social Security-registered Firms *(Estadística de Empresas Inscritas en la Seguridad Social)*. As in the previous section, the analysis focuses on the non-financial and non-agricultural market economy, and excludes firms with zero employees.

Chart 3.a shows that the distribution of the CBI sample of firms is considerably similar to that of the population in the period 2009-2021. Specifically, large firms (250 or more employees) and medium-sized enterprises (50-249 employees) accounted for 2% and 0.5%, respectively, of the total sample in both databases. However, microenterprises (1-9 employees) are slightly underrepresented in the CBI (83.3% of firms in this database, versus 84.9% in the Central Business Register), while small enterprises (10-49 employees) are slightly overrepresented (14.2% of firms in the CBI, versus 12.6% in the Central Business Register). In terms of the distribution of employment, Chart 3.b shows that the average percentage of workers at large firms in the period 2013-2021 is higher in

[19] When discussing firm size and its implications, the evidence suggests that the causal direction is productivity to size, not size to productivity. In other words, the most efficient and productive firms tend to gain market share and grow, while the least productive tend to lose market share.

[20] Banco de España (2024) and Gavilán and García Coto (2024).

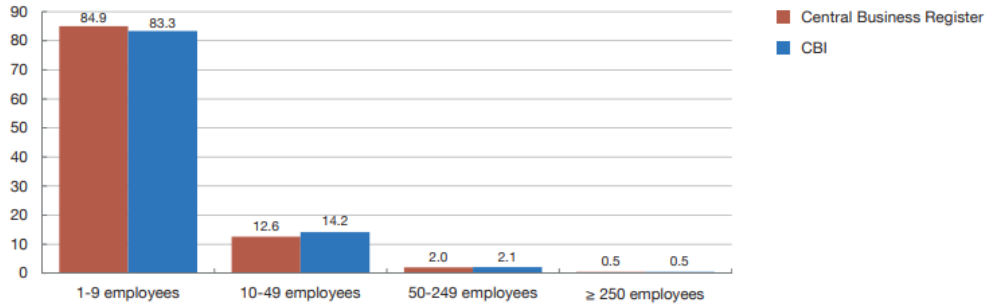
[21] We analyse firms incorporated as legal persons in the form of public limited companies, private limited companies and cooperatives.

[22] To define the size groups, the Central Business Register considers the total number of employees at year-end, without adjusting for temporary and part-time employees.

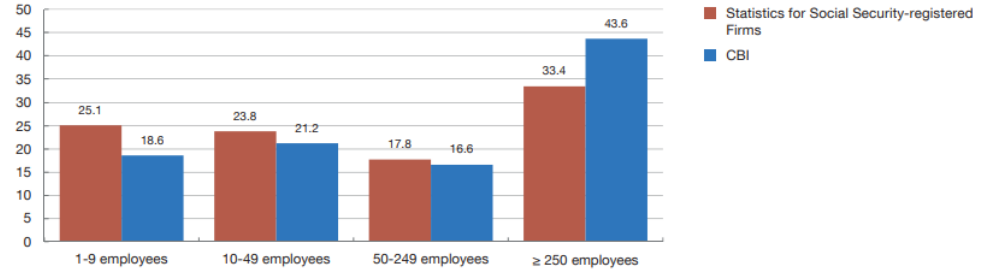
the CBI (43.6%) than in the social security data (33.4%).[23] By contrast, in the CBI the percentage of employment in the other size categories is lower than in the social security data, especially in the microenterprise segment (18.6% in the CBI, versus 25.1% in the social security data). These findings are in line with Almunia, López-Rodríguez and Moral-Benito (2018), who show that the left tail of the distribution of firms has a lower coverage rate in the CBI, due to poorer data quality in this size category, which stems, for example, from a higher relative concentration of missing data in the employment variable.

Chart 3. Distribution of the number of firms at national level, by firm size (a)

3.a Percentage of firms, by firm size. Average for the period 2009-2021



3.b Percentage of employment, by firm size. Average for the period 2013-2021



SOURCES: Banco de España, INE and Ministerio de Inclusión, Seguridad Social y Migraciones

(a) Including non-financial and non-agricultural market economy firms, excluding firms without any employees

In light of the CBI's representativeness when reflecting the national firm size distribution, next we examine the same distribution in the CBI, but at regional level. In terms of the number of firms, a CBI-Central Business Register comparison of the regional distribution of this variable by firm size is not possible, because the Central Business Register does not provide a breakdown of the population of firms by number of employees, legal form and region. In any event, Table 3.a highlights some cross-regional differences in firm size distribution. First, the Basque Country and Rioja have the lowest percentage of microenterprises (under 80%), while Castile-La Mancha, Extremadura and Castile-Leon have the highest proportion of these firms (over 85%). Second, Madrid stands out as the region with the highest percentage of

[23] The first year available with information broken down by sector and size in the Statistics for Social Security-registered Firms is 2013.

large firms (1.1%), which could be related to the "headquarters effect", which consists of firms – especially the largest ones – tending to locate their headquarters in major cities in order to harness the wealth of highly specialised professional profiles and the concentration of certain services.[24]

Table 3. Distribution of the number of firms and employment in the CBI, by firm size (a)

3.a Distribution of the number of firms, by firm size

Region	1-9 employees	10-49 employees	50-249 employees	≥ 250 employees
Andalusia	84.8	13.4	1.6	0.3
Aragon	82.9	14.8	2.0	0.4
Asturias	83.5	14.3	1.9	0.4
Balearic Islands	83.3	14.8	1.6	0.3
Canary Islands	80.3	16.6	2.7	0.4
Cantabria	80.5	16.8	2.2	0.5
Castile-Leon	85.3	13.1	1.4	0.2
Castile-La Mancha	86.0	12.5	1.3	0.2
Catalonia	82.5	14.6	2.4	0.5
Valencia	84.3	13.6	1.8	0.3
Extremadura	85.5	13.0	1.4	0.2
Galicia	85.2	12.8	1.7	0.3
Madrid	82.6	13.7	2.7	1.1
Murcia	81.9	15.6	2.2	0.4
Navarre	80.5	15.9	3.0	0.6
Basque Country	79.4	17.3	2.7	0.6
Rioja	79.7	17.9	2.2	0.2
Total	83.3	14.2	2.1	0.5

3.b Distribution of employment, by firm size

Region	1-9 employees	10-49 employees	50-249 employees	≥ 250 employees
Andalusia	30.3	31.3	18.3	20.2
Aragon	24.9	27.9	18.6	28.6
Asturias	23.7	25.3	17.3	33.8
Balearic Islands	25.5	29.3	15.4	29.7
Canary Islands	21.9	29.3	24.3	24.5
Cantabria	23.6	30.9	20.7	24.7
Castile-Leon	31.3	30.1	16.2	22.4
Castile-La Mancha	34.5	31.9	17.2	16.3
Catalonia	19.7	23.9	19.5	36.9
Valencia	24.5	26.1	17.6	31.8
Extremadura	36.8	32.8	17.2	13.1
Galicia	25.1	24.5	16.2	34.3
Madrid	8.7	10.5	10.8	69.9
Murcia	23.0	28.6	20.5	27.9
Navarre	19.6	26.2	23.4	30.9
Basque Country	17.8	25.0	19.6	37.6
Rioja	27.6	38.7	22.6	11.2
Total	19.0	21.6	16.1	43.3

SOURCE: Banco de España

(a) Average for the period 2009-2021. Including non-financial and non-agricultural market economy firms, excluding firms without any employees

[24] However, when considering firms overall, the coverage rate of the CBI in Madrid is lower than the national average (see Tables 1 and 2), suggesting that this region's lower representation is concentrated in the small enterprise category.

Turning to employment, Table 3.b shows the distribution of the number of employees by firm size according to the CBI. As was the case when analysing the number of firms, the regional distribution of employment by firm size cannot be compared with the Statistics for Social Security-registered Firms, as they do not provide information broken down by size or region. At national level, while firms with fewer than 50 employees accounted for 97.4% of the total on average in the period 2009-2021, they only made up 40.6% of employment.[25] Meanwhile, large firms accounted for just 0.5% of the total, but for 43.3% of employment. However, there are considerable cross-regional differences. First, the share of employment at large firms (250 or more employees) is almost 70% in Madrid, virtually doubling the proportion in the Basque Country (37.6%) and Catalonia (36.9%). Second, at the other end of the scale, in Castile-La Mancha and Extremadura microenterprises (1-9 employees) account for 34.5% and 36.8% of employment, respectively. Turning to the intermediate size categories, Rioja is the region with the most employment at small enterprises (38.7%), while the Canary Islands and Navarre are the regions with a higher share of employment at medium-sized enterprises (24.3% and 23.4%, respectively).

Against this backdrop, the CBI is a unique source in Spain for analysing regional heterogeneity in firm size distribution since other sources, such as the Central Business Register and the Statistics for Social Security-registered Firms, do not provide a breakdown of employment by firm size, sector and region. Accordingly, the CBI allows for analysis of the Spanish business sector from new dimensions, some of which are considerably important from an economic policy standpoint. By way of example, the remainder of this section uses the CBI to study the regional patterns of the concentration of employment at smaller enterprises. This is a matter that, as mentioned above, at national level constitutes one of the most important factors behind the low productivity growth in the Spanish economy.

In order to isolate the sectoral composition effect, Table 4 shows the share of employment concentrated at smaller enterprises (1-49 employees), by region and across all four sectors analysed.[26] At national level, the concentration of employees at smaller enterprises is higher in construction (66.8%) and in tourism-related services (47.2%), while in industry (36.1%) and in other services (36%) there is a lower share of employment in this size category. This indicates that regional differences in productive specialisation affect firm size distribution. In addition, within each sector there are also notable cross-regional differences in firm size, which, as mentioned above, can only be analysed thanks to the richness of the CBI firm-level data. For example, in industry, Asturias, Navarre and Madrid are the regions with the lowest share of employment at smaller enterprises (30.7%, 26.5% and 19%, respectively),

[25] These aggregate national figures are similar to those drawn from the Statistics for Social Security-registered Firms, according to which smaller enterprises accounted for 48% of employees in the period 2013-2021.

[26] The 50 employee threshold is important because academic literature has shown that the existence of rules and regulations that are applied to firms with 50 or more employees (such as the requirement to establish a workers' committee) may discourage business growth and lead to an over-allocation of resources to less productive firms. Maza (2024) examines how thresholds in accounting and financial reporting requirements influence businesses' growth decisions. Almunia, Jimeno, López-Rodríguez and Petit (2024) analyse the effects of the interaction of labour and fiscal regulations on the size of Spanish firms.

while in the Balearic Islands it is almost 70% and in Extremadura it stands at close to 60%.

Table 4. Share of employment at smaller firms (1-49 employees) in the CBI (%) (a)

Region	Industry	Construction	Tourism-related services	Other services	Total
Andalusia	50.0	74.5	70.1	59.8	61.6
Aragon	35.0	82.9	59.8	61.3	52.9
Asturias	30.7	56.3	77.3	51.0	49.0
Balearic Islands	67.4	84.3	40.6	58.7	54.9
Canary Islands	52.0	70.7	44.2	52.2	51.2
Cantabria	37.2	74.4	66.7	57.5	54.5
Castile-Leon	42.1	87.8	83.0	62.2	61.4
Castile-La Mancha	55.5	58.4	75.3	78.8	66.5
Catalonia	33.2	73.6	50.5	43.1	43.6
Valencia	49.4	82.0	64.7	42.2	50.7
Extremadura	58.5	87.7	82.5	65.8	69.6
Galicia	39.3	76.1	63.8	44.7	49.6
Madrid	19.0	40.2	22.9	16.3	19.3
Murcia	43.0	78.1	64.8	47.7	51.6
Navarre	26.5	72.7	67.1	69.6	45.7
Basque Country	33.0	64.5	54.3	43.7	42.9
Rioja	50.7	94.0	80.9	78.6	66.3
Total	36.1	66.8	47.2	36.0	40.6

SOURCE: Banco de España

(a) Average for the period 2009-2021. Including non-financial and non-agricultural market economy firms, excluding firms without any employees.

Turning to construction, according to the CBI around 90% of workers in Rioja, Castile-Leon and Extremadura are employed at smaller enterprises, versus the lowest percentages in Asturias (56.3%), Castile-La Mancha (58.4%) and Madrid (40.2%). Meanwhile, in tourism-related services, the highest percentage of employment at smaller enterprises exceeds 80% in Extremadura and Castile-Leon, in contrast to the Canary Islands and the Balearic Islands (around 40%). With regard to other services, the high share of employment at smaller enterprises in Rioja and Castile-La Mancha (almost 80%) contrasts with that in Madrid (16.3%) and Catalonia and Valencia (around 40%).

Given this high regional heterogeneity in the distribution of employment by firm size, we should consider the factors that are potentially behind these differences. In this regard, the first aspect to consider is that, broadly speaking, the share of employment at smaller enterprises tends to be lower in those sectors in which the region is more specialised (measured by the sector's share of total employment in each region), except for construction, where this correlation is not observed (see Chart 4).

The CBI firm-level data also allow us to examine whether the regional firm size distribution has changed over time. Indeed, one of the most notable developments in the Spanish economy in recent years has been the trend towards a lower concentration of employment at smaller enterprises. Specifically, according to the Statistics for Social Security-registered Firms, the percentage of employees at such firms fell from 43.3% in 2009 to 42.2% in 2021 and 39.5% in 2024.

The CBI allows for a more in-depth analysis of the changes in the firm size distribution of employment by enabling the regional dimension to be assessed and

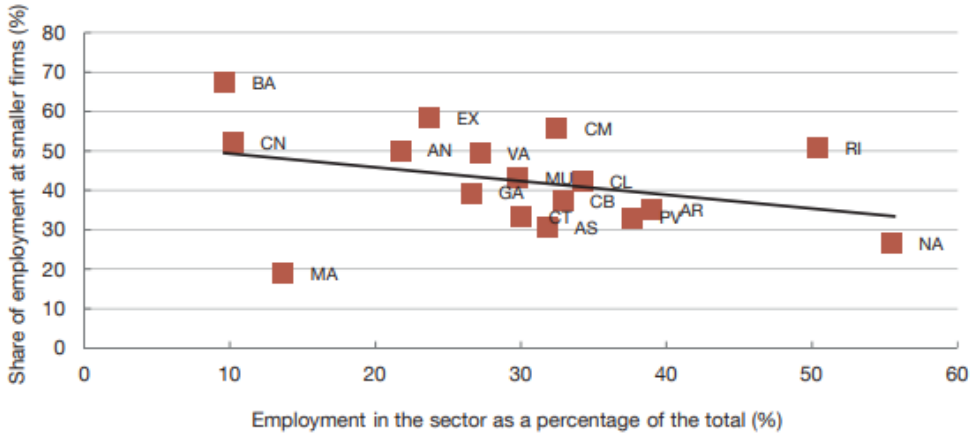
by isolating the effects of the different sectoral compositions across regions. Table 5 shows, for the 17 regions and four sectors under review, the change between the average shares of employment at smaller enterprises for the periods 2009-2014 and 2015-2021. By sector, the largest declines in the share of employment at smaller enterprises were in other services. These declines were particularly steep in Cantabria and Murcia. By contrast, in tourism-related services, the share of employment at smaller enterprises increased slightly in the country as a whole, albeit with a high level of dispersion across regions.

Employment at smaller enterprises decreased across all regions in the period under review, although it has tended to decline at a faster pace in the regions with a higher share in 2009 (see Chart 5). For example, in Castile-La Mancha and Extremadura, that share, which stood at around 75% in 2009, fell by 13 pp in both regions; all four sectors analysed contributed to this fall. At the other end of the scale, in the Basque Country and Madrid, this ratio fell very moderately (around 2 pp), given that the share of employment at smaller enterprises was already relatively low in 2009. In sum, this empirical evidence suggests that all regions appear to be trending towards a lower concentration of employment at smaller enterprises.

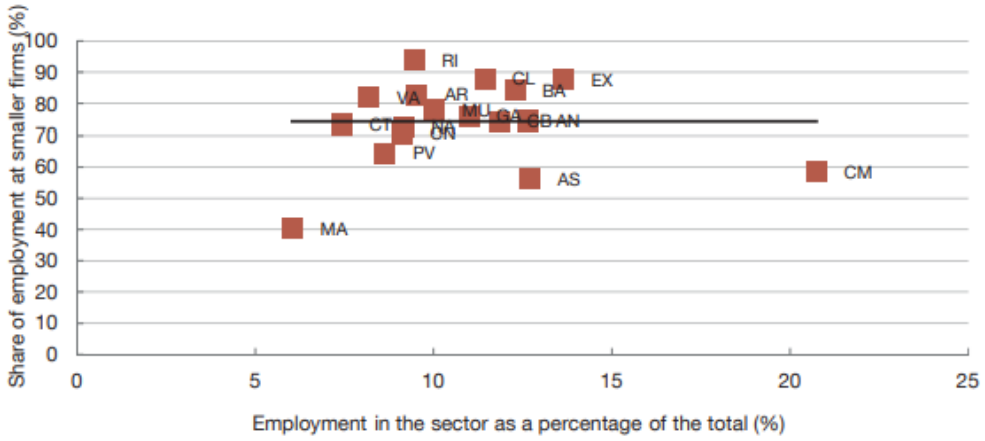
In short, the proportion of employees at smaller enterprises is considerably different across regions, even within the main sectors of activity. Madrid, followed by the Basque Country, Navarre and Catalonia, has the lowest proportion of employment at smaller enterprises, while at the other end of the scale, employment tends to be relatively more concentrated at such enterprises in Extremadura and Castile-La Mancha.

Chart 4. Relationship between the sectoral specialisation of each region and the share of employment at smaller firms (fewer than 50 employees) in each sector (a)

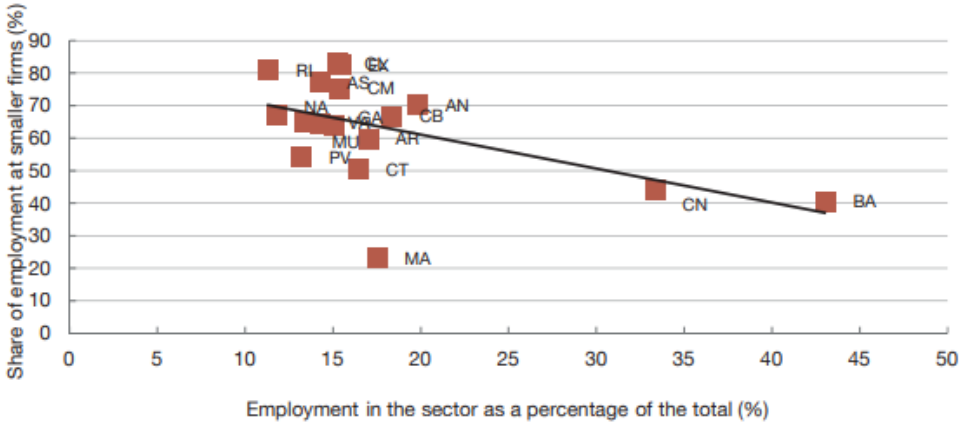
4.a Industry



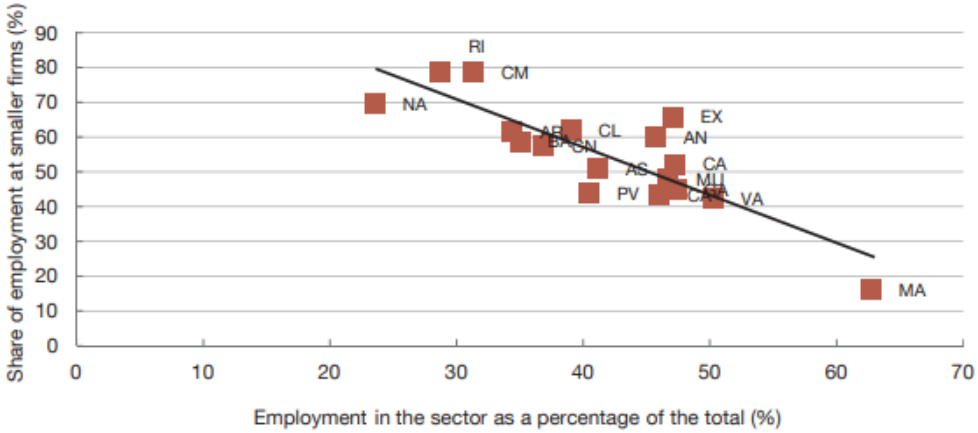
4.b Construction



4.c Tourism-related services



4.d Other services

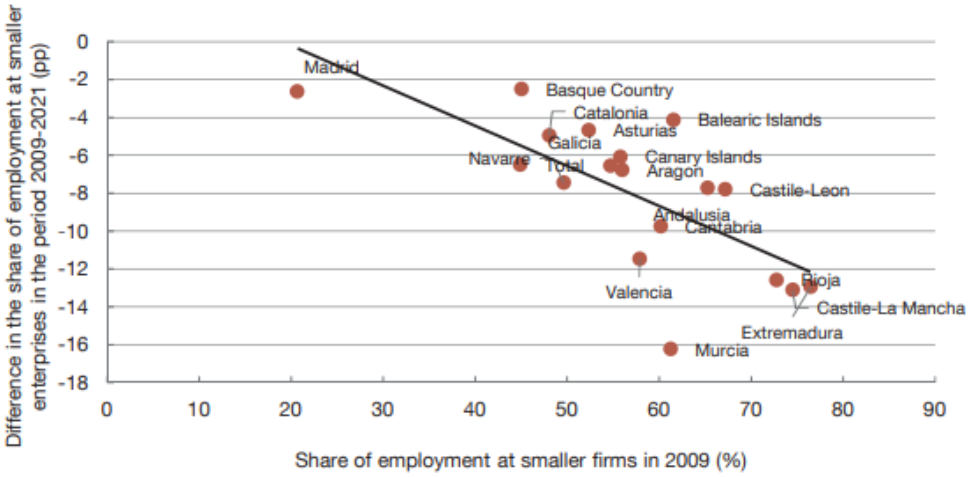


SOURCE: Banco de España.

(a) Average for the period 2009-2021. "Industry" includes the firms under NACE Rev. 2 sections B (mining and quarrying), C (manufacturing), D (electricity, gas, steam and air-conditioning supply) and E (water supply, sewerage, waste management and remediation). "Construction" includes firms in sector F (construction). "Tourism-related services" includes firms in sectors H (transportation), I (accommodation and food service

activities) and R (arts, entertainment and recreation). "Other services" includes firms in sectors G (wholesale and retail trade), J (information and communication), M (professional, scientific and technical activities) and N (administrative and support service activities). AN: Andalusia, AR: Aragon, AS: Asturias, BA: Balearic Islands, CN: Canary Islands, CB: Cantabria, CL: Castile-Leon, CM: Castile-La Mancha, CT: Catalonia, VA: Valencia, EX: Extremadura, GA: Galicia, MA: Madrid, MU: Murcia, NA: Navarre, PV: Basque Country, RI: Rioja.

Chart 5. Change in share of employment at smaller enterprises (1-49 employees) in the CBI (a)



SOURCE: Banco de España.

(a) Including non-financial and non-agricultural market economy firms, excluding firms without any employees.

Table 5. Change in the percentage of employment at smaller firms (1-49 employees) in the CBI between 2009-2014 and 2015-2021 (pp) (a)

Region	Industry	Construction	Tourism-related services	Other services	Total
Andalusia	-3.7	0.9	-2.0	-14.2	-7.7
Aragon	-7.5	-8.8	-14.5	2.0	-6.8
Asturias	1.1	1.8	-5.6	-12.5	-4.7
Balearic Islands	1.7	-3.0	-7.3	-7.4	-4.1
Canary Islands	-2.6	6.8	-2.2	-11.3	-6.1
Cantabria	-5.1	8.2	-2.9	-22.2	-9.7
Castile-Leon	-8.9	3.0	-6.5	-5.5	-7.8
Castile-La Mancha	-18.1	0.4	-17.3	-14.7	-13.1
Catalonia	-1.8	2.0	1.5	-9.5	-5.0
Valencia	-13.1	-0.2	-5.1	-11.5	-11.5
Extremadura	-13.6	6.2	-7.5	-19.7	-12.9
Galicia	-6.9	-4.3	-0.7	-6.9	-6.6
Madrid	-5.4	4.8	2.4	-3.0	-2.6
Murcia	-11.0	-0.1	-9.0	-22.1	-16.2
Navarre	-5.6	4.6	-8.5	-6.5	-7.4
Basque Country	-1.3	0.1	-4.3	-2.7	-2.5
Rioja	-12.0	3.2	-17.2	-15.0	-12.6
Total	-5.8	1.0	0.3	-8.2	-6.5

SOURCE: Banco de España.

(a) Including non-financial and non-agricultural market economy firms, excluding firms without any employees

5. Conclusions

This paper shows that the CBI database is a useful source of information for analysing the structure of the Spanish business sector from a regional perspective. The analysis of the coverage rates and the comparison of the productive systems, employment dynamics and firm size distribution with other official sources allow us to conclude that the sample of firms in the CBI is representative of the population of NFCs at regional level. In addition, the empirical exercises presented in this paper indicate that there are considerable cross-regional differences in firm size distribution. This highlights the importance of factoring in the regional dimension when tackling the challenge of the small size of Spanish firms. Although the analysis has been conducted at regional level, the high number of firms available in the CBI and the broad coverage in the small, single-province regions point to the potential use of these firm-level data at provincial level.

Thus, the availability of these firm-level data that are representative at regional level for the Spanish economy opens the door to various lines of research. In particular, the financial and economic information contained in the CBI enables analysis of regional heterogeneity in different key areas. For example, the regional analysis of the determinants of business size and growth, productivity, profit margins and investment dynamics are some of the areas of interest that represent a priority line of work for the Banco de España.

References

[1] Almunia, Miguel, David López-Rodríguez and Enrique Moral-Benito. (2018). "Evaluating the Macro Representativeness of a Firm-Level Database: An Application for the Spanish Economy". Documentos Ocasionales, 1802, Banco de España. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3132982

[2] Almunia, Miguel, Juan F. Jimeno, David López-Rodríguez and Borja Petit. (2024). "Size-dependent regulations in Spain". Mimeo.

[3] Banco de España. (2022a). "Nota metodológica". In Banco de España, Resultados anuales de las empresas no financieras 2022. pp. 15-61. https://www.bde.es/f/webbe/SES/Secciones/Publicaciones/PublicacionesAnuales/CentralBalances/22/Fich/ceba22nm.pdf

[4] Banco de España. (2022b). "Resultados anuales de las empresas no financieras 2022. Suplemento metodológico". https://www.bde.es/f/webbe/SES/Secciones/Publicaciones/PublicacionesAnuales/CentralBalances/22/Fich/ceba22sm.pdf

[5] Banco de España. (2024). "Chapter 2. Structural challenges facing the Spanish economy". In Banco de España, Annual Report 2023, pp. 91-159. https://repositorio.bde.es/handle/123456789/36537

[6] Fernández Cerezo, Alejandro, Matías Pacce and Isabel Sánchez. (2024). "La heterogeneidad regional en la evolución reciente de la inflación en España". Boletín Económico - Banco de España, 2024/T3, 08. https://doi. org/10.53479/37592

[7] Gavilán, Ángel, and Domingo García Coto. (2024). "El reto del tamaño empresarial en España". Documento de Trabajo, 35, Fundación del Instituto Español de Analistas. https://institutodeanalistas.com/wp-content/uploads/DOC.-No-35-El-reto-del-tamano-empresarial-en-Espana.pdf

[8] Kalemli-Ozcan, Sebnem, Bent Sorensen, Carolina Villegas-Sanchez, Vadym Volosovych and Sevcan Yesiltas. (2015). "How to Construct Nationally Representative Firm Level Data from the ORBIS Global Database". NBER Working Papers, 21558, National Bureau of Economic Research. https://doi.org/10.3386/w21558

[9] Maza, Luis Ángel. (2024). "Una reflexión sobre los umbrales cuantitativos en los modelos de depósito de las cuentas anuales y su posible impacto en el tamaño empresarial en España". Documentos Ocasionales, 2419, Banco de España. https://doi.org/10.53479/36639

[10] Mora-Sanguinetti, Juan S., and Isabel Soler. (2022). "La regulación sectorial en España. Resultados cuantitativos". Documentos de Trabajo, 2202, Banco de España. http://dx.doi.org/10.2139/ssrn.4035629

[11] Moral-Benito, Enrique. (2018). "Growing by learning: firm-level evidence on the size-productivity nexus". SERIEs, 9(1), pp. 65-90. https://doi.org/10.1007/s13209-018-0176-2

[12] Prades Illanes, Elvira, and Patrocinio Tello Casas. (2020). "The heterogeneous economic impact of COVID-19 among euro area regions and countries". Economic Bulletin - Banco de España, 2/2020, Analytical Articles. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/InformesBoletinesRevistas/ArticulosAnaliticos/20/T2/descargar/Files/be2002-art17e.pdf

BANCO DE ESPAÑA
Eurosistema

THE REGIONAL PERSPECTIVE OF THE BANCO DE ESPAÑA'S CENTRAL BALANCE SHEET DATA OFFICE (CBSO)

Alejandro Fernández Cerezo
Borja Fernández-Rosillo San Isidro
Natividad Pérez Martín

WORKSHOP: "NEW INSIGHTS FROM FINANCIAL STATEMENTS"

BANCO DE ESPAÑA
Madrid, 17th october 2024





BANCO DE ESPAÑA
Eurosistema

CONTENT

1. Introduction
2. “BExplore Regional Central Balance Sheet Data Office”
3. Regional coverage and representativeness of the CBI database
4. Exploring heterogeneity in firm size distribution at regional level
5. Conclusions

- Firm-level databases enable a deeper understanding of economic developments, also at regional level:
  - Regional differences in corporate variables may reveal trends not observable at national level.
  - Shocks affecting the whole country (pandemic/energy) may have asymmetric impacts across regions.
  - In decentralized countries, different levels of regulation across autonomous communities (CCAA).

- Almunia, López-Rodríguez and Moral-Benito (2018) evaluated the macro-representativiness of the Banco de España's Integrated Central Balance Sheet Database (CBI, in Spanish).
  - The CBI was able to replicate, at national level, the growth rate of employment in the Spanish private sector and the firm size distribution.

- In Fernández-Cerezo, Fernández-Rosillo and Pérez-Martín (2024), we show that the CBI database is representative also at regional (CCAA) level:
  - The CBI exhibits high coverage of firms with respect to the population of firms in each CCAA.
  - The CBI replicates the sectoral distribution of employment across CCAA.
  - The CBI replicates the dynamics of aggregate employment at CCAA level.

What is the CBI* database?

Information based on financial accounts of Spanish NFCs:
balance sheet, income statement, # employees, **location****…

Sources of information of CBI

Voluntary questionnaire directly from firms +
Mandatory deposits of annual accounts in Mercantile Registries

Data treatment & cleansing to raw firm data

From 1M NFCs received per year to 840k firms suitable for analysis after data cleansing by CBSO staff

Main characteristics of the CBI

Unit of analysis: individual enterprise (identified by anonymised tax id)
Frequency: annual, from 1995 to 2022

How can CBI data be accessed and visualized?

For researchers: microdata through BELab data laboratory
For general public: aggregate data through "BExplore" visualizations

* *CBI: Integrated Central Balance Sheet Database*
** *Location:* each firm is located in the province where it is headquartered.

2. "BEXPLORE REGIONAL CENTRAL BALANCE SHEET DATA OFFICE"

An interactive website to provide the general public with easy access to aggregate CBI regional data

BANCO DE ESPAÑA
Eurosistema

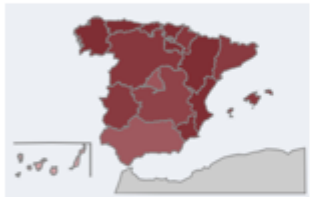
In order to bring regional CBI analysis closer to users, an interactive tool named **"BExplore Central Balance Sheet Data Office Regional"** has been developed

MAIN CONTENT

Available statistical information:

Nº of firms (% coverage) | Weighted Average Value | Median value (Q2)

about:



17 autonomous communities (CCAA) + Ceuta/Melilla

for:



52 provinces



14 sectors of activity



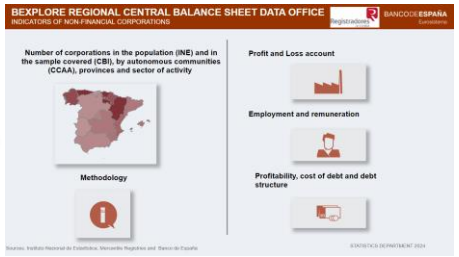
4 company sizes



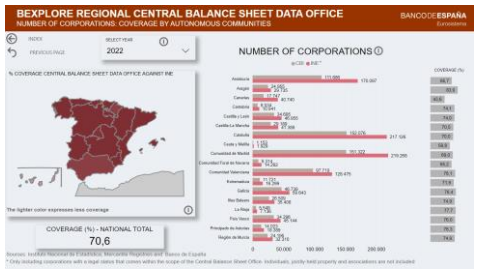
15 years (2008-2022)



11 variables

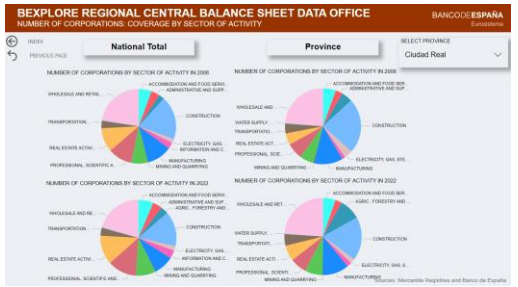


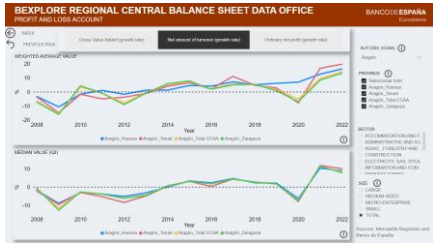
Power BI



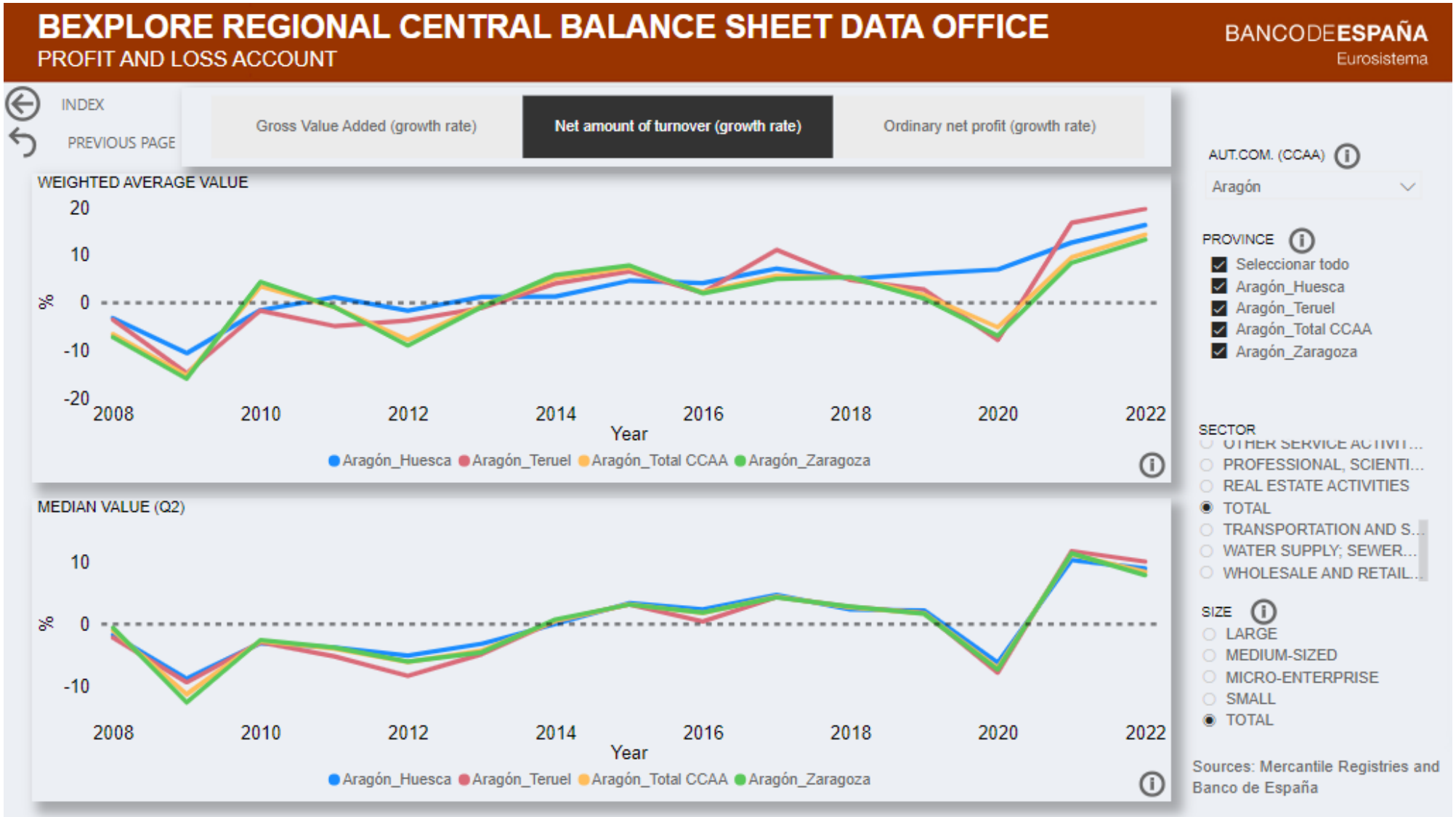
Coverage rates

Sectoral breakdown





-P&L variables
-Employment and labor costs variables
-Financial ratios: margins, financial cost, debt ratio...


BEXPLORE REGIONAL CENTRAL BALANCE SHEET DATA OFFICE
PROFIT AND LOSS ACCOUNT
BANCO DE ESPAÑA
Eurosistema
INDEX
PREVIOUS PAGE
Gross Value Added (growth rate)
Net amount of turnover (growth rate)
Ordinary net profit (growth rate)
WEIGHTED AVERAGE VALUE
20
10
0
-10
-20
%
2008
2010
2012
2014
2016
2018
2020
2022
Year
Aragón_Huesca
Aragón_Teruel
Aragón_Total CCAA
Aragón_Zaragoza
MEDIAN VALUE (Q2)
10
0
-10
%
2008
2010
2012
2014
2016
2018
2020
2022
Year
Aragón_Huesca
Aragón_Teruel
Aragón_Total CCAA
Aragón_Zaragoza
AUT.COM. (CCAA)
Aragón
PROVINCE
Seleccionar todo
Aragón_Huesca
Aragón_Teruel
Aragón_Total CCAA
Aragón_Zaragoza
SECTOR
PROFESSIONAL, SCIENTI...
REAL ESTATE ACTIVITIES
TOTAL
TRANSPORTATION AND S...
WATER SUPPLY; SEWER...
WHOLESALE AND RETAIL...
SIZE
LARGE
MEDIUM-SIZED
MICRO-ENTERPRISE
SMALL
TOTAL
Sources: Mercantile Registries and Banco de España

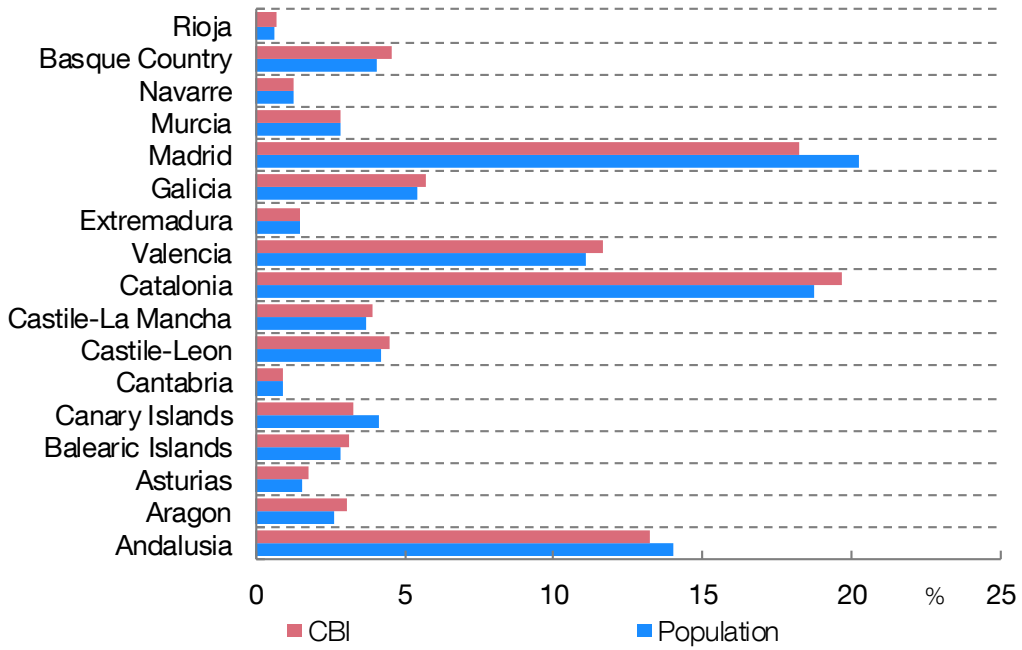
3. REGIONAL COVERAGE AND REPRESENTATIVENESS OF THE CBI DATABASE

The CBI database is reasonably representative of the population of Spanish NFCs at regional level

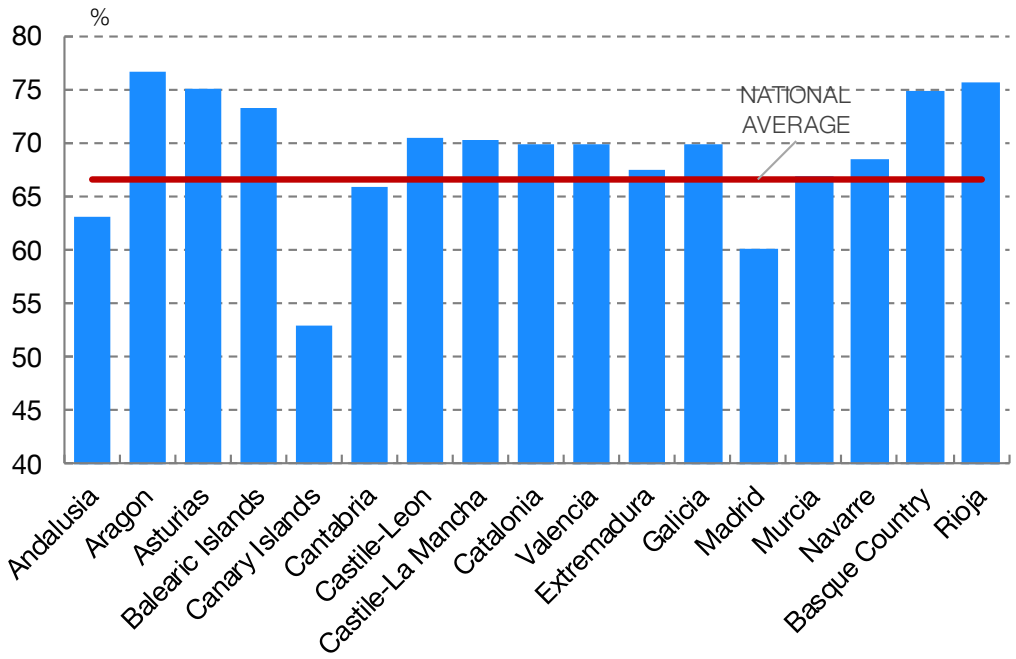
The regional distribution of the # of Spanish NFCs in the CBI adequately replicates that of the population

The coverage rate at national level is 66.5% on average, exceeding 70% in several CCAA

DISTRIBUTION OF THE NUMBER OF NFCs IN CBI AND IN POPULATION BY CCAA (a)



COVERAGE RATES BY CCAA (# NFCs IN CBI / # NFCs IN POPULATION) (a)



Sources: Banco de España and INE

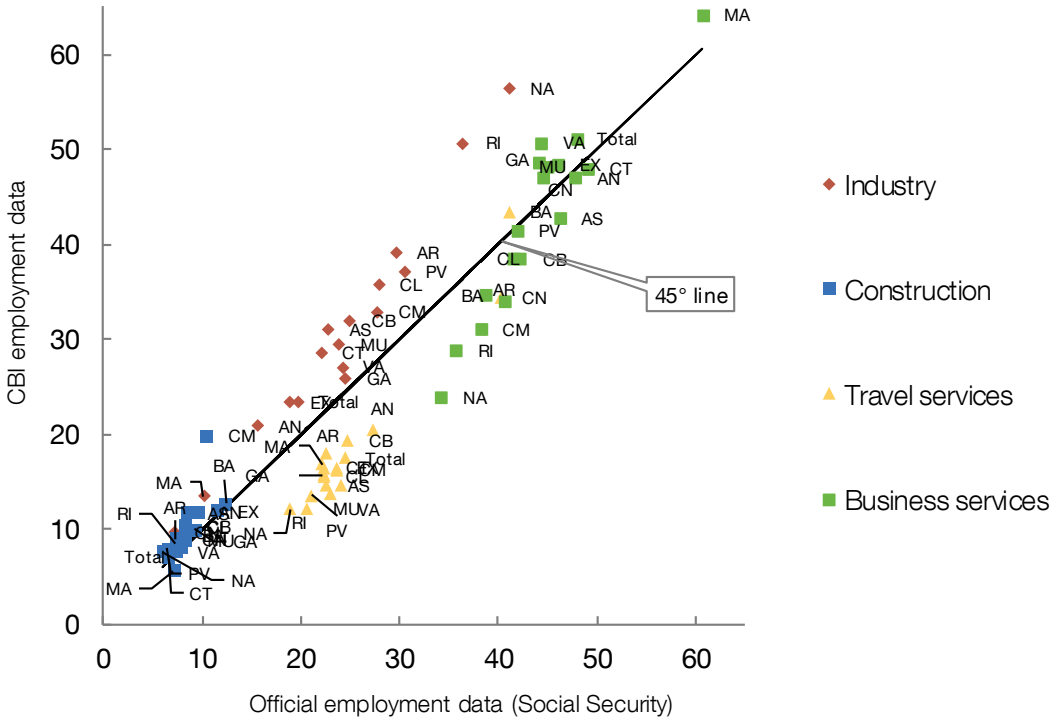
(a): Average for the period 2009-2021. The number of firms comprising the population of NFCs in Spain comes from the Central Business Register, prepared by the National Statistics Institute (INE), which in turn feeds Eurostat's structural business statistics.

3. REGIONAL COVERAGE AND REPRESENTATIVENESS OF THE CBI DATABASE

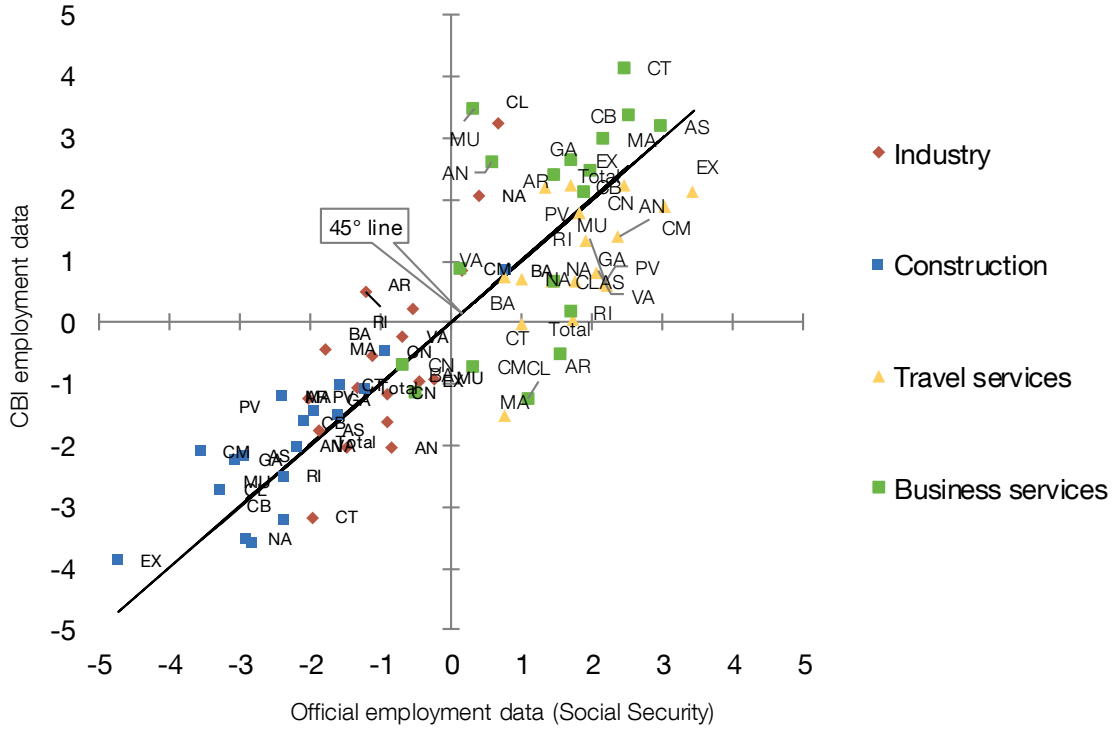
The CBI database replicates the sectoral distribution of employment and the changes over time across CCAA

BANCO DE ESPAÑA
Eurosistema

SECTORAL SHARES OF EMPLOYMENT BY CCAA (% OF TOTAL EMPLOYMENT IN EACH REGION) (a) (b)



CHANGES IN SECTORAL SHARES OF EMPLOYMENT BY CCAA (DIFFERENCES BETWEEN THE AVERAGES FOR THE PERIODS 2015-2021 AND 2009-2014 IN PP) (b)



Sources: Banco de España and Ministry of Inclusion, Social Security and Migration.
(a): Average for the period 2015-2021.
(b) "Industry" includes the firms under NACE Rev. 2 sections B, C, D and E. "Construction" includes firms in F. "Travel services" includes firms in H, I and R. "Business services" includes firms in G, J, M and N. AN: Andalusia, AR: Aragon, AS: Asturias, BA: Balearic Islands, CN: Canary Islands, CB: Cantabria, CL: Castile-Leon, CM: Castile-La Mancha, CT: Catalonia, VA: Valencia, EX: Extremadura, GA: Galicia, MA: Madrid, MU: Murcia, NA: Navarre, PV: Basque Country, RI: Rioja.

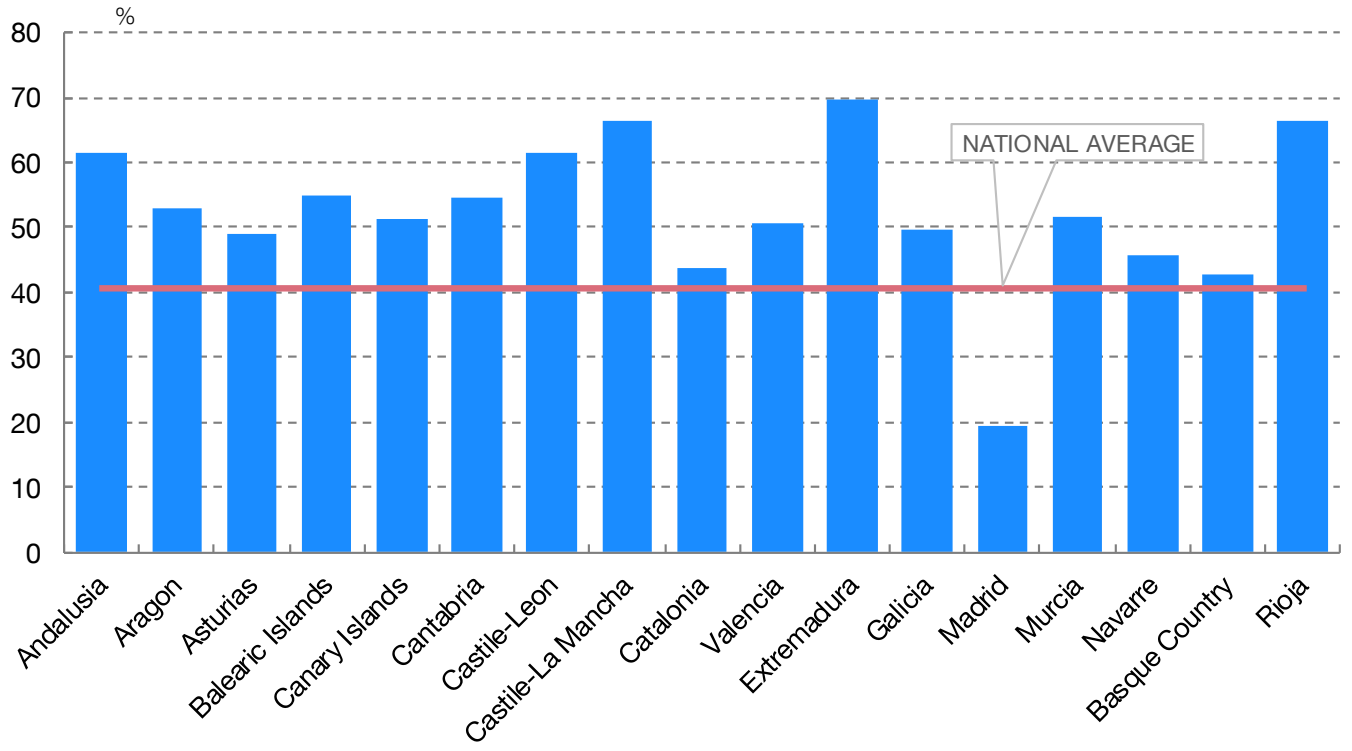
4. EXPLORING HETEROGENEITY IN FIRM SIZE DISTRIBUTION AT REGIONAL LEVEL

The CBI is a unique source in Spain for analysing regional heterogeneity in firm size distribution

BANCO DE ESPAÑA
Eurosistema

- According to CBI, small firms (<50 employees) account for 70% of employment in Extremadura and only 20% in Madrid in the non-financial and non-agricultural market economy in Spain (vs 40% in the national average).

SHARE OF EMPLOYMENT AT SMALLER FIRMS (1-49 EMPLOYEES) IN CBI (a)



Source: Banco de España.
(a): Average for the period 2009-2021. Including non-financial and non-agricultural market economy firms, excluding firms without any employees.

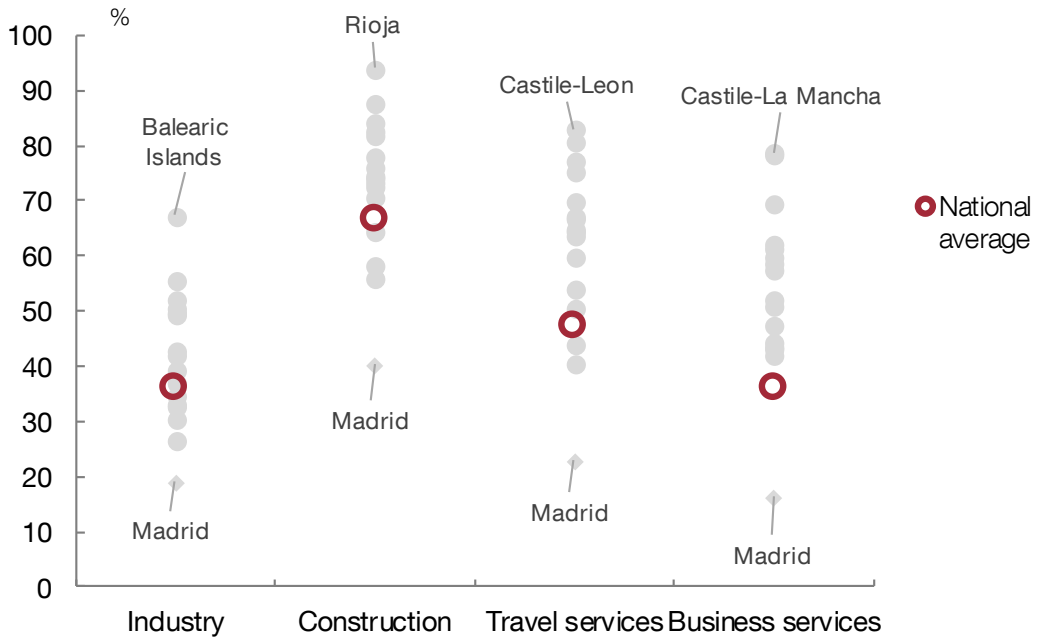
4. EXPLORING HETEROGENEITY IN FIRM SIZE DISTRIBUTION AT REGIONAL LEVEL

CBI provides a breakdown of employment by firm size, sector and region

BANCO DE ESPAÑA
Eurosistema

- Within each sector there are notable cross-regional differences in firm size.

SHARE OF EMPLOYMENT AT SMALLER FIRMS (<50 EMPLOYEES) IN CBI (a)



Source: Banco de España.
(a): Average for the period 2009-2021. Including non-financial and non-agricultural market economy firms, excluding firms without any employees. "Industry" includes the firms under NACE Rev. 2 sections B, C, D and E. "Construction" includes firms in F. "Travel services" includes firms in H, I and R. "Business services" includes firms in G, J, M and N. AN: Andalusia, AR: Aragon, AS: Asturias, BA: Balearic Islands, CN: Canary Islands, CB: Cantabria, CL: Castile-Leon, CM: Castile-La Mancha, CT: Catalonia, VA: Valencia, EX: Extremadura, GA: Galicia, MA: Madrid, MU: Murcia, NA: Navarre, PV: Basque Country, RI: Rioja.

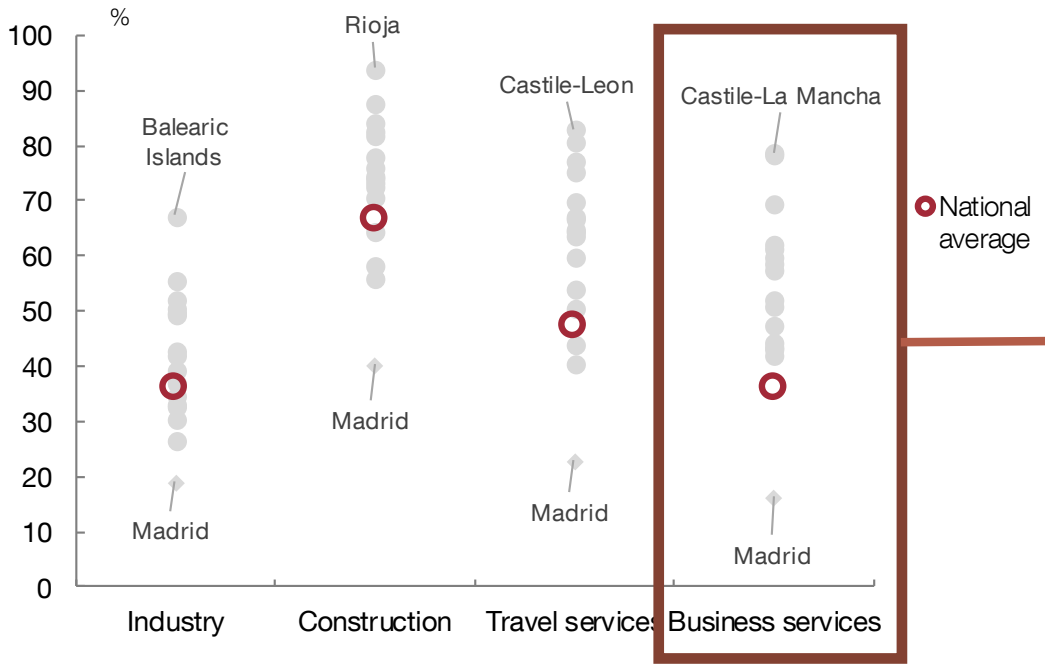
4. EXPLORING HETEROGENEITY IN FIRM SIZE DISTRIBUTION AT REGIONAL LEVEL

CBI provides a breakdown of employment by firm size, sector and region

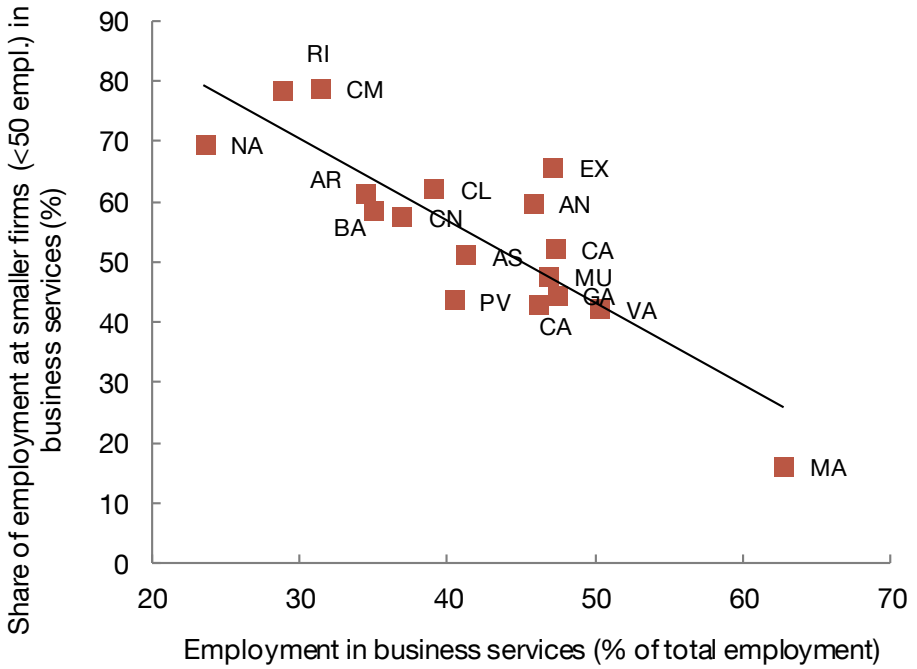
BANCO DE ESPAÑA
Eurosistema

- Within each sector there are notable cross-regional differences in firm size.
- The share of employment at smaller enterprises tends to be lower in those sectors in which the region is more specialized.

SHARE OF EMPLOYMENT AT SMALLER FIRMS (<50 EMPLOYEES) IN CBI (a)



SECTORAL SPECIALISATION AND THE SHARE OF EMPLOYMENT AT SMALLER FIRMS IN THE BUSINESS SERVICES SECTOR (a)



Source: Banco de España.

(a): Average for the period 2009-2021. Including non-financial and non-agricultural market economy firms, excluding firms without any employees. "Industry" includes the firms under NACE Rev. 2 sections B, C, D and E. "Construction" includes firms in F. "Travel services" includes firms in H, I and R. "Business services" includes firms in G, J, M and N. AN: Andalusia, AR: Aragon, AS: Asturias, BA: Balearic Islands, CN: Canary Islands, CB: Cantabria, CL: Castile-Leon, CM: Castile-La Mancha, CT: Catalonia, VA: Valencia, EX: Extremadura, GA: Galicia, MA: Madrid, MU: Murcia, NA: Navarre, PV: Basque Country, RI: Rioja.



Representative database

- The CBI database is a representative source of information for analyzing the structure of the Spanish business sector from a regional perspective.



Interactive tools

- In an effort to ease both external researchers' and the general public's access to Bank of Spain's databases, we are developing new user-friendly, interactive tools to present the potential uses of the CBI databases.



Cross-regional differences

- We showed that there are considerable cross-regional differences in firm size distribution.



Diverse potential research lines

- The availability of CBI firm-level data opens the door to various lines of research for analysis of regional heterogeneity in different key areas: e.g. the determinants of business size and growth, productivity, profit margins and investment dynamics.

BANCO DE ESPAÑA
Eurosistema

THANKS FOR YOUR ATTENTION



IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

How resilient companies navigated through Covid-19 pandemics[1]

António Costa and Pedro Cordeiro,
Banco de Portugal

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

How resilient companies navigated through Covid-19 pandemics

António Costa, Pedro Cordeiro

Abstract

Based on the financial statements of Portuguese non-financial corporations between 2015 and 2022, this study aims to assess the resilience of these companies in the face of external shocks. One of the most significant shocks in recent history was the Covid-19 pandemic, which began in the first quarter of 2020 and significantly altered the lives of people and the operations of companies due to the restrictions implemented as the pandemic worsened. This clearly affected how companies operated across different segments of activity, leading to an abrupt transition to digital practices, which was not possible at the same pace for all sectors.

Given that the pandemic had varying impacts on the companies within the Portuguese economy, depending on their sector of activity and their ability to adapt to new ways of operating in the market, this exploratory study aims to analyze how companies classified as resilient during the period 2015-2019 reacted to this shock. The study considers all companies that provided information for all the years from 2015 to 2019. Based on the financial statements reported annually by the companies in Annex A of the Simplified Corporate Information (IES), economic and financial indicators were calculated and analyzed to determine the resilience of the companies, initially, and based on the 2022 financial reports understand how the performance of these companies was affected, or not, by the pandemic crisis.

Keywords: companies' financial statements, Covid-19, resilience

JEL classification: G23, G32

Table of contents

How resilient companies navigated through Covid-19 pandemics 1

1. Introduction 3

2. Methodology 4

3. Data analysis 6

3.1. Full database characterization 6

3.2. Main results 6

3.2.1 Transportation and Storage 6

3.2.2 Accommodation and Food Service Activities 10

4. Conclusion 14

Bibliography 15

1. Introduction

Financial statements are one of the key elements used in analyses of different nature, in particular economic and financial analysis, not only to provide useful information for the company's own management but also for financial markets and its users / investors to assess companies' potential risks and their resilience, among others.

Using financial statements as basis, the aim of this study is to analyze the behavior of companies considered resilient before the Covid-19 pandemic and how they performed during and after this period. For this purpose, the sectors of activity most affected by the pandemic – Transportation and Storage and Accommodation and Food Service Activities (section H and I of NACE – Statistical Classification of Economic Activities) are going to be the focus of this analysis. The goal was to find evidence on how the companies that comprise those sectors managed to navigate through the challenges posed during the Covid-19 years comparing the key financial figures from 2019 and 2022.

To conduct the analysis, annual data from the Central Balance sheet database, which includes micro data of all Portuguese non-financial corporations, was used. The primary source of information of this database is the Annex A of the Simplified Corporate Information (IES). For study purposes we selected the years from 2015 to 2022, since this period provides a comprehensive overview of the companies' financial health prior to the pandemic, which is essential for assessing their resilience according to the defined criteria. Additionally, to observe the evolution of the proportion of resilient companies in different periods a second dataset from 2012 to 2019 was defined.

The Covid-19 pandemics, which began in the first quarter of 2020, significantly changed people's lives and company operations due to the restrictions introduced as the pandemic worsened. This significantly impacted how companies operated across different segments of activity, leading to an abrupt transition to digital practices, which was not possible for all sectors and at the same pace.

This paper is organized as follows. Chapter 2 presents a brief explanation of the methodology used to assess the resilience of a company and the rationale for the ratios used in the analysis. Chapter 3 identifies and describes the sectors most affected by the Covid-19 pandemic, comparing their financial situation between 2019 and 2022. Chapter 4 highlight the main conclusions.

2. Methodology

This chapter provides a brief description of the assumptions considered in the definition of the dataset, the rationale for the financial ratios used in the analysis, the underlying criteria employed to classify a company as resilient and the selected sectors of activity analyzed in the subsequent section.

Our dataset covers a time span from 2012 to 2022, comprising all the non-financial corporations (S11 companies according to the ESA 2010 definition) with financial statements available for each year within this period[1]. The classification of company size follows the criteria outlined in the EU recommendation 2003/361, and comprises micro, small, medium, and large dimensions.

After establishing the dataset, we selected five financial ratios that are most commonly used in financial companies' analysis. These ratios cover different aspects, including (i) financial and indebtedness structure since "*companies' financial debt is not only crucial for companies' investments and expansion as well as to finance its current activity*" (Barbosa, L. and de Pinho, P. (2016)), (ii) liquidity "*because it gives us a perspective if short-term debt has or not, a greater coverage by assets that can be converted into net financial assets also in short term*" (Martins A. and Santos M. (2022)), and (iii) operational performance / profitability since "*EBITDA, ROE and ROA are financial performance indicators that measure respectively, operating return, revenue and cost management capacity and the ability to generate returns on the total investments*" (Carvalho et al (2016)). It is also important to highlight the relevance of the selected ratios in risk assessment of companies, as it can be observed, for example, in the S&P Global Ratings' methodology for rating corporate industrial companies and utilities.

Thus, the financial ratios considered were:

- Equity to assets ratio $= \frac{Equity}{Total\ Assets}$;
- Financial Leverage $= \frac{Financial\ Debt}{Financial\ Debt + Equity}$;
- Current ratio $= \frac{Current\ Assets}{Current\ Liabilities}$;
- EBITDA margin $= \frac{EBITDA\ (Earnings\ Before\ Interest,\ Taxes,\ Depreciation\ and\ Amortization)}{Revenue}$;
- Financial Debt / EBITDA;

To assess whether a company is resilient or not in 2019 we compare individual companies' financial performance against their respective sector ratios, aggregating companies according to their NACE sections.

A company is considered to be resilient in 2019 if it meets the following criteria:

- For the last two years before pandemic (2018 and 2019), the company must have a positive EBITDA. Additionally, between 2015 and 2017, the company must record a positive EBITDA at least in an additional year.

[1] Companies born during these years (2015 to 2019) are excluded from the dataset. On the other hand, firms that ceased its activity and do not report its financial statements will not be considered as resilient.

- Therefore, for each year a score will be recorded and it will be considered positive if a company presents all of its ratios equal or above the weighted average of the sector in which it operates. Additionally, it is also relevant to highlight that:
  - Regarding the current ratio, this ratio needs to be higher than 100%. A low ratio (less than 100%) indicates difficulty to meet short-term financial obligations and the inability to take advantage of opportunities requiring quick cash;
  - For entities that have financial debt equal to "0", in the ratios in which this variable was used, these entities will be considered as being above the sector, since they will not face any constrains in meeting at least financial obligations.

Therefore, if a company presented a positive score in the last two years before pandemic (2018 and 2019) and one more positive score in the period from 2015 to 2017 and also satisfies the EBITDA criterion, then company is considered resilient.

Using the same criteria as previously outlined, but now applying them to financial statements from 2018 to 2022, the aim was to observe and analyze the impact of the COVID-19 pandemic on companies' resilience in the specific NACE sections and that will be examined in the next Chapter. By extending the timeframe to include the pandemic years, we can assess how the crisis affected companies' financial resilience within these sectors.

Also using the same criteria, an additional sample was constructed for the period between 2012 and 2016, in order to observe which companies were resilient in 2016, as well as between 2015 and 2019. By comparing the proportion of companies that keep this classification between 2016 / 2019 and 2019 / 2022 we can also check the Covid-19 impact in companies' resilience.

The next Chapter will focus on the NACE sections that were considered to be the most significantly impacted by the COVID-19 pandemic. To define these sectors of activity, we based on internal information available in Banco de Portugal, more precisely a survey jointly conducted by Banco de Portugal and the Statistics Portugal (Instituto Nacional de Estatística). It is possible to observe that the H (Transportation and Storage) and I (Accommodation and Food Service Activities) NACE sections were the most affected not only in terms of revenue reduction but also in what concerns the average time needed to recover to the regular activity level. As such we considered that it would be more beneficial to look deeper into these two NACE sections and consequently will be the focus of the subsequent analysis.

3. Data analysis

3.1. Full database characterization

As mentioned above, the sections that will be subject to a more detailed analysis will be H and I. It is also important to highlight that the dataset from 2015 to 2019 of total non-financial corporations presented a total of around 300 thousand companies in each year, of which, based on the criteria already mentioned, 13.73% were considered resilient in 2019.

3.2. Main results

3.2.1 Transportation and Storage

Initial characterization

In 2019, Section H (Transportation and Storage) comprised a total of 14,442 companies, with NACE 49 (Land transport and transport via pipelines) accounting for approximately 86% of the sector under review. It is also noteworthy that, in terms of company size, Section H is primarily composed of micro companies, representing around 86% of the total.

Between 2019 and 2022, 1,382 companies ceased its activity, of which 1,196 were classified as non-resilient (86.54%), of which 1,288 were classified as micro companies (93.20%). During the same period, 165 companies exited this section, while 176 companies entered, resulting in a net increase of 11 companies.

Sectoral Analysis

With respect to resilient companies, they accounted for 21.61% of the total companies within this sector (3,121 companies), a figure notably higher than the overall dataset, where 13.73% of firms were deemed resilient.

A preliminary analysis shows that, overall, section H demonstrated a strengthened position in 2022 compared to 2019, except for liquidity levels, which are below the levels of 2019, although they remain above 100%, as it can be observed in table 1.

General overview 1

	2022	2019	Δ 2019 /2022
Equity to assets ratio	27.11%	23.27%	3.84p.p.
Financial Leverage	58.41%	67.55%	-9.15p.p.
Current Ratio	123.41%	137.74%	-14.33p.p.
EBITDA margin	17.92%	16.53%	1.39p.p.
Financial Debt / EBITDA	3.14x	4.50x	-1.37x

It is also important to note that, when analyzing the section by company size, only micro companies do not follow the aforementioned trend, showing a slight deterioration in the Equity to assets ratio and Financial Leverage (please refer to table 2).

In this regard, medium-sized companies, with a total of 341 in this section, also stand out as they performed worse compared to the rest of the section in 2019. Between 2019 and 2022, this group of companies strengthened its capitalization, reduced debt levels, and improved its

profitability ratios in higher magnitude when compared to the overall section's performance as it can be checked in table 3.

Micro companies			2
	2022	2019	Δ 2019 /2022
Equity to assets ratio	37.11%	40.72%	-3.61p.p.
Financial Leverage	40.34%	38.61%	1.73p.p.
Current Ratio	163.37%	194.70%	-31.33p.p.
EBITDA margin	10.63%	9.39%	1.24p.p.
Financial Debt / EBITDA	3.72x	5.14x	-1.43x

Medium-sized companies			3
	2022	2019	Δ 2019 /2022
Equity to assets ratio	20.73%	11.41%	9.32p.p.
Financial Leverage	69.45%	84.81%	-15.36p.p.
Current Ratio	130.28%	133.26%	-2.98p.p.
EBITDA margin	28.69%	23.04%	5.65p.p.
Financial Debt / EBITDA	3.01x	6.09x	-3.09x

From the perspective of companies considered resilient in 2019, a slightly deterioration was observed across the majority of the ratios considered in this study (as it can be verified in table 4), with large and micro companies being the primary contributors to this trend. This evolution is also observed with a higher magnitude in the companies resilient in 2019 and non-resilient in 2022 as it is showed in table 5.

Additionally, it is important to highlight that of the 3,121 companies classified as resilient in 2019, 49.54% of these companies (1,546) no longer hold this classification after the pandemic period. Within this subset: i) 558 companies reported negative or zero EBITDA in 2022; ii) 92 had negative Equity (of which 71 also reported negative EBITDA); and iii) 186 ceased operations. In terms of company size, the vast majority of entities that lost their resilient status in 2022 were micro companies (1,455). It is noteworthy that, for both micro and large enterprises, approximately half of the companies within these categories lost their resilient status. Conversely, about 26% of medium-sized companies are no longer considered resilient.

When comparing the companies that were resilient in both periods (please refer to table 6), we observe that, across all metrics considered in the study, there was an improvement when compared to 2019.

Resilient companies 2019			4
	2022	2019	Δ 2019 /2022
Equity to assets ratio	69.90%	70.49%	-0.59p.p.
Financial Leverage	15.27%	14.34%	0.93p.p.
Current Ratio	382.41%	419.14%	-36.73p.p.
EBITDA margin	35.44%	37.84%	-2.40p.p.
Financial Debt / EBITDA	0.72x	0.65x	0.07x

Resilient companies 2019 non resilient 2022			5
	2022	2019	Δ 2019 /2022
Equity to assets ratio	54.54%	64.78%	-10.24p.p.
Financial Leverage	32.02%	20.05%	11.96p.p.
Current Ratio	139.14%	242.54%	-103.40p.p.
EBITDA margin	23.87%	34.52%	-10.65p.p.
Financial Debt / EBITDA	2.86x	1.28x	1.58x

Resilient companies 2019 and 2022			6
	2022	2019	Δ 2019 /2022
Equity to assets ratio	77.51%	74.09%	3.41p.p.
Financial Leverage	7.30%	10.82%	-3.52p.p.
Current Ratio	592.64%	561.05%	31.59p.p.
EBITDA margin	39.41%	39.24%	0.18p.p.
Financial Debt / EBITDA	0.28x	0.42x	-0.14x

Analysis of the main subsectors of activity

Although a general favourable trend in ratios can be observed in Transportation and Storage, when looking at the two main NACE's (49320 - Occasional passenger transport in light vehicles and 49410 - Road freight transport), which together account for approximately 80% (49320 – 5,835 companies, 49410 – 6,080 companies) of the companies in this section, an opposite trend was observed between the two.

Thus, while in 2019 the occasional passenger transport sector in light vehicles showed more robust indicators than the ones observed in the road freight transport sector (as can be observed in tables 7 and 8), it is noted that, as a result of mobility restrictions during the pandemic, which led to the temporary suspension of activity for some of these entities, a negative effect was felt among this set of companies. Consequently, at the end of year 2022 it was not possible to recover the levels reached in 2019, despite the relaxation of the restrictive measures that started to happen during the first quarter of 2022, which led to a significant increase in the number of incoming visitors to Portugal and consequently boosted the operational activity of these entities.

Companies from NACE 49320			7
	2022	2019	Δ 2019 /2022
Equity to assets ratio	40.27%	57.70%	-17.43p.p.
Financial Leverage	40.85%	25.00%	15.85p.p.
Current Ratio	272.85%	321.62%	-48.77p.p.
EBITDA margin	10.24%	11.36%	-1.12p.p.
Financial Debt / EBITDA	3.31x	2.45x	0.86x

Companies from NACE 49410			8
	2022	2019	Δ 2019 /2022
Equity to assets ratio	39.19%	39.30%	-0.11p.p.
Financial Leverage	41.71%	41.29%	0.42p.p.
Current Ratio	156.52%	153.23%	3.29p.p.
EBITDA margin	9.54%	8.53%	1.01p.p.
Financial Debt / EBITDA	2.30x	2.51x	-0.21x

It is also worth noting the deterioration of the companies' capitalization in NACE 49320, reflected in a reduction of approximately 17 percentage points in the Equity to assets ratio, which was influenced by the losses recorded in 2020 and 2021, as well as the increase in Financial Leverage. This was driven not only by the reduction in Equity but also by the increase in Financial Debt, since these entities faced a decline in its operational activity and the level of fixed costs did not decrease in the same proportion.

Finally, it should be noted that between 2019 and 2022, the number of companies with negative equity in the occasional passenger transport sector in light vehicles increased by around 37%, rising from 1,291 in 2019 companies to 1,766 in 2022.

Given the aforementioned developments, and in a sector that is predominantly represented by micro and small companies, of the 1,383 companies considered resilient in 2019, only 35.57% of them maintained this classification, highlighting the negative impact of the Covid-19 pandemic, and also the closure of 82 resilient companies.

Regarding NACE 49410, it is observed that this group of companies experienced a less turbulent period during the Covid-19 pandemic, not only because the transport of goods continued to operate with fewer restrictions but also due to the significant growth in online commerce driven by the previously mentioned restrictions, which boosted the operational activity of these transport companies.

As a result, no significant variations were observed in the financial ratios of these companies, although a reduction in the number of companies classified as resilient was observed, albeit to a significatively lesser extent (approximately 64% of companies remained resilient).

It is worth also noting that among the 821 companies in section H that were newly classified as resilient in 2022, around 52% of them are concentrated in the road freight transport sector.

Finally, in what regards the proportion of companies that keep its resilience classification in both periods for the two samples (2016 / 2019 and 2019 /2022) a decrease of -5.98p.p. (from 56.44% to 50.46%) was observed which indicates that company's resilience is affected by Covid-19 pandemics. This situation is more prominent in NACE 49320 where a greater decrease was

observed (from 53.75% to 35.57%), due to the reasons mentioned above. On the opposite side, NACE 49410 observed an increase from 59.44% to 63.85%.

3.2.2 Accommodation and Food Service Activities

Initial characterization

Section I (Accommodation and Food Service Activities) comprised a total of 26,950 companies in 2019, with the majority (approximately 79%) categorized under NACE 56 (Food and Beverage Service Activities). In terms of size classification, the section was predominantly composed of micro-enterprises (82.00%) and small enterprises (15.80%).

Between 2019 and 2022, 3,275 companies ceased operations, of which 3,092 were deemed non-resilient in 2019, with the vast majority being micro-enterprises (2,879). During this period, the sector saw a net increase of 1,431 companies, accounting for 1,716 new entries and 285 exits.

Resilient companies, representing 13.75% of the total sector (3,706 companies), were in line with the overall companies' universe, which also stood at 13.73%.

Sectoral Analysis

Starting with a general analysis, Section I strengthened its position compared to 2019, as shown in Table 9. The EBITDA margin is particularly noteworthy, primarily due to operational factors. Since the mid of the first quarter of 2022, the easing of restrictions allowed Portugal to welcome tourists again, which significantly boosted the sector's performance. As a result, operational metrics, such as the number of overnight stays, nearly returned to 2019 levels, positively impacting the activity of this sector.

General overview 9

	2022	2019	Δ 2019 /2022
Equity to assets ratio	38.63%	35.60%	3.03p.p.
Financial Leverage	51.13%	54.56%	-3.43p.p.
Current Ratio	113.21%	100.18%	13.03p.p.
EBITDA margin	37.20%	26.90%	10.30p.p.
Financial Debt / EBITDA	4.25x	6.46x	-2.21x

When analyzing this section by company size, micro companies demonstrated the strongest performance across all metrics between 2019 and 2022 (see Table 10). However, it is important to note that despite this upward trend, these companies remained below the sector average across all metrics in both 2019 and 2022.

Additionally, large companies were the only group to report a current ratio below 100% in both periods, despite a slight improvement in 2022, as shown in Table 11.

Micro companies			10
	2022	2019	Δ 2019 /2022
Equity to assets ratio	19.87%	13.08%	6.79p.p.
Financial Leverage	70.42%	80.11%	-9.69p.p.
Current Ratio	112.96%	97.63%	15.33p.p.
EBITDA margin	16.30%	8.86%	7.44p.p.
Financial Debt / EBITDA	7.43x	14.74x	-7.32x
Large companies			11
	2022	2019	Δ 2019 /2022
Equity to assets ratio	44.10%	46.78%	-2.68p.p.
Financial Leverage	38.43%	34.83%	3.60p.p.
Current Ratio	94.07%	89.31%	4.76p.p.
EBITDA margin	22.40%	18.51%	3.90p.p.
Financial Debt / EBITDA	1.80x	2.09x	-0.30x

Regarding resilient companies, there was a decline in all the key metrics between 2019 and 2022. The current ratio dropped from 306.99% in 2019 to 292.11% in 2022, primarily driven by micro and small companies. Additionally, financial leverage increased from 16.00% in 2019 to 20.54% in 2022, with contributions from companies of all sizes with exception of large companies. These trends are illustrated in Tables 12, 13, and 14, below.

It is also noteworthy that nearly 72% of the 3,706 companies classified as resilient in 2019 no longer held this status by 2022. Of these 2,658 companies, 183 ceased operations, 450 reported a negative EBITDA in 2022, and 135 had negative equity in at least 2022, with 91 also recording a negative EBITDA. Moreover, it's important to highlight that over 1,300 of these companies had at least one year of negative EBITDA between 2020 and 2021.

In terms of company size, it is observed that the vast majority (2,658) of entities that lost their resilient status in 2022 were micro companies. This category experienced the highest percentage of companies losing resilience (76.04%), followed by large enterprises (62.50%).

Resilient companies 2019			12
	2022	2019	Δ 2019 /2022
Equity to assets ratio	68.33%	72.01%	-3.68p.p.
Financial Leverage	20.54%	16.00%	4.53p.p.
Current Ratio	292.11%	306.99%	-14.89p.p.
EBITDA margin	29.15%	30.11%	-0.96p.p.
Financial Debt / EBITDA	1.12x	0.80x	0.32x

Resilient micro companies 2019			13
	2022	2019	Δ 2019 /2022
Equity to assets ratio	59.43%	65.35%	-5.92p.p.
Financial Leverage	28.72%	22.26%	6.46p.p.
Current Ratio	341.10%	389.23%	-48.13p.p.
EBITDA margin	19.94%	22.45%	-2.50p.p.
Financial Debt / EBITDA	2.00x	1.25x	0.75x

Resilient large companies 2019			14
	2022	2019	Δ 2019 /2022
Equity to assets ratio	72.46%	69.49%	2.97p.p.
Financial Leverage	15.15%	15.72%	-0.57p.p.
Current Ratio	161.70%	133.99%	27.71p.p.
EBITDA margin	41.52%	37.76%	3.76p.p.
Financial Debt / EBITDA	0.48x	0.55x	-0.08x

Regarding the companies that were resilient in 2019 but not resilient in 2022 (Table 15), there is a general decline in the indicators under analysis, although they remain above the section's average. As for the companies that remained resilient in both periods, most of the ratios for 2022 are in line with the values observed in 2019.

Resilient companies 2019 non resilient 2022			15
	2022	2019	Δ 2019 /2022
Equity to assets ratio	60.68%	68.27%	-7.59p.p.
Financial Leverage	27.89%	19.37%	8.52p.p.
Current Ratio	222.23%	258.06%	-35,83p.p.
EBITDA margin	25.07%	27.53%	-2.46p.p.
Financial Debt / EBITDA	1.58x	0.94x	0.63x

Resilient companies 2019 and 2022			16
	2022	2019	Δ 2019 /2022
Equity to assets ratio	77.39%	76.89%	0.49p.p.
Financial Leverage	12.22%	11.73%	0.49p.p.
Current Ratio	415.21%	400.05%	15.16p.p.
EBITDA margin	35.21%	34.51%	0.70p.p.
Financial Debt / EBITDA	0.64x	0.61x	0.03x

Analysis of the main subsectors of activity

Given that this section experienced an overall positive growth between 2019 and 2022 and considering that the NACE codes within this section pertain to similar activities, we have divided the section into its two constituent NACE codes (55 – Accommodation and 56 – Food and Beverage Service Activities) to analyze their respective performance between 2019 and 2022.

As observed in the tables below (17 and 18), both NACE codes exhibited similar trends across the ratios considered in this study. However, it is important to highlight the significant increase in the EBITDA margin within the accommodation sector, whereas the food and beverage service sector maintained its EBITDA margins in line with 2019 levels. The increase in EBITDA obtained in the accommodation sector also contributed to a decrease in the Financial Debt / EBITDA ratio. Lastly, both activities showed similar improvements in the Current Ratio, supporting the overall positive trend within the section.

Companies from NACE 55			17
	2022	2019	Δ 2019 /2022
Equity to assets ratio	38.63%	35.60%	3.03p.p.
Financial Leverage	51.13%	54.56%	-3.43p.p.
Current Ratio	113.21%	100.18%	13.03p.p.
EBITDA margin	37.20%	26.90%	10.30p.p.
Financial Debt / EBITDA	4.25x	6.46x	-2.21x

Companies from NACE 56			18
	2022	2019	Δ 2019 /2022
Equity to assets ratio	28.65%	27.33%	1.33p.p.
Financial Leverage	52.39%	51.71%	0.69p.p.
Current Ratio	126.91%	108.14%	18.77p.p.
EBITDA margin	9.95%	9.22%	0.73p.p.
Financial Debt / EBITDA	2.44x	2.21x	0.23x

Despite the positive performance of the NACE codes under analysis between 2019 and 2022, a significant reduction in the number of resilient entities was observed in both during this period. Specifically, in NACE 55, only 38.93% of companies maintained their resilient status in 2022, while in NACE 56, this decline was even more pronounced, with only 23.56% of companies achieving resilient status in 2022. This trend is largely attributed to the negative impact of the COVID-19 pandemic, as many entities experienced negative results due to restrictions on both domestic and international movement. Some entities even closed their facilities temporarily, with some undertaking renovation projects.

Finally, it is worth noting that in 2022, 342 companies became classified as resilient, with 55.85% of these entities belonging to NACE 56.

As in section H, an impact by Covid-19 pandemics was observed in a greater magnitude, since the decrease in the companies that were resilient from 2016 to 2019 and 2019 to 2022 was around 40p.p. (from 68.27% to 28.28%).

4. Conclusion

Through the study that we conducted on the two most impacted sectors by the COVID-19 pandemic, based on the financial statements of the companies (the primary source of information of this study) it can be observed that there was a significant decline in the number of resilient companies between 2019 and 2022 (Section H: 49.54% reduction, and Section I: 71.72% reduction). It can also be noted the decline of the proportion of companies that kept its resilience status in the two samples considered. This situation was exacerbated by the years 2020 and 2021, as a significant portion of these entities operated under furlough schemes and, in some cases, temporarily ceased operations, leading to negative results. Nonetheless, it is important to highlight that with the easing of restrictions on the movement of people in 2022, there was an increase in tourist inflows into the country. Coupling this with high inflation, which was largely passed on to the end consumer, this resulted in an increase in operational activity for companies within these two sectors, allowing them to generally recover the levels observed in 2019.

Bibliography

European Comission (2003) *Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises.* Obtained from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32003H0361European Comission (2003).

Martins A. and Santos M. (2022) *The economic and financial resilience of Portuguese companies in the face of exogenous shocks: the COVID-19 pandemic and the invasion of Ukraine.* Obtained from: https://www.gee.gov.pt/pt/?option=com_fileman&view=file&routed=1&name=TE%20105%20-%20Resili%C3%AAncia%20das%20empresas%20portuguesas.pdf&folder=estudos-e-seminarios%2Ftemas-economicos&container=fileman-files

Barbosa, L. and de Pinho, P. (2016) *Corporate financing structure.* Obtained from: https://www.bportugal.pt/sites/default/files/anexos/papers/re201601_p.pdf

Carvalho, A.O., Ribeiro, I., Cirani, C.B.S., Cintra, R.F. (2016). *Organizational resilience: a comparative study between innovative and non-innovative companies based on the financial performance analysis.* Obtained from: https://www.redalyc.org/journal/4991/499151079006/499151079006.pdf

Banco de Portugal (2020 and 2021) *Rapid and Exceptional Survey of Enterprises.* Obtained from: https://www.bportugal.pt/page/quais-os-impactos-da-covid-19-nas-empresas

S&P Global Ratings' methodology for rating corporate industrial companies and utilities. Obtained from: S&P Global Ratings (spglobal.com)



HOW RESILIENT COMPANIES NAVIGATED THROUGH COVID-19 PANDEMICS

WORKSHOP ON "NEW INSIGHTS FROM FINANCIAL STATEMENTS"

17 OCT. 2024| MADRID

ANTÓNIO COSTA | PEDRO CORDEIRO

BANCO DE PORTUGAL
EUROSISTEMA

Introduction| **Scope of the study**

- Primary goal: Analyze how resilient companies reacted to the Covid-19 pandemics shock
- Main source: Financial statements, since they are critical components for economic and financial analysis, providing insights into a company's performance and financial health
- Sections analyzed: The two most affected sectors in Portuguese economy “Transportation and Storage” and “Accommodation and Food Service Activities”

Dataset| **Definition**

- Primary source: Annex A of the Simplified Corporate Information (IES)
- Dataset timespan
  - Main dataset: From 2015 to 2022
  - Second dataset: From 2012 to 2019
- Companies included: Non-financial corporations (S11), with financial statements available for each year in each dataset
- Rationale: This period provides a comprehensive overview of the companies' financial health prior to the pandemic, which is essential for assessing their resilience according to the defined criteria

Methodology | **Resilience**

Ratios

- Equity to assets ratio = $\frac{Equity}{Total\ Assets}$
- Financial Leverage = $\frac{Financial\ Debt}{Financial\ Debt+Equity}$
- Current Ratio = $\frac{Current\ Assets}{Current\ Liabilities}$
- EBITDA margin = $\frac{EBITDA}{Revenue}$
- Financial Debt / EBITDA

Resilience Criteria

- Last two years of each dataset, company must have a **positive EBITDA**. Additionally, at least in one of the other three years of each dataset one more positive EBITDA must be recorded;
- In each year a **positive score will be assigned if a company presents all of its ratios equal or above the weighted average of the sector** in which it operates and:
  - **Current ratio** must be higher than 100%;
  - Entities that have **financial debt 0**, will be considered as being above the sector in the ratios in which this variable is used.

Methodology

If a company presents a **positive score in the last two years of each dataset and one more on the other three years** considered and fulfil the EBITDA criteria the company is considered resilient.

Main results| **General Overview**

Evolution of all companies in Section H

Transportation and Storage			
	2022	2019	Δ 2019 /2022
Equity to assets ratio	27.11%	23.27%	3.84p.p.
Financial Leverage	58.41%	67.55%	-9.15p.p.
Current Ratio	123.41%	137.74%	-14.33p.p.
EBITDA margin	17.92%	16.53%	1.39p.p.
Financial Debt / EBITDA	3.14x	4.50x	-1.37x

- **Strengthened position** in 2022 when compared to 2019
- **Liquidity levels are the exception**, although remain above 100%

Evolution of all companies in Section I

Accommodation and Food Service Activities			
	2022	2019	Δ 2019 /2022
Equity to assets ratio	38.63%	35.60%	3.03p.p.
Financial Leverage	51.13%	54.56%	-3.43p.p.
Current Ratio	113.21%	100.18%	13.03p.p.
EBITDA margin	37.20%	26.90%	10.30p.p.
Financial Debt / EBITDA	4.25x	6.46x	-2.21x

- **Strengthened position** in 2022 when compared to 2019
- **EBITDA margin** is particularly noteworthy, primarily due to operational factors. Since the mid of the first quarter of 2022, the easing of restrictions allowed Portugal to welcome tourists again, which significantly boosted the sector's performance

Main results | **Main subsectors of activity section H**

Occasional passenger transport in light vehicles			
	2022	2019	Δ 2019 /2022
Equity to assets ratio	40.27%	57.70%	-17.43p.p.
Financial Leverage	40.85%	25.00%	15.85p.p.
Current Ratio	272.85%	321.62%	-48.77p.p.
EBITDA margin	10.24%	11.36%	-1.12p.p.
Financial Debt / EBITDA	3.31x	2.45x	0.86x

Road freight transport			
	2022	2019	Δ 2019 /2022
Equity to assets ratio	39.19%	39.30%	-0.11p.p.
Financial Leverage	41.71%	41.29%	0.42p.p.
Current Ratio	156.52%	153.23%	3.29p.p.
EBITDA margin	9.54%	8.53%	1.01p.p.
Financial Debt / EBITDA	2.30x	2.51x	-0.21x

- More robust indicators in 2019 compared to Road freight transport subsector
- Covid-19 negatively affected this subsector, not being able to recover the 2019 levels in 2022

- Subsection less affected by Covid-19, since it continued its operations with fewer restrictions and due to the significant growth in online commerce
- No material changes observed between the periods considered

Main results| **Resilient companies 2019**

Transportation and Storage			
	2022	2019	Δ 2019 /2022
Equity to assets ratio	69.90%	70.49%	-0.59p.p.
Financial Leverage	15.27%	14.34%	0.93p.p.
Current Ratio	382.41%	419.14%	-36.73p.p.
EBITDA margin	35.44%	37.84%	-2.40p.p.
Financial Debt / EBITDA	0.72x	0.65x	-0.07x

Accommodation and Food Service Activities			
	2022	2019	Δ 2019 /2022
Equity to assets ratio	68.33%	72.01%	-3.68p.p.
Financial Leverage	20.54%	16.00%	4.53p.p.
Current Ratio	292.11%	306.99%	-14.89p.p.
EBITDA margin	29.15%	30.11%	-0.96p.p.
Financial Debt / EBITDA	1.12x	0.80x	0.32x

- **Slightly deterioration** observed across the majority of the ratios considered
- **Evolution was softened** by the companies that remained **resilient in both periods,** since they **strengthened** its position **across all metrics** between 2019 and 2022
- A **decline** can also be observed **in all the key metrics between 2019 and 2022**
- The companies that remained **resilient in both periods, kept its ratios in line** in 2019 and 2022

Main results| **Resilient companies**

Evolution of resilient companies in Transportation and Storage sector

		2019		
		Resilient	Non-Resilient	Total
2016	Resilient	1,774	1,369	3,143
	Non-resilient	980	10,036	11,016
	Total	2,754	11,405	14,159

		2022		
		Resilient	Non-Resilient	Total
2019	Resilient	1,575	1,546	3,121
	Non-resilient	821	10,500	11,321
	Total	2,396	12,046	14,442

Evolution of resilient companies in Accommodation and Food Service Activities

		2019		
		Resilient	Non-Resilient	Total
2016	Resilient	1,493	694	2,187
	Non-resilient	1,265	20,029	21,294
	Total	2,758	20,723	23,481

		2022		
		Resilient	Non-Resilient	Total
2019	Resilient	1,048	2,658	3,706
	Non-resilient	342	22,902	23,244
	Total	1,390	25,560	26,950

- In transportation and storage, the **proportion of companies that kept its resilience status** in both periods for the two samples **decrease around 6 p.p.** (from 56.44% to 50.46%)
- In accommodation and food service the **proportion of companies that kept its resilience status** in both periods for the two samples **decrease around 40 p.p.** (from 68.27% to 28.28%)
- Covid-19 pandemics impacted Portuguese non-financial companies' resilience

Conclusions

Did the COVID-19 pandemic impacted companies' performance?

- **Significant decrease in resilient companies** in the two most impacted sectors by the COVID-19 pandemic between 2019 and 2022 (Transportation and storage: - 49.54%, and Accommodation and food service activities: - 71.72%

- **Decline of the proportion of companies that kept its resilience status in the two samples considered**. This situation was exacerbated by the years 2020 and 2021, since a significant portion of these entities operated recurred to layoff programs and, in some cases, temporarily ceased operations, leading to negative results

- **Recovery in 2022 of the levels observed in 2019**: with the easing of restrictions on the movement of people in 2022, there was an increase in tourist inflows into the country. Coupling this with high inflation, which was largely passed on to the end consumer, this resulted in an increase in operational activity for companies within these two sectors

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

The pass-through of cost shocks to firms' prices and the impact on value added[1]

Fabio Parlapiano,
Bank of Italy

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

The pass-through of cost shocks to firms' prices and the impact on value added

Fabio Parlapiano

Abstract

The post-pandemic surge in intermediate goods and energy prices has intensified scrutiny of firms' pricing policies for their role in feeding the inflationary dynamics and accommodating the sequence of cost shocks. This work combines firm-level balance sheet and price data to estimate the contributions of prices (relative changes in output and intermediate input prices) and quantities (relative changes of output and input quantities) to the nominal growth rate of value added. We document that between 2016 and 2022 the transmission of intermediate input price changes to firms' output prices was less than one-to-one, resulting in a negative contribution of pricing policies to aggregate value added. The primary driver of value added growth, including its strong surge in the aftermath of the pandemic shock, is instead the dynamics of quantities. Moreover, estimates suggest that, at the individual level, a greater transfer of input price increases to output prices is associated to higher value added for the firm; this relationship is stronger for large corporates and for those more dependent on intermediate inputs. Moreover, we document that pricing policies provide a limited hedge against fluctuations in input prices: firms adjust their prices in response to idiosyncratic cost shocks but these adjustments do not result in an increase in value added.

Keywords: firms' price setting, operating pass-through, economic margins.

JEL classification: G00, G30, L1.

Contents

The pass-through of cost shocks to firms' prices and the impact on value added 1

1. Introduction and main findings 2

2. Modelling framework 4

3. Data 8

4. Results 12

5. Results 17

6. References 17

1. Introduction and main findings

The resurgence of inflation in the years 2021-22, in the context of cost-push shocks due to the energy crisis, supply chain disruptions and the post-pandemic pent-up demand, prompted renewed interest in firms' pricing policies. Firms' prices represent mostly what consumer pay, thus their role in the propagation of the inflationary dynamic is important. At the same time, when pricing policy are set as a spread over marginal costs, they define the extent firms' bear (or share) the burden of soaring production costs with their customers; thus they can be viewed as an operational hedging tool against cost-shocks.

This paper analyses the relationship between firms' pricing policies and their value added dynamic, highlighting the contributions made by changes in prices and quantities to the growth rate of value added. The analysis is based on detailed firm-level survey data on output and intermediate input price changes, which are available for a sample of Italian non-financial firms. These data are used to decompose the annual variation of value added into three additive components: i) pricing policy, ii) intermediate input productivity and iii) the interaction between the two. Differently from metrics derived from national accounts, firm-level analysis helps explaining heterogeneities depending on firms' exposure to input price shocks, firms' reliance on intermediate input and their capacity to shift a price shock along the supply chain to consumers' markets.

We complement the evidence derived from our decomposition with regression analysis of the response of firms' pricing policies and value added dynamics to idiosyncratic intermediate input price shocks. This enables us to characterise the economic resilience of firms, in terms of their cost-transfer capacity to output prices.

The analysis is based on Italian firms' survey data for the 2016-22 period drawn from the Bank of Italy's Survey on Inflation and Growth Expectations (SIGE). The sample covers mostly unlisted companies with at least 50 employees in the manufacturing, service and construction sectors. It accounts for approximately 1,500 observations per year. Furthermore, we utilise company balance sheet data from the Cerved database to gather firms' revenues, intermediate costs, and value added. It is important to note that the use of value added, defined as the difference between revenues and the cost of intermediate goods and services, allows for a nuanced understanding of economic margins. In particular, it enables for the identification of the role played by volatile and generally large shares of total costs: in the euro area, the share of intermediate goods and energy costs over total costs ranges from 80 to 60 percent in the case of manufacturing and service sectors, respectively. Although value added is highly correlated with operating profit, such as gross operating profits, unit profits or mark-ups measures, it differs from operating profit; the latter accounts for all firms' cost factors (since it takes into account also personnel expenses).

Previous studies analyzed the relationship between firms' pricing behaviour and their economic margins and, related to our work, two complementary research questions have been investigated: first, how does an inflationary environment affect firms' profits and, second, how do profits feedback into the inflationary dynamics.

The former strand of research suggests that when inflation is high profits tend to expand. However, if inflation is particularly high, profits decelerate and eventually

drop. Andler and Kovner (2022) use quarterly firm-level data for US listed companies to show that in the 2021-22 period profit margins growth was more pronounced in industries with higher inflation. However, in the long run, the authors' document that in high inflation regimes, firms' gross profits growth weakens and, eventually, falls. These results are consistent with the intuitive mechanism documented in Moore (1983), that is, differences in the pace of output and input price variations lead profits to expand only when output prices grow faster; this is often the case in not too high inflation regimes.

The second strand investigates if and to what extent firms' pricing policies, measured via mark-ups, the premium above the total cost that provides the seller with a profit, contribute to inflation. The guiding narrative is that inflation is linked to, or perhaps caused by, enterprises charging higher mark-ups, when raising prices above what would be justified by the increase in expenses. In the face of unprecedented increases in energy and intermediate input prices as a result of 2020-22 shocks, the expected drop in firms' profits did not materialize and, in contrast, non-financial firms operating profits rebounded substantially. This situation raised some concerns about corporations increasing prices more than their costs, a mechanism that would intensify inflation pressures.

Empirical evidence, however, indicates that the surge in inflation during the specified period cannot be attributed to pricing policies, as mark-ups mostly remained stable or returned to pre-pandemic levels. In Italy, Colonna et al. (2023) utilized national accounts data to demonstrate that despite the growing proportion of gross operating surplus in relation to value added (i.e., the profit share), mark-ups remained constant overall. Similarly, Leduc et al. (2024) provided comparable evidence for the United States. Additionally, Ganapati et al. (2020) and Champion et al. (2023) revealed that in both the United States and Australia: a) the transmission of cost-shocks to output prices generally is incomplete, resulting in approximately a 0.7 percent increase in output prices for every 1 percent rise in costs, and b) contrary to the expectations of competitive markets, the pass-through effect is lower in industries with higher levels of concentration.

In the cross-section of industries, high mark-up act as a buffer against fluctuations in consumer prices due to economic cycle. Firm level data for the Euro area (Kouvavas et al., 2021) show that inflation in sectors with high mark-up displays a lower sensitivity to shocks (oil supply, global demand and monetary policy). It can be argued that firms with greater market power are less likely to pass on cost-shocks to end prices. Similarly, Kharroubi et al. (2023) document a diminished impact of cost-push shocks (global oil supply shocks) to producers' price inflation in sectors where firms apply higher mark-ups. However, not all shocks are alike and high mark-up provide little cushion against prices pressures stemming from inflationary oil shocks. Conversely, dis-inflationary shocks are passed on to consumers more weakly by high mark-up firms; which are instead more likely to increase their revenues, and hence their profits.

Our work makes several contributions to the existing literature. Firstly, unlike previous studies, we specifically consider the dynamics of intermediate input prices and, importantly, its heterogeneity across the spectrum of firms. By using granular firm-level data on intermediate input and output prices we are able to analyse how changes in prices impact firms' value added, while also recognizing that these impacts may vary among firms. Secondly, although our sample is limited to a subset of Italian

non-financial corporations, we use survey weights to ensure that our findings are representative of the entire population, thus allowing for generalization to the broader Italian business landscape. Lastly, the use of survey data provides us with a unique advantage in examining how firms' pricing decisions are influenced by cost-shocks, defined as differences between actual and expected changes in intermediate input prices.

The results of our analysis suggest that firms' pass-through of intermediate input to output prices is, on average, incomplete. A significant portion of cost increases are borne by firms, with no industry or size class exhibiting complete pass-through. The contribution of pricing policies (made by both changes in output and input prices) to the annual growth rate of value added is minimal and, generally, negative. However, the years 2021-22 indicate a notable shift in the relative contribution of prices. Through a higher operating pass-through, the pricing policies of large firms provide them with only a limited shield against cost hikes. Lastly, we document how pricing policies respond to adverse cost shocks, i.e. unexpected cost increases, yet firms' value added does not improve significantly. This suggests that firms strive to uphold their existing profitability levels and opt to share a portion of the economic losses with either end consumers or businesses in order to retain their market share.

The rest of the paper proceeds as follows. Section 2 sets out the modelling framework and section 3 describes the data. Section 4 discusses the decomposition of value added growth rate at various levels of aggregation (for the economy as a whole, by firm size class and sector); the paragraph then illustrates results from regression analysis. Section 6 concludes.

2. Modelling framework

This section outlines a decomposition of the nominal growth rate of firm-level value added yielding three additive components: a pricing policy or 'price' component, an intermediate input productivity or 'quantity' component and the interaction between the two 'price and quantity'. Descriptive evidence on the relative importance of these components is provided in time series for the whole sample of firms as well as by size groups and sectoral affiliation.

We use then these components in a regression setting, to estimate the extent to which changes in a firm's pricing policies correlate with its value added dynamic; while controlling for firm and time effects.

Lastly, we investigate how a price shock to intermediate input (being energy or shortages in materials, for instance) translates in a change in a firm's pricing policy and how it affects its prospective value added growth rate. To this end, idiosyncratic price shocks are used as an instrument to retrieve exogenous changes in a firm's pricing policy.

2.1 Price and quantity contributions to value added

To single out the contribution of pricing policies to the dynamics of value added, we decompose firms' nominal growth rates of revenues, intermediate costs and value added into a price and quantity components (holding previous period quantities and prices fixed, respectively). Firms' reported change in input and output prices over the past 12 months and the realized revenues and intermediate costs growth rates are used for this purpose.

Formally, the annual growth rate (Δ) of revenues can be decomposed as follows:

$$\Delta Rev_{t,\,t-1} = \frac{p^o_t q^o_t - p^o_{t-1} q^o_{t-1}}{p^o_{t-1} q^o_{t-1}} \tag{1}$$

$$\Delta Rev_{t,\,t-1} = \Delta p^o_{t,\,t-1} + \Delta q^o_{t,\,t-1} + \Delta p^o q^o_{t,\,t-1}$$

where, $\Delta p^o_{t,\,t-1}$ is the (observable from survey data) firms' reported change in output prices between t and t-1, $\Delta q^o_{t,\,t-1} = \left[(1+\Delta Rev_{t,t-1})/\left(1+\Delta p^o_{t,\,t-1}\right)-1\right]$ is the implied change in output quantities and $\Delta p^o q^o_{t,\,t-1} = \left[\Delta p^o_{t,\,t-1} * \Delta q^o_{t,\,t-1}\right]$ is the interaction between price and quantity components. Similarly, the annual growth rate of intermediate costs ($\Delta Cost_{t,\,t-1}$) can be decomposed in an intermediate input price changes $\left(\Delta p^i_{t,\,t-1}\right)$, a quantity and an interaction components as in Eq. (1). Rearranging the revenues and intermediate costs decompositions yields, in short, the following decomposition for the growth rate of value added:

$$\Delta VA_{t,t-1} = \frac{VA_t - VA_{t-1}}{VA_{t-1}} = \frac{1}{VA_{t-1}}\left[(\Delta Rev * Rev_{t-1}) - (\Delta Cost * Cost_{t-1})\right]$$

$$\Delta VA_{t,\,t-1} = \Delta p^{VA}_{t,\,t-1} + \Delta q^{VA}_{t,\,t-1} + \Delta pq^{VA}_{t,\,t-1} \tag{2}$$

where: the contribution made by firms' pricing policies (both output and intermediate input prices), productivity and the interaction term between prices and quantities are defined in Eq. (3) to (5).

$$\Delta p^{VA}_{t,t-1} = \frac{1}{VA_{t-1}}\left[\left(Rev_{t-1} * \Delta p^o_{t,\,t-1}\right) - \left(Cost_{t-1} * \Delta p^i_{t,\,t-1}\right)\right] \tag{3}$$

$$\Delta q^{VA}_{t,t-1} = \frac{1}{VA_{t-1}}\left[\left(Rev_{t-1} * \Delta q^o_{t,\,t-1}\right) - \left(Cost_{t-1} * \Delta q^i_{t,\,t-1}\right)\right] \tag{4}$$

$$\Delta pq^{VA}_{t,t-1} = \frac{1}{VA_{t-1}}\left[\left(Rev_{t-1} * \Delta p^o_{t,\,t-1} * \Delta q^o_{t,\,t-1}\right) - \left(Cost_{t-1} * \Delta p^i_{t,\,t-1} * \Delta q^i_{t,\,t-1}\right)\right] \tag{5}$$

The contribution of prices (3) and quantities (4) to the nominal value added growth rate is thus pinned down by three factors: (i) the difference between output and input price changes ($\Delta p^{o}_{t,\,t-1} - \Delta p^{i}_{t,\,t-1}$) a proxy for a firm's pricing policy (PP) or, equivalently, its operating pass-through ratio; (ii) the difference between output and intermediate input quantity changes ($\Delta q^{o}_{t,\,t-1} - \Delta q^{i}_{t,\,t-1}$), a proxy for the productivity of intermediate inputs (IIP) and; (iii) the difference between revenues and intermediate input costs ($Rev_{t-1} - Cost_{t-1}$) a proxy for the intensity of intermediate inputs (III). Under the assumption of a positive elasticity of intermediate goods and services to output quantities, for firms with low intermediate input intensity, intermediate costs per unit of revenues weigh less and even a substantial increase in their price could have little impact on marginal costs.

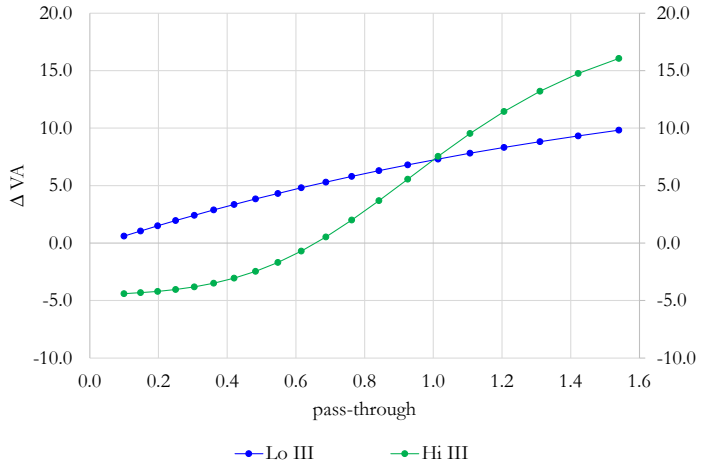
Figure 2 illustrates the comparative statics of the relationship between firms' pricing policies, measured for ease of interpretation by the ratio between output and input changes, and the growth rate of value added. It is assumed that the share of intermediate input used to produce 1 unit of output is 0.1 and 0.9, respectively. These two polar cases demonstrate how a firm's economic performance is influenced by the interaction of its production function and pricing policy; output quantities are held constant thus changes in value added are driven exclusively by changes in pricing policies.

As the figure suggests, an increase in value added is compatible with pass-through below one -in contrast to what perfect competition hypothesize- especially when firms' intensity in the use of intermediate input is low. Nonetheless, even for firms with high intermediate input intensity positive changes in value added are achievable when the pass-through reaches a certain threshold (e.g. 0.6 in our case). It also noteworthy that firms with high III experience a more substantial impact on their value added as a result of changes in pass-through; this is evident from the steeper slope of the curve (we will point out this particular feature in the empirical section).

In the context of the extraordinary increase in firms' value added over the years 2021-22 and consequential policy debate regarding its sources, Figure 1 adds two elements to this discussion that are helpful to note: (1) evidence of firms' higher profits is not unambiguously informative of greedy pricing policies (firms' passing on their cost increases to customers more than one-to-one). In turn, pass-through lower than one are compatible with value added and profits growth, especially if the intensity of intermediate input is not too high and productivity gains (or, equivalently, changes in the input mix) allow savings in the use of intermediate inputs; (2) high (low) intensity of intermediate input can amplify (dampen) the response of value added growth rate to changes in pricing policies and, equivalently, gains (losses) in productivity.

Change in pass-through and value added

Figure 1



[1] Figure Note: The figure illustrates the relationship between a firm's pricing policy, measure by the ratio between output and intermediate input price changes (a proxy for pass-through) and value added for two hypothetical firms having low and high intermediate input intensity. Low (high) intermediate input intensity are defined by setting intermediate input quantities equal to 0.1 (0.9) for each unit of output.

Sources: Our elaboration.

2.2 Response of firms' pricing policy to cost-shocks

The decomposition of value-added growth rate presented in the previous section provides a basis for investigating how changes in a firm's pricing policy affects its economic performance. This is achieved through the use of reduced-form approach in a panel data regression setting, wherein firm and time effects are controlled for.

Investigating this relationship in causal setting is a challenging task due to endogeneity concerns which include: (i) simultaneity bias, whereby variations in value added are jointly determined by firms' pricing policies (see Eq. 2); and (ii) omitted variables bias, whereby third factors could cause both changes in pricing policies and value added. These factors include, for instance, a firm's market power, the extent of competition or exposure to international markets.

We start from estimates of a parsimonious semi-identity model in (6):

$$\Delta VA_{i,t,t-1} = \alpha + \beta_1 PP_{i,t,t-1} + \beta_2 IIP_{i,t,t-1} + \gamma_i + \tau + \varepsilon_{i,t,t-1} \quad (6)$$

where: $\Delta VA_{i,t,t-1}$ is the firm-level value added growth rate, $dPP_{i,t,t-1}$ is the firm's pricing policy measured by the difference between output and input price changes ($\Delta p_t^o - \Delta p_t^i$), and $IIP_{i,t}$ is the firm intermediate input productivity measured by the difference between output and input quantity changes ($\Delta q_t^o - \Delta q_t^i$) term, γ_i and τ are firm and time fixed effects. Simultaneity and omitted variables bias, are addressed in (7-8) using instrumental variable (IV) approach where idiosyncratic shocks to intermediate input costs occurred in the t-1 period ($Z_{i,t-1}$) are used in the first stage to isolate exogenous variations in a firm's pricing policy (see Figure 2):

$$\Delta VA_{i,t,t-1} = \alpha + \delta \widehat{dPP_{i,t,t-1}} + \gamma_i + \tau + \varepsilon_{i,t,t-1} \qquad \text{(7, second stage)}$$

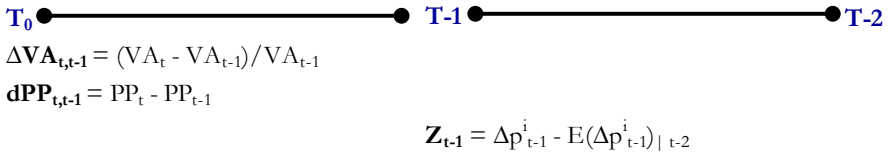
$$dPP_{i,t,t-1} = \alpha + \gamma Z_{i,t-1} * X_i + \gamma_i + \tau + \mu_{i,t,t-1} \qquad \text{(8, first stage)}$$

where: $Z_{i,t-1}$ is the lagged value of the difference between the realized $\Delta p^i_{i,t,t-1}$ and expected $\widehat{\Delta p}^i_{i,t,t-1}$ intermediate input price changes between t and t-1, and X_i is a vector of firms' characteristics which is used to test differences non-monotonicity in the response of PP.

The definition of shock that we follow aligns with the concept of surprise or unexpected change. The instrument is relevant under the assumption that firms adjust their current pricing policy following input price forecasting errors made in the previous period; in turn the correlation between firm-level input cost shocks and the change in value added is about 0.3. Exogeneity rests on both the independence between the instrument and the outcome variable and the assumption that effects on the outcome occur only via the endogenous variable (exclusion restriction). A nearly null correlation between the instrument and the outcome is suggestive that the former condition may hold. The latter condition implies that the input prices forecasting error made by a firm impacts its value added changes only via firm-specific changes to its pricing policy. Due to the idiosyncratic nature of the shock this assumption seems reasonable; in contrast, if the instrument involved a generalized cost shock, the impact on a firm value added could occur also via other channels such as market demand or supply changes.

Timeline of variables used in the regression

Figure 2



[1] Figure Note: The figure illustrates the main variables used in the regression and their temporal references.

Sources: Our elaboration.

3. Data

Information on firm-level prices is sourced from SIGE. The survey is conducted on a quarterly basis and collects information on consumer inflation expectations over different time horizons, changes in their own purchases and selling prices, as well as firms' sentiment on aggregate cyclical developments. The sample is representative of

medium-sized and large Italian firms is obtained using a stratified random sampling approach, which combines economic sector, firm size and geographical area information as strata. In order to derive aggregate statistics, a weighting coefficient is provided for each sample unit. This accounts for the ratio between the number of respondent companies and the number of companies in the reference universe.

The primary data for our analysis is derived from survey responses to the following questions:

- *What was the average variation of output price for the products or services sold by your company over the last 12 months?*
- *What was the average variation of intermediate input price for the product or services purchased by your company over the last 12 months?*

The SIGE sample was matched with balance-sheet information from Cerved on revenues and value added. The resulting dataset (Table 1) includes observations for the period 2016-22.[1]

Descriptive statistics

Table 1

	N. firms	Δ output price	Δ intermediate input price	Δ revenues	Δ intermediate costs	Δ value added
2016	586	0.16	0.94	3.52	4.09	3.56
2017	580	0.91	1.96	5.78	7.55	4.04
2018	628	0.91	1.69	3.15	4.18	2.07
2019	670	0.63	1.16	0.33	1.01	-1.44
2020	836	0.17	0.69	-8.76	-7.50	-10.12
2021	992	3.12	6.94	19.55	19.24	22.37
2022	867	6.67	11.48	16.52	19.39	12.49

[1] Table Note: The table reports the number of firms included in the matched SIGE-Cerved dataset and weighted averages (using survey weights) of relevant variables. Percentage changes (Δ). Value added is the difference between revenues and intermediate costs.

Sources: Bank of Italy's SIGE, Cerved.

The extent to which changes in intermediate input prices are passed on to customers via output prices, also defined as the operating pass-through, can be illustrated by the ratio between the overall 12-month output and intermediate input price variations. This is a simple upper bound proxy that does not account for a number of factors that would be included in a fully-fledged model of firms' price

[1] The years prior to 2016 were not considered in the analysis because intermediate input cost data were not collected.

setting behaviour, such as the state- and time-dependence features documented in Riggi and Tagliabracci (2022).[2]

In the 2016-23 period, the average firms' operating pass-through was incomplete. For a 1 percent change in intermediate input prices, only approximately 0.6 percent was passed on to sale prices in the same year (Figure 3 – panel a). Despite facing high inflation due to energy shock and supply chain disruptions, the average increase in output prices did not exceeded the increase in input intermediate input costs.[3] In late 2022, when energy price pressures abated, the average pass-through rebounded to its pre-pandemic level. This pricing pattern is in line with previous empirical evidence regarding two key aspects: (a) the persistence of output price changes, where future adjustments closely mirror past output price trends (rather than responding immediately to input price changes); and (b) the potential non-linear nature of firms' pricing policies, which depend on the nature of cost-shocks. In cases of inflationary cost-shocks, firms may choose to absorb some of the negative impact by reducing pass-through to customers (Kharroubi et al., 2023). Another noteworthy aspect of the price data is the reduced dispersion in operating pass-through since the onset of monetary tightening in early 2022. The reduced heterogeneity in firms' pricing policies, following a cost-push inflationary shock, lends support to the role of the input cost channel. This is the case when firms face similar changes in their costs and variations in their sale prices also tend to align.

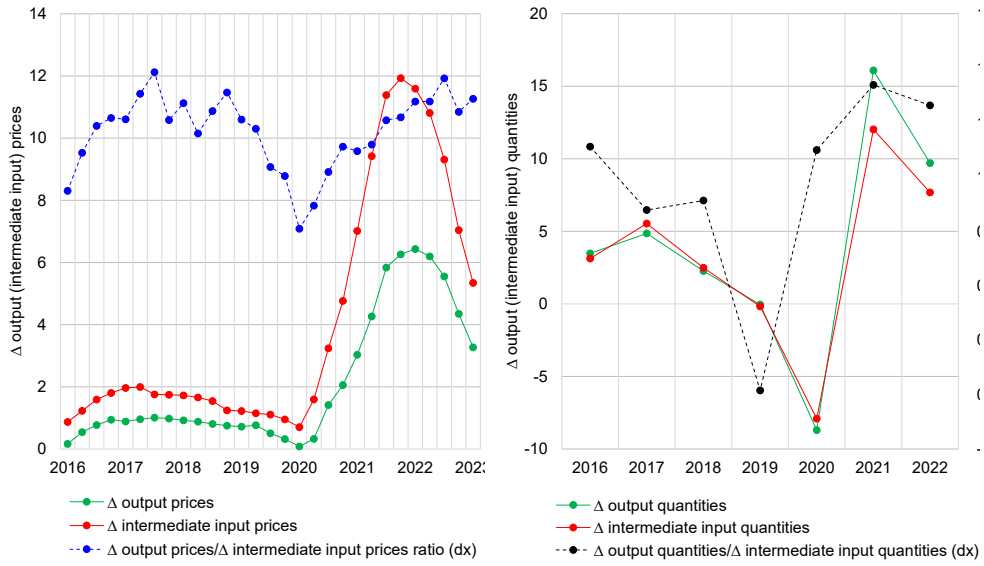
Firms' operating pass-through and intermediate input productivity

Figure 3



[2] The probability of a firm adjusting its prices can vary with *(i)* time and *(ii)* the macroeconomic or firm-specific conditions. The latter case belongs to the theoretical modelling framework of state-dependent models: changes in firm-level costs for intermediate products, revenues, utilization capacity are found to be important state variables (Lein, 2010). Empirical studies that look at the importance of both state- and time-dependent factors for firm-level price setting see also Dixon and Grimme (2022), and Riggi and Tagliabracci (2022).

[3] Riggi and Tagliabracci (2022) find that the pass-through of input costs to output prices is quantitatively limited over the period 2017-22: a one per cent increase in the prices of intermediate input leads to roughly a 0.2 per cent rise in Italian firm's selling prices.



[1] Figure Note: panel a) illustrates the reported 12-month percentage change in output and intermediate input prices and the ratio between the two for the whole sample of firms included in the SIGE; panel b) reports estimated output and intermediate input quantities and the ratio between the two for the sample of firms included in SIGE and Cerved. Percentage changes in output/input quantities are derived according to Eq. (1.b).

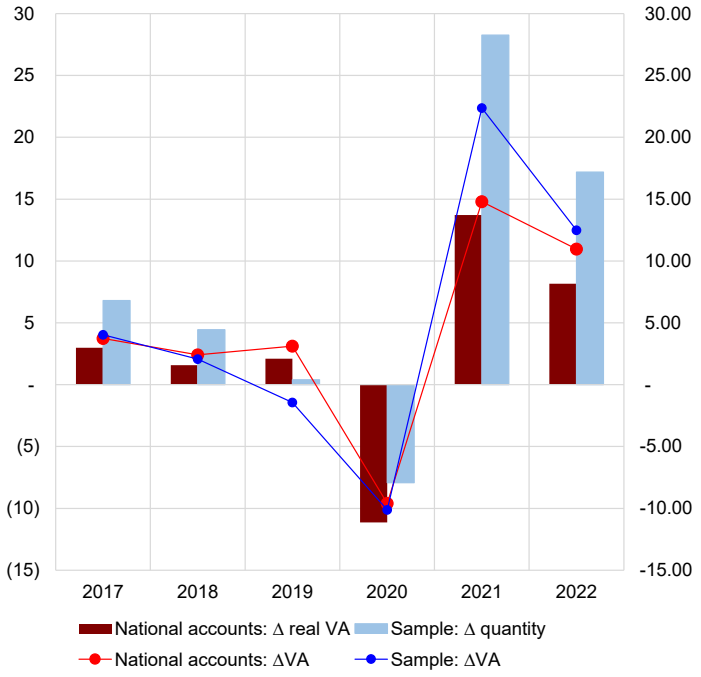
Sources: Bank of Italy's SIGE, Cerved.

Developments in output and intermediate input quantities, as implied by equation 1.b using both firms' prices and balance-sheet information, indicate that their variations were greatly aligned until the onset of the pandemic (Figure 3 – panel b). This suggests relative stability in the intensity with which intermediate input enter the production function. Subsequently, there was a shift in this pattern, with output quantities increasing at a faster rate than intermediate inputs; thus resulting in a higher level of productivity for intermediate inputs above pre-pandemic levels.

Our reference measure for firms' economic performance is represented by value added (VA). Unlike the gross operating margin, which provides a measure of firms' operating profits, VA is gross of labour costs. Due to the relative stickiness with which labour costs adjust to the consumers' price dynamics, changes in VA and operating profits are highly correlated. From SIGE, we are able to directly observe intermediate input price dynamics and, as a result, to obtain accurate information to decompose VA without having to resort to other estimates or proxies. Despite these advantages, due to the features of the survey, firms report the average input and output price changes across the range of products or services used or sold. While this may simplify the task of questionnaire respondents, it introduces some layers of inaccuracy as individual product price paths may differ substantially from the average.

To put our matched SIGE-Cerved sample into context, Figure 4 illustrates the nominal and real growth rate of value added from the national accounts and our sample, respectively. The two series are very close in level and their correlation exceeds 0.95 over the period, suggesting that the SIGE-Cerved sample is also representative of developments affecting the whole business sector.

Benchmarking SIGE with National accounts data Figure 4



[1] Figure Note: The figure shows the nominal and real value added growth rate for the private non-financial corporate sector (national accounts) and for the firms included in the matched SIGE-Cerved dataset. Survey weights are used to report sample observations to the population of Italian companies. Percentage changes (Δ). Value added is the difference between revenues and intermediate costs.

Sources: SIGE, Cerved, National accounts.

4. Results

4.1 Pricing policies over time

Figure 5 decomposes revenues and value added growth rates (solid lines) by their price and quantity components (histograms) given in equations (1) and (2). Firms' revenues (panel a) fluctuations over the 2016-22 period are correlated with output prices. The latter account on average for approximately one-third of revenues growth. However, during the period of economic recovery following the global pandemic, the notable surge in inflation also manifested itself in a substantial increase in the incidence of output price inflation on revenue growth rates.

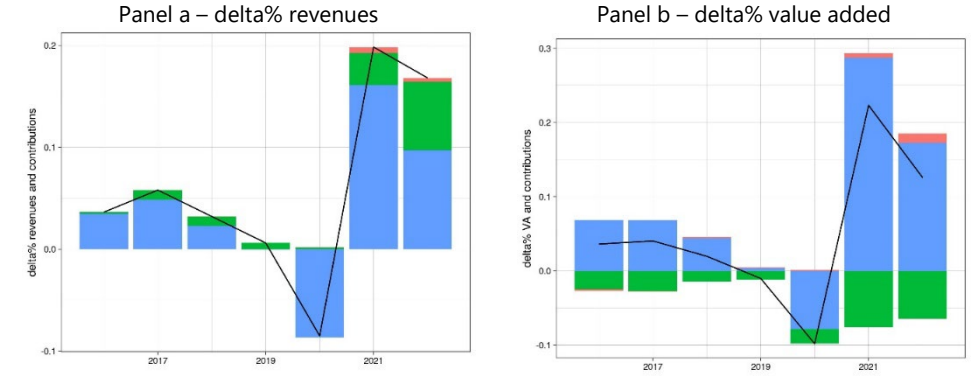
The aggregate dynamic of value added is only modestly affected by firms' pricing policies and their contribution is negative (panel b). This result suggests that, on average, the incomplete pass-through of input cost increases to output prices, given also the high extent of intermediate input intensity, typically does not provide firms with an operation hedging against changes in intermediate input prices. Moreover,

the growth in value added over time is predominantly driven by improvements in productivity of intermediate inputs; a technological shift in the production process or the substitutability of inputs –which could have become less available and relatively more expensive during the 2021-22 period- could lead to these results.

The aggregate results presented here contribute to the ongoing debate regarding the potential impact of firms' pricing policies in the recent inflationary episode. Our findings, which indicate a negative effect of pricing policies on the value added dynamic, are consistent with prior evidence from Italy, which suggests that firms' charging constant mark-ups did not contribute to soaring inflation (Colonna et al., 2023).

Price and quantity contributions to revenues and value added

Figure 5



[1] Figure Note: The histograms decompose the weighted average revenue and value added growth rates (solid lines) into a price, quantity and an interaction component derived according to equations (1) and (2).

Sources: Bank of Italy's Survey on Inflation and Growth Expectations and Cerved.

4.2 Size and sectoral differences

By examining firm-level data, it is possible to investigate potential heterogeneity between firms' pricing policies depending on their characteristics. Notably, firm size (given by the number of employees), can help differentiate between companies with different market power, exposure to cost fluctuations or international competition.

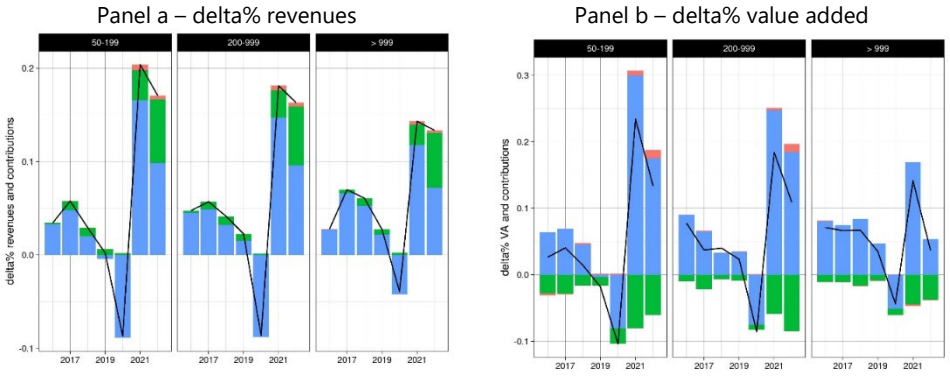
Figure 6 decomposes the nominal revenues and value added growth rate between medium (50-199), large (200-999) and very large (more than 999 employees) companies in their price, quantity and interaction components. For the sample period as a whole the average contribution of prices to revenues growth rates is modest for both medium-large and very large firms (accounting for about 2 and 1.7 per cent, respectively). As expected, this values align closely to inflation dynamic, given that firms' sale prices are covered to various degrees in consumer price indices.

Some differences are apparent with regard to the contribution of prices to the dynamics of value added across the firm-size spectrum. On average, for larger firms,

the contribution of pricing policies to value added is negative but nevertheless more muted with respect to the smaller ones. The incidence of the price component on the value added growth rate is approximately 40 percent for the former and about 70 percent for the latter. These discrepancies could be due to the fact that larger firms are less susceptible to fluctuations in intermediate input prices, since the impact of such price increases is approximately half that of smaller firms. Furthermore, the divergence in these outcomes became more pronounced following the impact of the pandemic.

Revenues and value added decompositions by firm size

Figure 6



[1] Figure Note: The histograms decompose the weighted average revenue and value added growth rates (solid lines) into a price, quantity and an interaction component derived according to the equations (1) and (2). Firm size classes are defined by the number of employees with medium firms 50-199, large 200-999 and very large >999.

Sources: Bank of Italy's Survey on Inflation and Growth Expectations and Cerved.

In contrast, differences across industries are less clear-cut. Sale price increases display a common pattern across sectors, and similarly to the aggregate trend, their contribution to the change in revenues increased notably only after the pandemic shock. The contribution of pricing policies to sectoral value added growth is on average negative, even across sectors. It became more severe only after the pandemic. This negative incidence of prices on value added growth rates was particularly pronounced in the construction sector.

4.3 Cost-shock driven pricing changes and the impact on value added

The evidence in the previous section, based on the accounting decomposition approach, seems to suggest a negative contribution of pricing policies alone on firms' economic margins. This could be due to both incomplete pass-through to output prices and to the high extent of intermediate input intensity. Existing research on the cyclical behaviour of mark-ups indicates that firms may encounter demand or technological constraints that inhibit their ability to fully adjust prices in response to

cost shocks. Furthermore, the competitive landscape and market structure can also significantly impact pricing decisions and the transmission of cost shocks.

In this section we utilize firm-level panel data covering the period from 2016 to 2022 to estimate the regression model specified in (2.2) and contrast pooled estimates with those derived from a panel with individual and time fixed effects.

The parsimonious model in (6), is used to delve more into the role played by pricing policies and intermediate input productivity terms (Table 2). The elasticity of the value added growth rate to contemporaneous IIP term is positive and significant; consistent with descriptive evidence provided in the previous section on the substantial positive contribution of quantity in driving value added dynamics. A one-standard deviation increase in IIP (13 percentage points) correlates with an increase by 11 percentage points in value added. Controlling for time invariant firm characteristics, within firms' estimates suggests that changes in PP correlate with increasing value added (column 3), but the magnitude of this effect is lower –with respect to the effect of productivity - and varies over time. A one-standard deviation increase in PP (5 percentage points) correlates with a 6 percentage points increase in value added. For the years 2021-22, however, increasing pass-through from rising intermediate input to output prices did not shield economic margins resulting in a drop of value added (column 4). This finding aligns with prior evidence indicating that firms have limited operational hedging capacity of their gross operating cash flows through sale price increases. Specifically, firms have limited capacity to transfer input cost increases to output prices, and changes in pricing policies result in a non-significant increase of firms' value added. Champion et al. (2023) document findings that are consistent with those presented here. They estimate a pass-through of 0.85, which implies that cost increases are only partly incorporated in sale prices which, in turn, leads to a fall in profit margins.

Pricing policies over time

Table 2

	Pooled	**Pooled**	**FE**	**FE**
	(1)	**(2)**	**(3)**	**(4)**
(Intercept)	0.0534 (0.0363)	0.0481 (0.0367)		
PPt	0.3824 (0.2123)		0.8918 *** (0.1602)	
IIPt	1.271 *** (0.1252)		1.204 *** (0.1192)	
PPt * 2016		0.5771 (0.5087)		-1.510 (0.8273)
PPt * 2017		1.093 (0.5671)		1.022 ** (0.3936)
PPt * 2018		1.270 * (0.5522)		0.6579 (0.7109)
PPt * 2019		2.555 *** (0.5795)		1.927 ** (0.6056)
PPt * 2020		1.502 *** (0.3492)		0.9890 (0.6775)
PPt * 2021		-1.486 *** (0.2867)		-1.895 *** (0.2968)
PPt * 2022		-0.7800 ** (0.2406)		-0.7694 ** (0.2801)
Fixed-Effects:	----------	----------	----------	----------
year	No	No	Yes	No
firm	No	No	Yes	Yes
S.E.: Clustered	by: firm & year	by: firm & year	by: firm & year	by: firm & year
Observations	5129	5129	5129	5129
R2	0.143	0.021	0.515	0.413

[1] Figure Note: The table reports OLS estimates of equation (6) using panel data from the SIGE-Cerved sample for the 2016-22 period. Observations are weighted using survey weights.

Sources: our estimates based on SIGE and Cerved data

Previous estimates obtained from the endogenous equation are suggestive of the extent of correlations between pricing policies and changes in value added. To overcome simultaneity and omitted variables bias inherent in previous results, 2SLS-IV approach is used. Input cost shocks (or surprises) occurred in the period t-1 are used to shift the operating pass-through in period t and recover local estimates of value added changes.

Results are reported in Table 3. First stage confirms that firms adjust their pricing policies following an idiosyncratic input cost shock (columns 2). A one-standard deviation cost shock (5 percentage points) results in a (partial) 3 percentage point increase in operating pass-through. This however does not result in a statistically significant impact on value added.

The instrument does not affect firms' pricing policies in a monotonous manner. By introducing an interaction term for the size class of the firm in the first stage, we can effectively consider the varying abilities of firms to convert cost shocks into output prices. Our findings indicate that larger firms possess a higher capability to endure cost fluctuations and enhance their operational pass-through (columns 4). Nevertheless, this alone does not lead to positive changes in value added.

The pass-through of cost shocks to value added

Table 3

	ΔVA_t	dPP_t	ΔVA_t	dPP_t	ΔVA_t
	(second stage)	*(first stage)*	*(second stage)*	*(first stage)*	*reduced form*
	(1)	(2)	(3)	(4)	(5)
dPP_t	0.0825 (1.054)		0.0527 (1.016)		
Input cost shock $_{t-1}$		0.6267 *** (0.1315)			0.0517 (0.6511)
$dIIP_t$			0.0019 (0.0020)	-3.31e-5 (0.0006)	
Input cost shock t-1 * Size =1				0.6611 *** (0.1385)	
Input cost shock t-1 * Size =2				0.4100 ** (0.1183)	
Input cost shock t-1 * Size =3				1.066 *** (0.1297)	
Fixed-Effects:	----------	----------	----------	----------	----------
year	Yes	Yes	Yes	Yes	Yes
firm	Yes	Yes	Yes	Yes	Yes
S.E.: Clustered	by: firm & year	by: firm & year	by: firm & year	by: firm & year	by: firm & year
Observations	2019	2019	2019	2019	2019
R2	0.45	0.53	0.45	0.54	0.45
F-test (1st stage)		595.29		209.85	
Wald (1st stage)		22.718		108.96	

[1] Figure Note: The table reports estimation of equations (6) and (7) using panel data with firm-time fixed effects. Robust standard errors in parenthesis.

Sources: our estimates based on SIGE and Cerved data.

5. Results

The strong increase in inflation over the past few years fed a vigorous debate around its sources and, consequently, the most appropriate policy response. One of the hypothesis under scrutiny, the "greed-flation", suggests that firms may have charged higher prices with respect to the increased costs. This narrative, which echoed in a number of discussions in Europe as well as the United States, does not find empirical support in our study.

We use firm-level data on output and intermediate input price changes to illustrate the contribution of firms' pricing policies and that of productivity to the growth rate of their value added. This decomposition accounts for a number of heterogeneities, generally hidden within aggregate statistics, such as: differences in firms' intermediate input intensity, exposure to input price changes and the capacity to transfer price changes from suppliers to consumer markets.

Overall, the robust growth in output quantities has driven the positive developments in value added observed during the post pandemic recovery. When examining heterogeneity across firms, we find that the pass-through of input price variations to end prices correlates with positive developments in value added especially for the larger firms and those with higher intermediate input intensity.

Finally, we isolate exogenous shifts in firms' pass-through due to idiosyncratic cost shocks. Following an unexpected increase in input prices, we find that firms' increase their pass-through, with larger firms doing so by a greater extent. However, this increase in pass-through is not necessarily a sign of resilience given that their value added does not improve significantly. This evidence is compatible with firms' aiming to maintain their current profitability and share part of the economic consequences with their counterparts (final consumers or other businesses) in order to retain market shares. The paper offers new evidence on firms' price setting behaviour and the extent to which firms and consumers share the burden of a cost shock.

6. References

Andler, M., and Kovner, A., 2022. Do Corporate Profits Increase When Inflation Increases?, Federal Reserve Bank of New York Liberty Street Economics, July 13, 2022.

Arce, O., Hahn, E., and Koester, G., 2023. How tit-for-tat inflation can make everyone poorer. The ECB Blog, 30 March 2023.

Champion, M., Edmond, C., Hambur, J., 2023. Competition, Markups, and Inflation: Evidence From Australian Firm-Level Data, Reserve Bank of Australia mimeo.

Colonna, F., Torrini, R., and Viviano, E., 2023. The profit share and firm markup: how to interpret them?, Banca d'Italia, Questioni di Economia e Finanza (Occasional Papers), N. 770, 2023.

Dixon, H. D., Grimme, C., 2022. State-dependent or time-dependent pricing? New evidence from a monthly firm-level survey: 1980–2017. European Economic Review, N. 150, 2022.

Hahn, E., 2023. How have unit profits contributed to the recent strengthening of euro area domestic price pressures?, ECB Economic Bulletin, N. 4, 2023.

Lagarde, C., 2023. The path ahead. The ECB and Its Watchers XXIII. Conference speech. March 2023.

Lane, P.R., 2023. Underlying inflation. Lecture at Trinity College Dublin. March 2023.

Lein, S. M., 2010. When do firms adjust prices? Evidence from micro panel data., Journal of Monetary Economics, N. 57, 2010.

Leduc, S., Li, H., Liu, Z. 2024. Are Markups Driving the Ups and Downs of Inflation?, FRBSF Economic Letter, N. 12, 2024.

Moore, H., Geoffrey, 1986. Chapter 'Inflation and Profits' in Business cycles, inflation and forecasting 2nd edition, Ballinger, 1986.

Moro. A., 2010, Biased technical change, intermediate goods and total factor productivity, working paper.

Riggi, M., Tagliabracci, A., 2022. Price rigidities, input costs, and inflation expectations: understanding firms pricing decisions from micro data, Questioni di Economia e Finanza (Occasional Papers), N. 733, 2022.

BANCO DE ESPAÑA
Eurosistema

Irving Fisher Committee on
Central Bank Statistics

BIS

eccbso
European Committee of Central Balance Sheet Data Offices

Workshop: "New Insights from financial statements"

The pass-through of cost shocks to firms' prices and the impact on value added

Fabio Parlapiano (Bank of Italy)

Madrid, 17 October 2024

BANCA D'ITALIA

Background and objectives

Background:

The resurgence of inflation in the years 2021-22, prompted renewed interest in **firms' pricing policies** for their role a) in the **propagation** of the inflationary dynamic b) setting the extent that firms' bear (or share) the **burden of soaring production costs** with their counterparties (hedging).

What the paper does:

Uses detailed firm-level **survey data** on output and intermediate input price changes matched with **balance-sheet** information to:

a. Analyses the **relationship between firms' pricing policies** and their **value added dynamic**;

b. Decomposes the annual variation of value added into three additive components: i) **price** (pricing policy), ii) **quantity** (intermediate input productivity) and iii) the **interaction** between the two;

c. Analyses the **response** of firms' pricing policies and value added dynamics to **idiosyncratic intermediate input price shocks**. Enabling us to characterise the **resilience** of firms, in terms of their cost-transfer capacity to output prices and, eventually, hedge their economic margins.

Related works

How does an inflationary environment affect firms' profits?

- When inflation is high profits tend to expand. However, **if inflation is particularly high, profits decelerate and eventually fall**. Intuitively, profit expand when output prices grow faster than input prices (Andler and Kovner, 2022; Moore, 1983).

To what extent firms' pricing policies, measured via mark-ups, contribute to inflation?

- Surging inflation in the 2021-22 period cannot be attributed to pricing policies, as **mark-ups mostly remained stable** or returned to pre-pandemic levels (Colonna *et al.*, 2023; Leduc *et al.*, 2024);

- **The transmission of cost-shocks to output prices is incomplete** (0.7 percent increase in output prices for every 1 percent rise in costs), and b) contrary to the expectations of competitive markets, the pass-through effect is **lower in industries with higher levels of concentration** (Ganapati *et al.*, 2020; Champion et *al.*, 2023);

- Firms with greater **market power** (high mark-up) **are less likely to pass on cost-shocks** to end prices (Kouvavas *et al.*, 2021; Kharroubi *et al.* 2023).

Modelling framework (1/2)

Decomposing Value Added (VA)

Nominal growth rate (Δ) of value added (VA) is **decomposed** as follows:

$$\Delta VA_{t,t-1} = \frac{VA_t - VA_{t-1}}{VA_{t-1}} = \frac{1}{VA_{t-1}} \left[\Delta Rev \cdot Rev_{t-1} - \Delta Cost \cdot Cost_{t-1}\right]$$

$$\Delta VA_{t,t-1} = \Delta p^{VA}_{t,t-1} + \Delta q^{VA}_{t,t-1} + \Delta pq^{VA}_{t,t-1} \quad \text{1)}$$

The **contribution** made by firms' pricing policies (both output and intermediate input prices), productivity and the interaction term between prices and quantities are defined:

$$\Delta p^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \Delta p^{o}_{t,t-1} - Cost_{t-1} \cdot \Delta p^{i}_{t,t-1}\right]$$

$$\Delta q^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \Delta q^{o}_{t,t-1} - Cost_{t-1} \cdot \Delta q^{i}_{t,t-1}\right]$$

$$\Delta pq^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \Delta p^{o}_{t,t-1} \cdot \Delta q^{o}_{t,t-1} - Cost_{t-1} \cdot \Delta p^{i}_{t,t-1} \cdot \Delta q^{i}_{t,t-1}\right]$$

Modelling framework (1/2)

Decomposing Value Added (VA)

Nominal growth rate of value added (VA) is decomposed as follows:

$$\Delta VA_{t,t-1} = \frac{VA_t - VA_{t-1}}{VA_{t-1}} = \frac{1}{VA_{t-1}} \left[\Delta Rev \cdot Rev_{t-1} - \Delta Cost \cdot Cost_{t-1}\right]$$

$$\Delta VA_{t,t-1} = \Delta p^{VA}_{t,t-1} + \Delta q^{VA}_{t,t-1} + \Delta pq^{VA}_{t,t-1} \qquad \textbf{1)}$$

The **contribution** made by firms' pricing policies (both output and intermediate input prices), productivity and the interaction term between prices and quantities are defined:

$$\Delta p^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \boxed{\Delta p^{o}_{t,t-1}} - Cost_{t-1} \cdot \boxed{\Delta p^{i}_{t,t-1}}\right] \qquad \textbf{PP,} \text{ pass-through}$$

$$\Delta q^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \boxed{\Delta q^{o}_{t,t-1}} - Cost_{t-1} \cdot \boxed{\Delta q^{i}_{t,t-1}}\right] \qquad \textbf{IIP,} \text{ input prod.}$$

$$\Delta pq^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \Delta p^{o}_{t,t-1} \cdot \Delta q^{o}_{t,t-1} - Cost_{t-1} \cdot \Delta p^{i}_{t,t-1} \cdot \Delta q^{i}_{t,t-1}\right]$$

Modelling framework (1/2)

Decomposing Value Added (VA)

Nominal growth rate of value added (VA) is decomposed as follows:

$$\Delta VA_{t,t-1} = \frac{VA_t - VA_{t-1}}{VA_{t-1}} = \frac{1}{VA_{t-1}} \left[\Delta Rev \cdot Rev_{t-1} - \Delta Cost \cdot Cost_{t-1}\right]$$

$$\Delta VA_{t,t-1} = \Delta p^{VA}_{t,t-1} + \Delta q^{VA}_{t,t-1} + \Delta pq^{VA}_{t,t-1} \qquad \textbf{1)}$$

The **contribution** made by firms' pricing policies (both output and intermediate input prices), productivity and the interaction term between prices and quantities are defined:

$$\Delta p^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \Delta p^{o}_{t,t-1} - Cost_{t-1} \cdot \Delta p^{i}_{t,t-1}\right]$$

$$\Delta q^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \Delta q^{o}_{t,t-1} - Cost_{t-1} \cdot \Delta q^{i}_{t,t-1}\right]$$

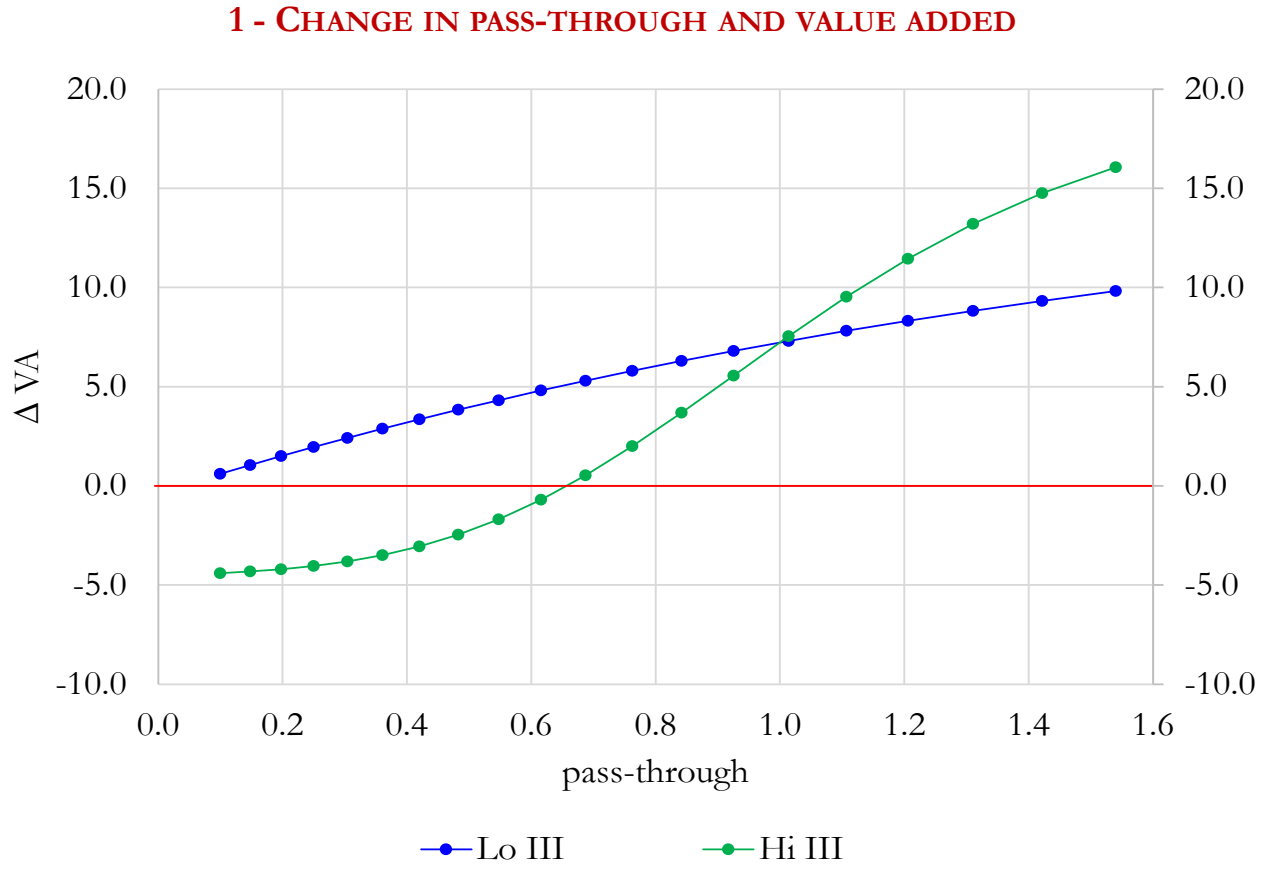
III,
Intensity input

$$\Delta pq^{VA}_{t,t-1} = \frac{1}{VA_{t-1}} \left[Rev_{t-1} \cdot \Delta p^{o}_{t,t-1} \cdot \Delta q^{o}_{t,t-1} - Cost_{t-1} \cdot \Delta p^{i}_{t,t-1} \cdot \Delta q^{i}_{t,t-1}\right]$$

Modelling framework (2/2)

Comparative statics: pricing policies and growth rate of value added

Firms' pricing policies (in ratio) and ΔVA in **two polar cases**: the share of intermediate input used to produce 1 unit of output is 0.1 (Low III) and 0.9 (High III); Δ output quantities are held constant.



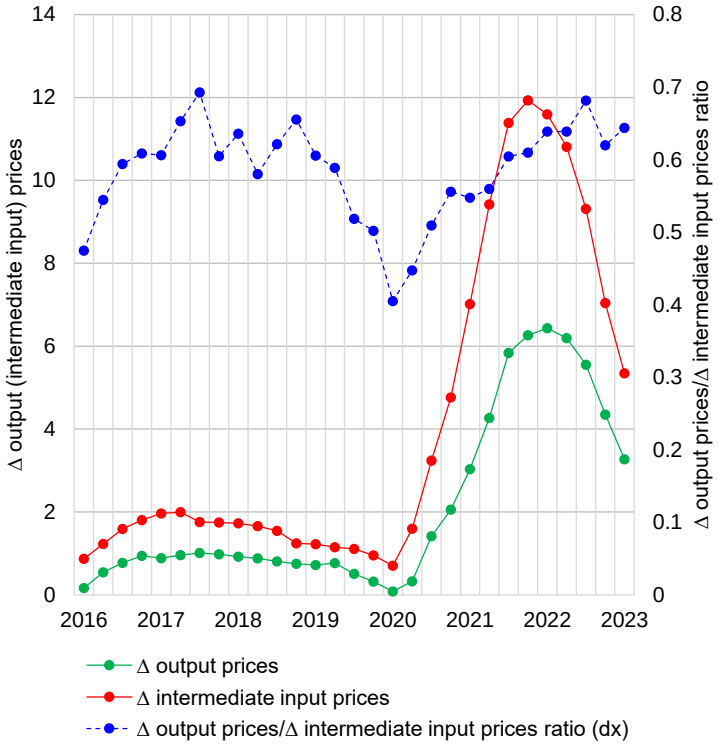
Data

Prices ($\Delta p^{i,o}$)

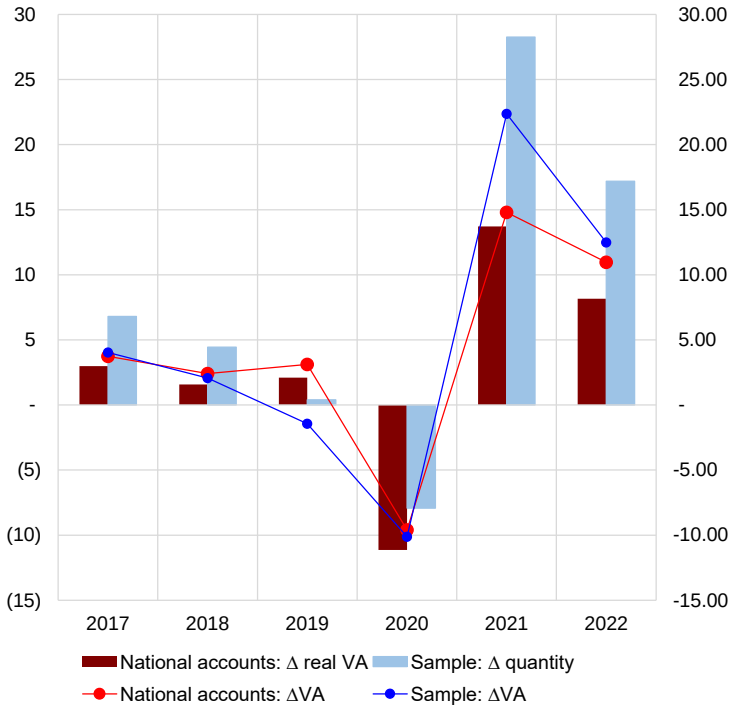
Firm-level price change information for the 2016-22 period sourced from the Survey on Inflation and Growth Expectations (SIGE); information collected at quarterly frequency from a sample of medium-sized and large Italian firms.

- What was the average variation of output (intermediate input) price for the products or services sold (purchased) by your company over the last 12 months?

2 - OPERATING PASS-THROUGH (%)

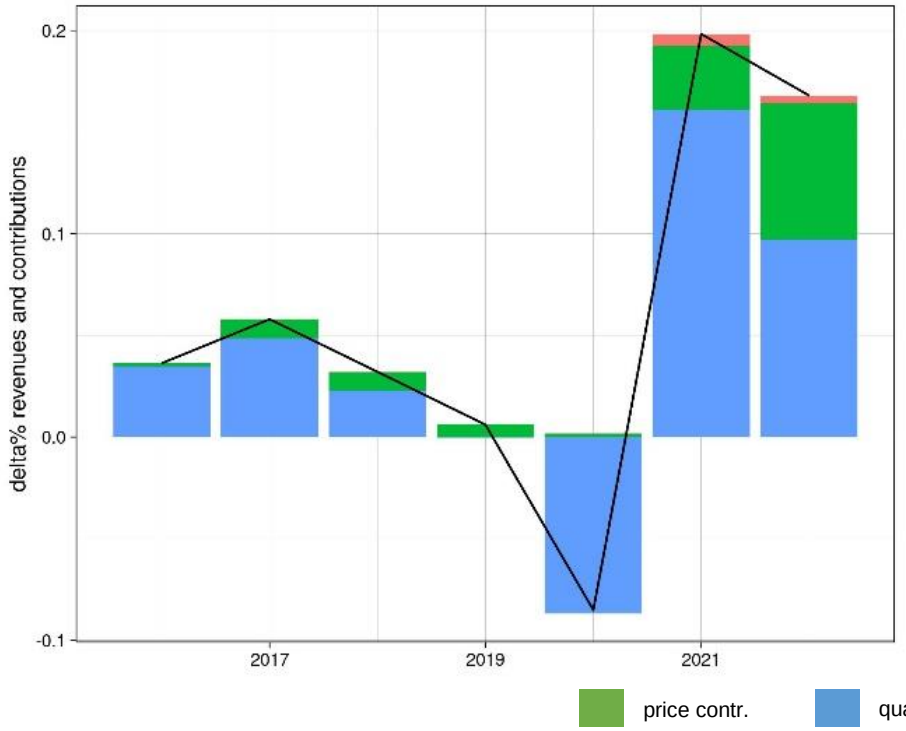


3 - SIGE AND NATIONAL ACCOUNTS DATA (%)

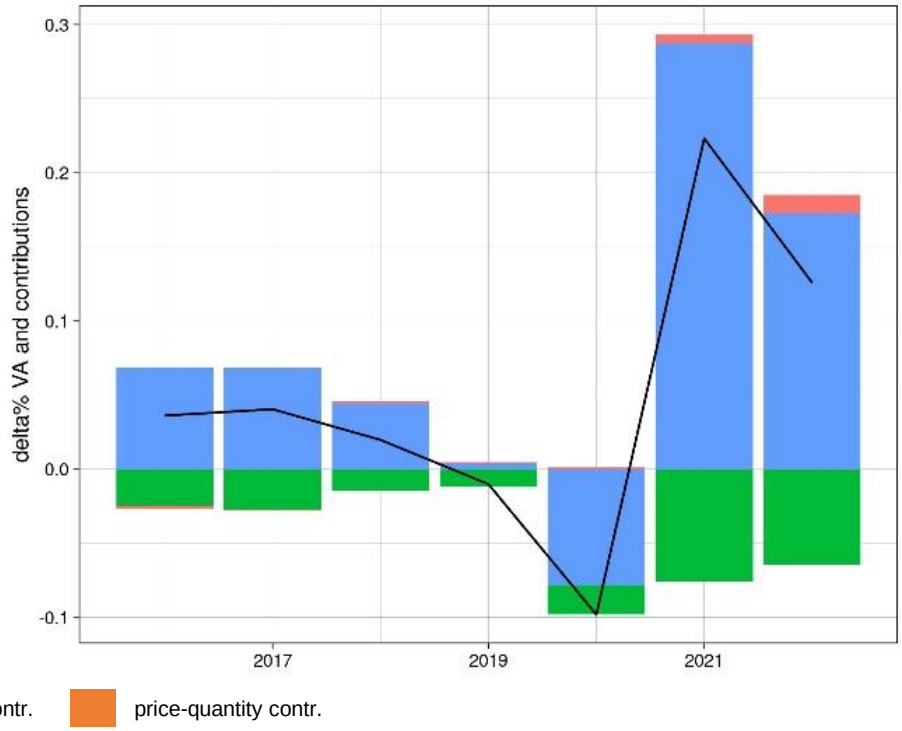


1. Results: pricing policies over time

4 - ΔREVENUES (%)



5 - ΔVALUE ADDED (%)

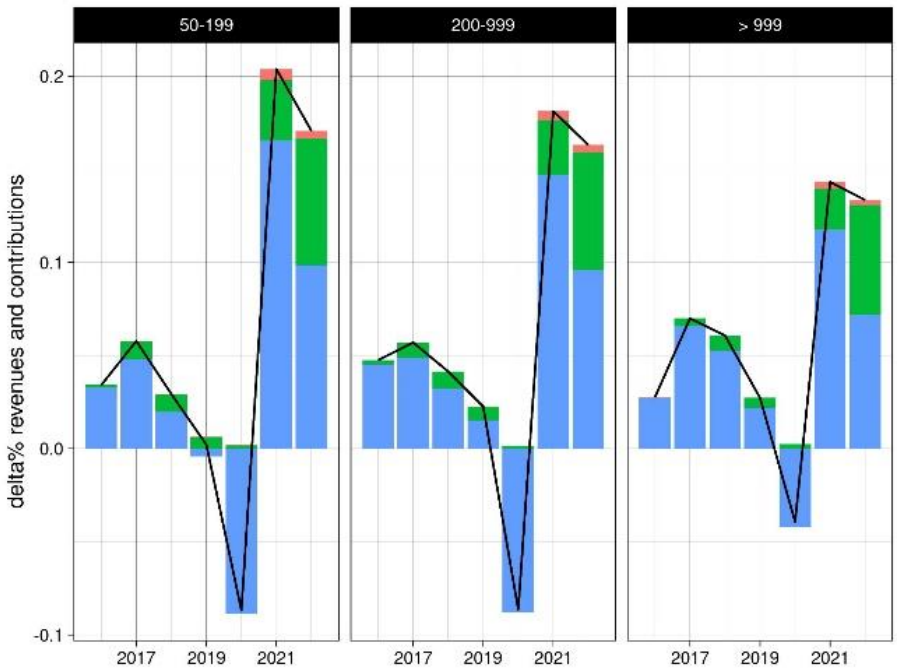


price contr. quantity contr. price-quantity contr.

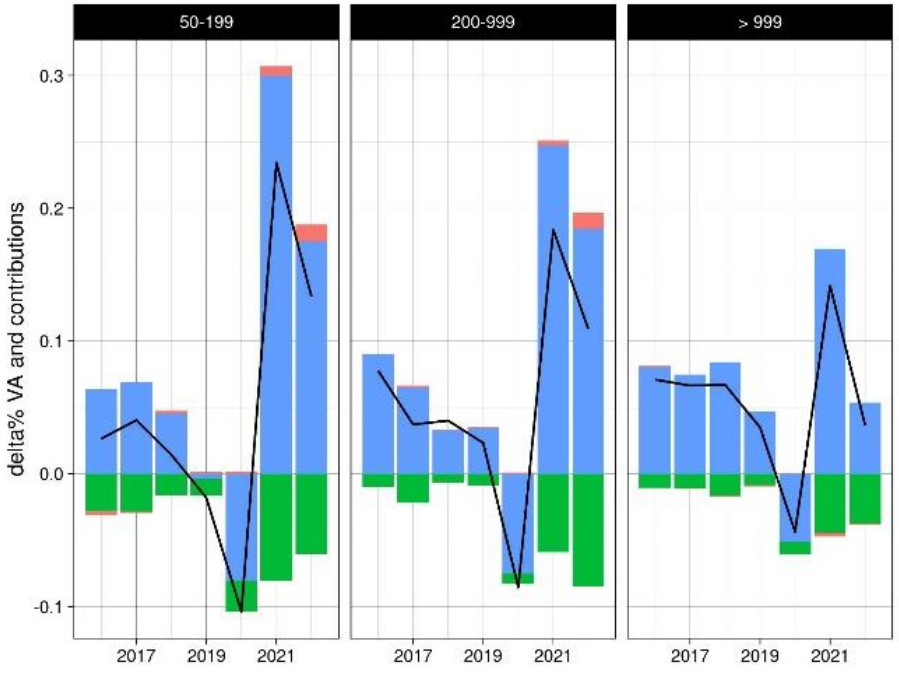
- **Firms' revenues** (**solid line**) fluctuations over the 2016-22 period are correlated with output **prices** (**green histogram**); accounting for 1/3 of revenues growth. **After the pandemic**, the notable surge in inflation also manifested itself in a **substantial increase in the incidence of output price** on revenue growth rates.
- Negative effect of **prices (Δp^{VA})** on ΔVA due to a) incomplete PT and b) the high extent of intermediate input intensity. The VA dynamic is predominantly driven by **quantities (Δq^{VA})**.

2. Results: differences accross sizes

6 - ΔREVENUES (%)



7 - ΔVALUE ADDED (%)



price contr. quantity contr. price-quantity contr.

- **Larger firms display a negative, but more muted, contribution of prices to ΔVA**; they are potentially less susceptible to fluctuations in intermediate input prices, since the impact of such price increases is approximately half that of smaller firms. Furthermore, the divergence in these outcomes became more pronounced following the impact of the pandemic.

3. Cost-shock driven pricing changes and the impact on value added

2SLS - IV setting

How changes in a firm's pricing policy affects its economic performance? **Simultaneity and omitted variables bias** addressed in IV setting:

- **Idiosyncratic shocks** (surprises) to **intermediate input costs** occurred in the $t-1$ period ($Z_{i,t-1}$) are used in the first stage to isolate exogenous variations in a firm's pricing policy (dPP):

$$Z_{i,t-1} = \Delta p^{i}_{i,t-1} - E\left(\Delta p^{i}_{i,t-1} | t-2\right) \qquad \text{INSTRUMENT}$$

$$dPP_{i,t,t-1} = \alpha + \gamma Z_{i,t-1} \cdot X_i + \gamma_i + \tau + \mu_{i,t,t-1} \qquad \text{FIRST STAGE}$$

$$\Delta VA_{i,t,t-1} = \alpha + \delta \widehat{dPP}_{i,t,t-1} + \gamma_i + \tau + \epsilon_{i,t,t-1} \qquad \text{SECOND STAGE}$$

- **Relevance:** if firms adjust their current **pricing policy** following **input price forecasting errors** made in the previous period (cor. Z, dPP = 0.3);
- **Exogeneity**: independence between the **instrument** and the **outcome variable** (cor. Z, ΔVA$\cong 0$);
- **Exclusion**: effects on ΔVA occurr only via dPP; input prices forecasting error made by a firm impacts its value added changes only via firm-specific changes to its pricing policy.

3. Cost-shock driven pricing changes and the impact on value added

2SLS - IV setting

- **Firms adjust their pricing policies following an idiosyncratic input cost shock**: 1-SD cost shock (5 p.p.) results in a 3 pp increase in operating pass-through (c.2). Impact on value added is however not statistically significant (c.1);
- **Larger** firms endure cost fluctuations and **enhance their operational pass-through by more**. This alone **does not lead to positive changes in value added**.

	ΔVA_t	dPP_t	ΔVA_t	dPP_t	ΔVA_t
	(second stage)	*(first stage)*	*(second stage)*	*(first stage)*	*reduced form*
	(1)	(2)	(3)	(4)	(5)
dPP_t	0.0825 (1.054)		0.0527 (1.016)		
Input cost shock $_{t-1}$		0.6267 *** (0.1315)			0.0517 (0.6511)
$dIIP_t$			0.0019 (0.0020)	-3.31e-5 (0.0006)	
Input cost shock t-1 * Size =1				0.6611 *** (0.1385)	
Input cost shock t-1 * Size =2				0.4100 ** (0.1183)	
Input cost shock t-1 * Size =3				1.066 *** (0.1297)	
Fixed-Effects:	----------	----------	----------	----------	----------
year	Yes	Yes	Yes	Yes	Yes
firm	Yes	Yes	Yes	Yes	Yes
S.E.: Clustered	by: firm & year	by: firm & year	by: firm & year	by: firm & year	by: firm & year
Observations	2019	2019	2019	2019	2019
R2	0.45	0.53	0.45	0.54	0.45
F-test (1st stage)		595.29		209.85	
Wald (1st stage)		22.718		108.96	

Conclusion

- The narrative for which firms have charged higher prices with respect to the cost increase (*i.e.* greedy pricing) **does not find empirical support** in our study.
  - Firms' **pass-through** of intermediate input to output prices is, on average, **incomplete**.

- The robust growth in **output quantities** has driven the **positive developments** in value added observed during the **post pandemic recovery**.

- Shifts in firms' pass-through due driven by idiosyncratic cost shocks **do not translate into improvements** in their value added dynamics.

- This finding aligns with firms' **limited operational hedging capacity** through sale price increases; firms strive to uphold their existing profitability levels and **opt to share** a portion of the economic losses with either end consumers or businesses in order to **retain their market shares**.

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Corporate debt structure and heterogeneous monetary policy transmission[1]

Marie Alder,
European University Institute

Nuno Coimbra,
Bank of France

Urszula Szczerbowicz,
SKEMA Business School

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.



This paper contains research conducted within the network "Challenges for Monetary Policy Transmission in a Changing World Network" (ChaMP). It consists of economists from the European Central Bank (ECB) and the national central banks (NCBs) of the European System of Central Banks (ESCB).

ChaMP is coordinated by a team chaired by Philipp Hartmann (ECB), and consisting of Diana Bonfim (Banco de Portugal), Margherita Bottero (Banca d'Italia), Emmanuel Dhyne (Nationale Bank van België/Banque Nationale de Belgique) and Maria T. Valderrama (Oesterreichische Nationalbank), who are supported by Melina Papoutsi and Gonzalo Paz-Pardo (both ECB), 7 central bank advisers and 8 academic consultants.

ChaMP seeks to revisit our knowledge of monetary transmission channels in the euro area in the context of unprecedented shocks, multiple ongoing structural changes and the extension of the monetary policy toolkit over the last decade and a half as well as the recent steep inflation wave and its reversal. More information is provided on its website.

Corporate debt structure and heterogeneous monetary policy transmission

Marie Alder* Nuno Coimbra** Urszula Szczerbowicz***

October 2024

Abstract

Using French firm data, we show that corporate debt structure plays a significant role in monetary policy transmission. In addition to interest rate policy, we analyse the impact of a novel ECB-induced sovereign spread shock that is related to bond liquidity and show that both types of policy tightening diminish French firms' investment. The transmission of conventional shocks is stronger for firms with higher shares of bank debt, but contractionary bond spread shocks lower investment more for firms with higher shares of bond debt. Bond liquidity tightening leads to higher bond-bank loan interest rate spreads and lower bond issuance.

Keywords: Monetary policy transmission, Corporate debt structure, Investment.
JEL Classification: E22, E43, E44, E52.

*European University Institute (email: marie.alder@eui.eu); **Banque de France, CEPR (email: nuno.coimbra@banque-france.fr); ***SKEMA Business School (email: urszula.szczerbowicz@skema.edu).
We thank Anne Duquerroy, Saki Bigio, Fiorella di Fiore, Valère Fourel, Refet Gürkaynak, Fédéric Holm-Hadulla, Guillaume Horny, Sergio Mayordomo, Christoph Grosse-Steffen, Julia Schmidt and the participants at the 1st Workshop of the ChaMP ESCB research network, 2024 IJCB annual research conference, 40th Symposium on Money Banking and Finance (GdRE), 2024 EABCN "New Challenges in Monetary Economics", SNDE Symposium 2024, Doctoriales Macrofi 2024, EEA-ESEM Congress 2024 and Banque de France internal seminars for helpful comments and suggestions. We are also grateful to Olivier Gonzalez for his input on Fiben data. The views presented in this paper are the authors' and do not reflect those of the Banque de France or the Eurosystem.

1. Introduction

Since the Global Financial Crisis, the share of debt securities in total non-financial corporation (NFC) debt has increased significantly.[1] This increase was particularly striking in the Euro area, traditionally dependent on bank-based finance, where the bond share of corporate debt almost doubled between 2007 and 2021 (from 9% to 16.6%). In France, the share of bond debt in total firm debt rose from 19% to 30% in the same period, but there were other countries where the increase was even more dramatic. In Spain, for example, market debt as a share of total firm debt went from 3% in 2007 to 14.7% in 2021.

Financial instruments that firms use to finance their activity have different characteristics, making them imperfect substitutes. Previous literature has shown that bank loans and bond securities respond differently to monetary policy shocks (Kashyap, Stein, and Wilcox, 1993, Becker and Ivashina, 2014, Lhuissier and Szczerbowicz, 2022). As such, corporate debt structure can matter for the monetary policy transmission.

In this paper, we not only show the importance of debt composition in the transmission of conventional monetary policy to French firms' investment but, using a novel approach, we also investigate the effect of unconventional monetary policy arising from movements in sovereign Bond SPreads (BSP) and tightly linked to liquidity. We show that the impact on firm investment depends jointly on the type of monetary policy and the firms' debt structure. Firms that are more dependent on bank finance react more to conventional monetary policy (CMP) shocks, while firms more reliant on market finance are more reactive to BSP shocks.

The importance of debt structure of NFC for monetary policy transmission was highlighted by Philip Lane and Isabel Schnabel, yet with apparently different conclusions. On one hand, Lane (2022) argued that a large bank share of NFC debt may dampen conventional monetary transmission due to slower speed of pass-through of policy rate changes into bank lending rates, when compared with corporate bond yields. On the other hand, Schnabel (2021) claimed that CMP shocks should have a stronger impact on the rates charged for bank loans than for corporate bonds, so the real effects of CMP strengthen with the share of bank finance in the economy. These contrasting views highlight that the importance of debt structure for the monetary transmission is not yet fully understood.

[1] Adrian, Colla, and Shin (2013), Darmouni and Papoutsi (2022).

The canonical New Keynesian channel of monetary policy focuses on real rates and their impact on demand, via intertemporal consumption optimization. In such models, the financing structure of firms is irrelevant, as typically the Modigliani-Miller theorem holds. Recent literature has focused on financial frictions, where additional channels are present. For example, monetary policy has been shown to affect NFC investment through a "balance sheet channel" (Bernanke and Gertler, 1995). This channel implies that policy rate increases can make it more expensive for the firms to borrow externally and raise the firm-specific user cost of capital, decreasing their investment. Higher policy rates increase the "external finance premium" because they reduce asset values, and thus decrease the value of firms' balance sheets and their net worth.[2] Most of this literature has focused on loans or is silent about the distinction between market and bank debt.

But in the presence of financial frictions, the pass-through of monetary policy to bank and market debt could be quite different. In that case, central bank's rates would affect firms differently, depending on their access to different types of external financing. The firm-specific debt structure becomes an important factor of this heterogeneity, as long as monetary policy has an uneven impact on the costs of bank lending and debt securities (Holm-Hadulla and Thürwächter, 2021) and there is imperfect substitution between the two types of credit. Both bank loans and corporate bond markets can be subject to different frictions, independently reinforcing or attenuating the monetary transmission mechanism.[3]

According to the bank lending channel, CMP tightening leads to more restrictive bank credit conditions. In these circumstances, bond markets can provide an alternative to bank financing to NFC that have access to that "spare tyre" (Kashyap, Stein, and Wilcox, 1993, Adrian, Colla, and Shin, 2013).[4] If monetary tightening decreases bank loans but stimulates corporate bond issuance, then the effectiveness of monetary policy could be hampered. The investment will fall less after interest rates hikes for the firms

[2]The external finance premium is the difference between the cost of capital raised by firms externally and the cost of capital raised using cash flows generated internally.

[3]Bank loans tend to be more costly and more exposed to cyclical shifts in credit supply (Becker and Ivashina, 2014). Bonds on the other hand are held by a dispersed number of investors, which makes them difficult to renegotiate existing credit contracts in times of financial distress, impeding efficient restructuring (Bolton and Scharfstein, 1996, Crouzet, 2017). Another difference is that in France, corporate bonds have longer maturities than bank loans on aggregate, and they are more likely to have fixed interest rates (Gueuder and Ray, 2022).

[4]The term "spare tyre" was used by Greenspan (1999) in his speech "Do efficient financial markets mitigate financial crises?", where he referred to capital markets as substitutes for the loss of bank financial intermediation.

with better access to bond markets (Crouzet, 2021). Moreover, the burden of adjustment will fall disproportionately on firms that do not have access to the bond market "spare tyre", leading to possibly unwanted competitive effects of monetary policy. However, as argued by Darmouni, Giesecke, and Rodnyansky (2020), when frictions in bond financing are important, a bond lending channel can potentially dominate and firms with more bond financing would be *more* negatively affected by monetary tightening. In this paper, using a panel of micro data of French firms, we show that this is the case for bond spread shocks, but not for conventional ones which have a stronger impact on firms that rely relatively more on bank financing.

The implementation of unconventional monetary policy (UMP) has added an additional dimension to monetary policy and its transmission. In this paper, we acknowledge the development in central bank toolkits and investigate whether there are differences in the transmission of CMP and (a particular type of) UMP with respect to corporate debt structure. UMP have already been shown to have heterogeneous effects on issuance and cost of each debt instrument compared to CMP.[5] Quantitative easing, in particular, reduced risk premia on debt securities, which stimulated corporate bond issuance rather than bank lending to NFC.[6] Therefore, conventional and unconventional monetary policies have potentially different effects on NFC investment.[7] In this paper, we focus on the bond credit channel of unconventional monetary policies and show that firms which rely relatively more on market finance are more sensitive than those that rely more on bank finance, while the converse holds true for conventional monetary policy.

More specifically, we use firm-level panel data for France (FIBEN) to investigate the relevance of corporate debt structure for ECB monetary policy transmission. Our dataset consists of more than 11,000 distinct firms and around 80,000 observations, over the period 1999-2019. We rely on high frequency surprises around ECB announcements to identify monetary policy shocks. For CMP, we use the updated surprises from

[5]Lhuissier and Szczerbowicz (2022) show that an expansionary CMP in the United States leads to a rise in aggregate loans and a decline in debt securities issuance, while an expansionary UMP generates a decline in loans and a rise in debt issuance.

[6]There is evidence that central bank asset purchases lowered the NFC debt securities cost relative to the cost of bank funding, thus encouraging companies to switch from bank to bond financing (see for instance Arce, Mayordomo, and Gimeno, 2020, De Santis and Zaghini, 2021, Grosse-Rueschkamp, Steffen, and Streitz, 2019).

[7]Holm-Hadulla and Thürwächter (2021) study the effects of the ECB shocks to short-term and long-term rate on euro area countries' GDP. They show that a higher bond share goes along with a weaker transmission of short-term policy rate shocks to GDP, but the transmission of longer-term yields policy shocks is stronger.

Jarociński and Karadi (2020), who separate conventional monetary policy shocks from central bank informational shocks. These surprises are based on risk-free yield changes of maturity up to one year around ECB announcements, which allows them to capture both interest rate decisions and (Odyssean) forward guidance. For UMP, we use the high frequency changes of 10-year sovereign spreads between French and German bonds around ECB announcements, in order to study the effect of monetary policy shocks linked to French bond market liquidity.[8]

We begin by identifying a novel UMP shock extracted from movements in French-German 10-year sovereign spreads around ECB announcements (BSP shocks). Since conventional monetary policy could also have an impact on bond spreads, we orthogonalize the 10-year French-German spread surprises[9] with respect to CMP surprises. We then show that ECB-induced BSP tightening shocks reduce French sovereign bond market liquidity, as measured by bid-ask spreads, across *all* available maturities and their first principal component.

We use local projections (LP) proposed by Jordà (2005) to evaluate the average impact of ECB conventional and unconventional monetary policy on French firms' investment. We find that both CMP and BSP shocks have an economically and statistically significant negative effect on French firms' investment. Then we proceed to estimate the heterogeneous effect of both types of monetary policy depending on firms' debt structure. We control for firm fixed effects to capture permanent differences across firms and also for sector-time fixed effects in order to capture differences in how sectors respond to aggregate shocks. We provide evidence that monetary policy transmission to firm investment is a function of each firm's share of market debt and the specific type of monetary policy being used. Conventional monetary policy has a stronger impact on firm investment when the firm is more reliant on bank loans, while unconventional policies that increase liquidity in bond markets (such as quantitative easing) have a stronger effect when firm financing is more market-based.

To shed light on the mechanism of transmission, we also show that contractionary BSP shocks transmit to a lower share of bond debt in new issuance, with negative effects on NFC investment. Moreover, we also show that after a contractionary BSP shock the relative cost of bonds compared to bank loans increases, indicating that the

[8]The French sovereign bond market is comparable to the German market in terms of credit rating, currency and amounts outstanding in the individual bonds (Ejsing and Sihvonen, 2009). Moreover, Schwarz (2018) finds that liquidity is an important driver of Euro area sovereign spreads.

[9]Using data from Altavilla et al. (2019).

transmission of BSP to funding costs is stronger for market debt.

The contribution of this paper is twofold. First, we identify a bond-liquidity channel of MP, associated with BSP shocks, and provide evidence on its impact on the liquidity of the French bond market. Second, we study the role of corporate debt structure in the transmission of both types of monetary policy to investment. By uncovering the relative importance of bond and bank credit supply shocks induced by these two types of monetary policy, we provide novel empirical evidence on the credit and liquidity channels of different forms of monetary policy.

The paper is organised as follows. Section 2 provides an overview of the literature. Section 3 describes the data used, while Section 4 compares the aggregate effects of both types of monetary policy and explores the role of the debt structure in the heterogeneous responses of firms to the two monetary policy shocks. Section 5 uses complementary data sources to shed light on the transmissions channels and Section 6 concludes.

2. Literature

Our paper relates to the literature on the credit channel of monetary policy, both from the firm balance sheet and the bank lending channel perspectives.[10] This literature links the heterogeneous response of firms to monetary policy shocks in the presence of financial frictions, related both to banks and NFC balance sheets. Ottonello and Winberry (2020) find that firms with low default risk are the most responsive to monetary shocks. Other studies argue that the firm-level response also depends on their size (Gertler and Gilchrist, 1994), their holdings of liquid assets (Jeenas, 2023) and the type of financial constraints faced by the firms (Chitu et al., 2023). Cloyne et al. (2023) use the firm's age and dividend payouts as a proxy for financial constraints, finding that financial frictions account for about one third of the aggregate investment response to conventional monetary policy.

The imperfect substitutability of different instruments of corporate debt generates additional frictions that affect monetary policy transmission. In particular, the share of floating-rate debt and the debt maturity were shown to affect the transmission of monetary policy to firms' investment and stock prices (Ippolito, Ozdagli, and Perez-

[10]For the firm balance sheet channel, see Ashcraft and Campello (2007), Bernanke and Gertler (1995), and Gertler and Gilchrist (1994). For the bank lending channel, see Bernanke and Blinder (1988), Bernanke and Gertler (1995), Jiménez et al. (2012), Stein and Kashyap (2000), Kishan and Opiela (2000).

Orive, 2018, Gürkaynak, Karasoy-Can, and Lee, 2022, Jungherr et al., 2022).[11]

Another important aspect of debt heterogeneity is related to the loan-bond composition of corporate debt.[12] The firm-level evidence from the United States shows that a higher share of bonds in corporate financing attenuates the impact of conventional monetary policy on firms' stock prices and investment, in line with the bank lending channel (Crouzet, 2021, Ippolito, Ozdagli, and Perez-Orive, 2018). In this context, the possibility of issuing corporate bonds can hamper the effectiveness of interest rate increases, as the NFC can substitute bank loans with bond financing, even if only partially so. On the other hand, Darmouni, Giesecke, and Rodnyansky (2020), using firm-level data, highlight that stock prices and investment of listed euro area firms with higher bond to asset ratios are more affected by conventional monetary policy shocks than their counterparts, pointing to the importance of bond market frictions in the euro area.

We contribute to this literature by investigating the reaction of French firms investment to conventional monetary policy shocks. French firms have the highest share of bond financing in the EA, and as such we provide evidence that is more in line with the US evidence. This reinforces the idea that bond market depth is important to explain differences between US and EA-wide results. More importantly, we study here not only the role of bond-loan debt structure for CMP transmission but also for UMP transmission, with a focus on the bond liquidity channel.[13] To do this, we use high-quality microeconomic data on French firms. Earlier literature found that unconventional monetary policy reduces corporate bond yields and risk premia, stimulating corporate bond issuance (Wright, 2012, Altavilla and Giannone, 2017, Lo Duca, Nicoletti, and Vidal Martínez, 2016, Lhuissier and Szczerbowicz, 2022). Giambona et al. (2020) used microeconomic data to study the effect of QE on investment. They find that investment by firms with access to the bond market increases. Using aggregate data in a panel of EA countries, Holm-Hadulla and Thürwächter (2021) show that the share of aggregate bond financing plays an opposite role in conventional and unconventional monetary policy transmission. It weakens the transmission of short-term policy rate shocks to GDP but strengthens the effects of monetary policy shocks to longer-term yields, which tend to be more responsive to UMP measures. This suggests that the bank lending

[11] See also Bräuning, Fillat, and Wang (2020), Barclay and Smith Jr (1995), Diamond and He (2014).

[12] The composition of debt instruments and related financing costs play an important role in firms' investment dynamics. See Dees et al. (2022), De Fiore and Uhlig (2015), Adrian, Colla, and Shin (2013).

[13] Our sample also includes firms that are not publicly listed.

channel is not the main transmission channel for UMP. In this paper, we explore a separate channel of UMP related to bond liquidity, exploiting firm-level data.[14] We provide a novel way to identify unconventional monetary policy shocks that affect liquidity, and show that firms which are more reliant on corporate bonds markets for their external finance are more impacted by it than firms that are more reliant on bank lending.

Finally, our paper is related to the literature studying the impact of monetary policy shocks using high-frequency identification (Kuttner, 2001, Gürkaynak, Sack, and Swanson, 2005, Gertler and Karadi, 2015, Gerko and Rey, 2017, Jarociński and Karadi, 2020, Nakamura and Steinsson, 2018, Altavilla et al., 2019 among others). We add to this literature by constructing high frequency surprises for the bond liquidity channel of monetary policy, which was particularly important during ECB asset purchase programs. We identify shocks that impact bond liquidity from movements in 10-year French-German sovereign spreads around ECB policy announcements that are orthogonal to CMP shocks.

3. Data and summary statistics

3.1. Monetary policy shocks

We rely on high frequency surprises to identify monetary policy shocks. For CMP, we use the updated surprises from Jarociński and Karadi (2020) who separate conventional monetary policy shocks from central bank informational shocks. These (updated) surprises around ECB announcements are based on risk-free asset changes of maturity up to one year, capturing both interest decisions and near-term forward guidance.[15]

Unconventional monetary policy is a large set of tools that encompasses anything that goes beyond the use of policy rates. This can include very diverse instruments such as forward guidance, asset purchases or lending operations. Since we want to examine the role of credit channels in bank and bond financing of French firms, we want to capture unconventional monetary policy shocks that are most directly connected to

[14]Recent work by Lee and Engel (2024) highlights a link between QE, liquidity and investments in risky foreign assets. Here, we focus on the impact of unconventional monetary policy on the domestic economy.

[15]The updated MP shocks are currently available on Marek Jarocinski's webpage. The updated series are based on the 1st principal component of the Monetary Event-window changes in OIS with maturities 1, 3, 6 months and 1 year. The Monetary Event-window is defined as in Altavilla et al. (2019).

French bond markets. To do this, we use the high frequency movements in the 10-year French-German sovereign spread (BSP shock). To remove any possible systematic effect of CMP on these spreads, we also orthogonalize these surprises with respect to the CMP shocks. Figure A3 displays the time series of the two types of shocks. As expected, CMP shocks seem more frequent until 2008, while BSP shocks are more prominent during the zero lower bound (ZLB) period.

Table 1: Largest BSP Shocks

Date	BSP shock	Sources	Explanations
12/03/2020	0.117	ECB statement, Financial Times	Christine Lagarde stated it was not the ECB's role to respond to movements in government debt markets.
08/12/2011	0.057	ECB press release, Financial Times	Announcement of 3-year LTRO. Markets were disappointed by the downplaying of the prospect of renewed sovereign bond purchases.
02/08/2012	0.054	ECB statement, Reuters	ECB disappoints markets which expected a more immediate OMT implementation.
04/06/2020	−0.053	ECB statement, Financial Times	Increase of pandemic emergency purchase programme (PEPP) envelope by €600 billion (including undisclosed amount of corporate bonds).
09/02/2012	−0.051	ECB press release, ECB statement, Reuters	ECB eases eligibility criteria for collateral used in Eurosystem credit operations.
06/11/2008	−0.048	ECB statement, Reuters	After a 50 basis points cut, analysts suggest door to further monetary policy easing is likely to remain open.
07/07/2011	0.047	ECB statement (Q&A), Reuters	Jean-Claude Trichet refused to discuss further steps if Euro zone crisis worsens, generating uncertainty about acceptance of Greek collateral
06/05/2010	0.034	ECB statement (Q&A), The New York Times	Jean-Claude Trichet declares that purchases of Greek bonds were not discussed at this meeting.
06/06/2012	−0.030	ECB statement, Reuters	ECB decided to continue conducting its main refinancing operations at fixed-rate tender procedures with full allotment for as long as necessary.
03/11/2011	−0.029	CBPP programme announcement	ECB announces details of its new covered bond purchase programme (CBPP2).
03/03/2011	−0.029	ECB statement, Financial Times	QT postponed: "ECB shelved further steps to unwind the exceptional support for eurozone banks."
10/05/2001	0.028	ECB statement, Monthly Bulletin	Lower M3 growth than previously announced.

To better understand the nature of these unconventional shocks, we also provide a narrative description of the events associated with the windows where we observe the largest intra-day BSP shocks. Table 1 provides information on what was communicated at the dates of the 12 largest BSP shocks (in absolute terms). For example, the largest shock in our sample occurs on March 12, 2020 after the COVID-19 pandemic outburst. At this date, markets were disappointed by the modest strengthening of the APP and the statement by President Lagarde reaffirming that the ECB is "not here to close spreads". In other dates, there was also significant movement in the 10-year French-German spread during conferences that featured announcements regarding asset purchase programs, suggesting these are likely important drivers of the spread. Empirical

evidence shows indeed that the announcements of the ECB asset purchases such as OMT and PEPP reduced the 10-year French-German spread (Szczerbowicz, 2015, Hubert et al., 2024).

Both French and German sovereign bond markets are comparable in terms of credit rating, currency and amounts outstanding in the individual bonds (Ejsing and Sihvonen, 2009), but most importantly there is also evidence that movements in French-German sovereign spreads reflect mostly changes in liquidity premia.[16] Later, in Section 4.1., we will show that BSP surprises have a strong and consistent impact on French sovereign bond liquidity across all maturities and their first principal component. This strong link to liquidity is evidence that a bond market liquidity channel is present and constitutes one of the drivers of firm-level impact.

There are also important reasons to believe movements in the sovereign spread can be important to firms, and in particular to corporate bond markets. First, ECB asset purchases were shown to spill over to corporate bonds (Altavilla, Carboni, and Motto, 2021). Second, sovereign bond yields are an important benchmark for corporate bond pricing. They usually represent a floor for corporate bond yields or, in terms of bond prices or credit ratings, a ceiling, known as the "sovereign-ceiling hypothesis" (Borensztein, Cowan, and Valenzuela, 2013, Adelino and Ferreira, 2016, Almeida et al., 2017).

3.2. Firm-level data

We measure the impact of the ECB monetary policy on French firms' investment using firm-level data on French companies from the Banque de France's FIBEN (*Fichier Bancaire des Entreprises*) database. We rely on the consolidated database as investment and financing choices are often decided at the group level. We combine two consolidated databases for each of the accounting standards under which French companies can publish their results (French standards and International Financial Reporting Standards).

Companies are identified by their SIREN number, which is an Insee[17] code identifying uniquely each company, organization or association operating in France. Results are typically reported once a year, so data is annual, but the reporting date and the length of the fiscal year can vary. To avoid double-counting we exclude observations

[16] ECB (2009), box 4: "New evidence on credit and liquidity premia in selected euro area sovereign yields".

[17] French National Institute of Statistics and Economic Studies.

whenever there are multiple entries with the same SIREN-date pair, and for consistency we exclude those for which the duration of the fiscal year is different from 12 months. This ensures that any observation is for a complete year of business.

For the remaining observations, if the reporting month is between January and June, the observation year is considered as occurring the preceding year for the purpose of aligning it with time-sector fixed effects or annual aggregate controls. Otherwise, it is considered as occurring in the year of reporting. This allows us to consider the year in which most of the activities described in the observation take place.

Given that our monetary policy shocks are daily, we follow Durante, Ferrando, and Vermeulen (2022) and align the yearly aggregation of shocks to match the reporting month of each firm. Yearly aggregation follows the formula below, yielding firm-specific monetary policy shocks based on their reporting month.[18]

$$Y_{f,t} = \sum_{j=r_{f,t}+1}^{12} m_{j,t-1} + \sum_{j=1}^{r_{f,t}} m_{j,t} \tag{1}$$

where $Y_{f,t}$ is the firm-f specific shock at year t, $r_{f,t}$ is the reporting month of firm f in year t and $m_{j,t}$ is the MP shock in month j of year t.

Finally, we exclude observations with negative equity, negative assets or with leverage above the 99th percentile to exclude firms close to or in default. We also winsorize the remaining firm-level variables at the 1st and 99th percentiles for the regressions, as is standard in the literature.

3.3. Descriptive statistics

Our working sample contains 81 358 observations from 11 478 distinct groups (henceforth *firms*). Our sample covers the years after the introduction of the Euro in 1999 to 2019.[19] There is substantial heterogeneity among the firms of our sample. We report summary statistics in Table 2 and in Figure 1. The average share of bond debt in total debt (henceforth bond ratio) is on average 0.05, due to the presence of a high number of firms that do not finance themselves through bonds. This is about half of the ratio

[18] In Appendix A, we show that results are robust to using a more straightforward calendar year aggregation.

[19] We remove the observations from 2020, so as not to incorporate the Covid-19 pandemic period.

found by Darmouni, Giesecke, and Rodnyansky (2020) in their dataset of large firms that enter the EURO STOXX 50 index. Around 80% of the observations in our sample are from firms declaring no bond debt during that reporting year.

Table 2: Summary statistics

	Mean	Std.dev.	Min	Max
Monetary policy shocks (pp)*				
CMP shocks	0.04	0.10	-0.09	0.26
BSP shocks	-0.01	0.04	-0.10	0.08
Dependent variable				
Investment rate** (%)	1.57	5.87	-19.58	43.25
Aggregate control variables				
French output gap	-0.006	1.60	-2.6	2.79
French inflation	1.38	0.77	0.07	2.81
VIX	19.9	7.06	11.04	40
10y French sovereign rate	3.01	1.65	0.13	5.39
3m interbank rate	1.75	1.76	-0.36	4.63
Firm-specific control variables				
Leverage	0.25	0.17	0	0.79
Total assets (in bn)	0.39	1.72	0	25.72
Cash flows to total assets	0.10	0.07	-0.19	0.75
Bond ratio	0.05	0.16	0	1
Maturity ratio***	0.56	0.36	0	1

** MP shock moments calculated under a calendar-year aggregation ** Investment rate is defined as the difference in net tangible assets with respect to lagged total assets. *** Maturity ratio is defined as the share of debt with maturity above 1 year.*

Figure 1 displays histograms from the subsample of firms that finance themselves at least partly through bonds.[20] Within the group that has access to the bond market, there are more firms with low bond ratios than there are with large ones. Despite this pattern, the distribution is more even for bond ratios than it is for bond debt over assets. There is a non-negligible number of observations across all possible values of bond ratios, allowing us to explore this dimension of the panel data. For bond debt over assets, we observe a more concentrated distribution. This is expected, since firms at the higher end of the distribution need to combine high leverage ratios with high bond ratios. In Figure A1 of Appendix A, we also provide histograms for the logarithm of assets, the share of firm debt with maturity above 1 year (henceforth *maturity ratio*) and

[20] As mentioned before, there is a large mass point at 0, such that including it masks the heterogeneity within the remaining firms.

leverage for the full sample of firms.

Figure 1: Distribution of bond ratios and bond debt over total assets

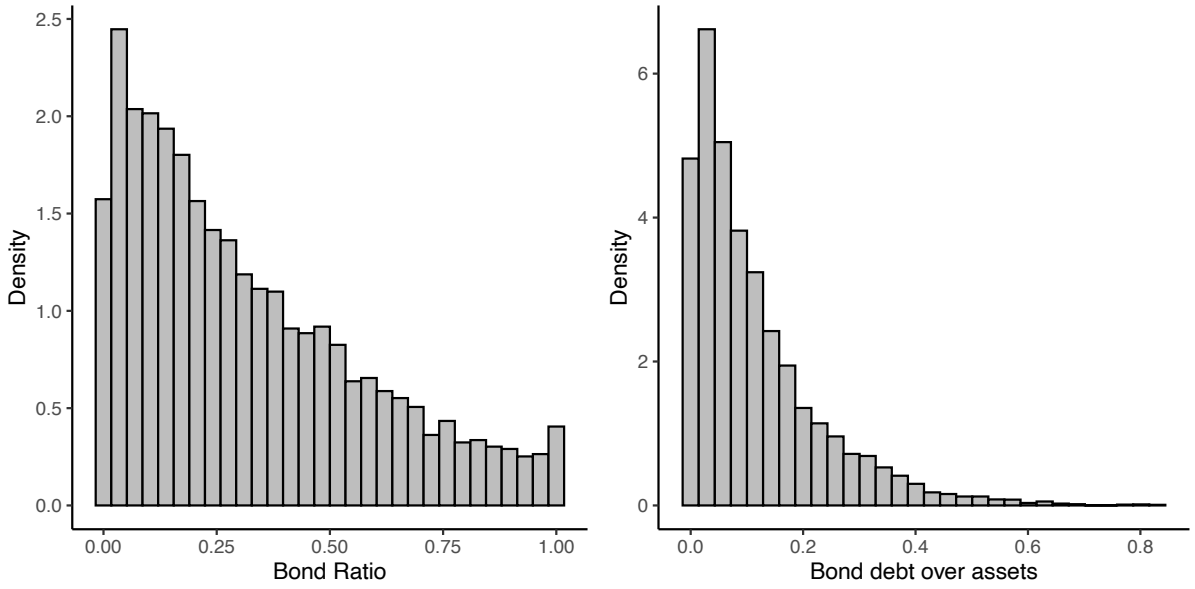
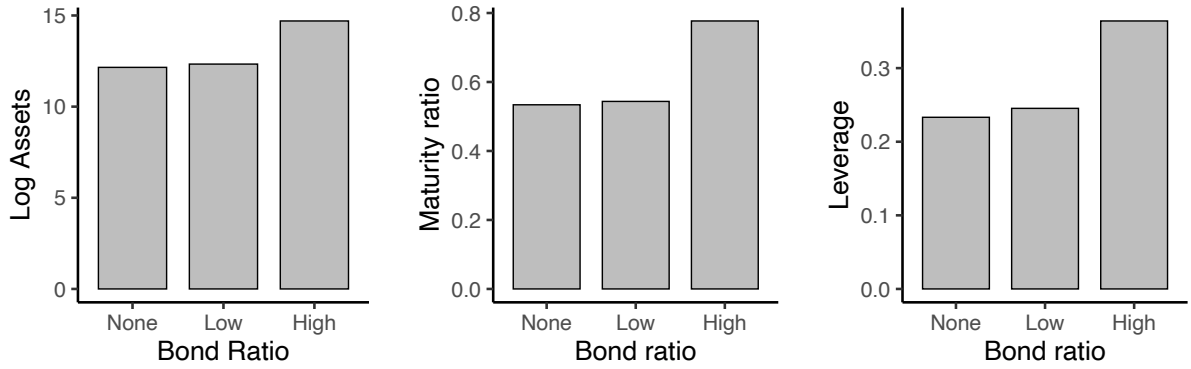


Figure 2 provides information on firms' assets, maturity ratio and leverage, according to their corporate debt structure. For each variable, we indicate the average across 3 categories of firms: those with a bond ratio equal to zero, those with a bond ratio below a cut-off value and those with a bond ratio above it. The cut-off value is the median bond ratio of firms with non-zero bonds. Firms with a bond ratio higher than the (conditional) median are on average significantly larger than those within the other two categories. They are also more highly levered and have higher maturity ratios relative to the other two groups.

Figure 2: Corporate debt structure and firm characteristics



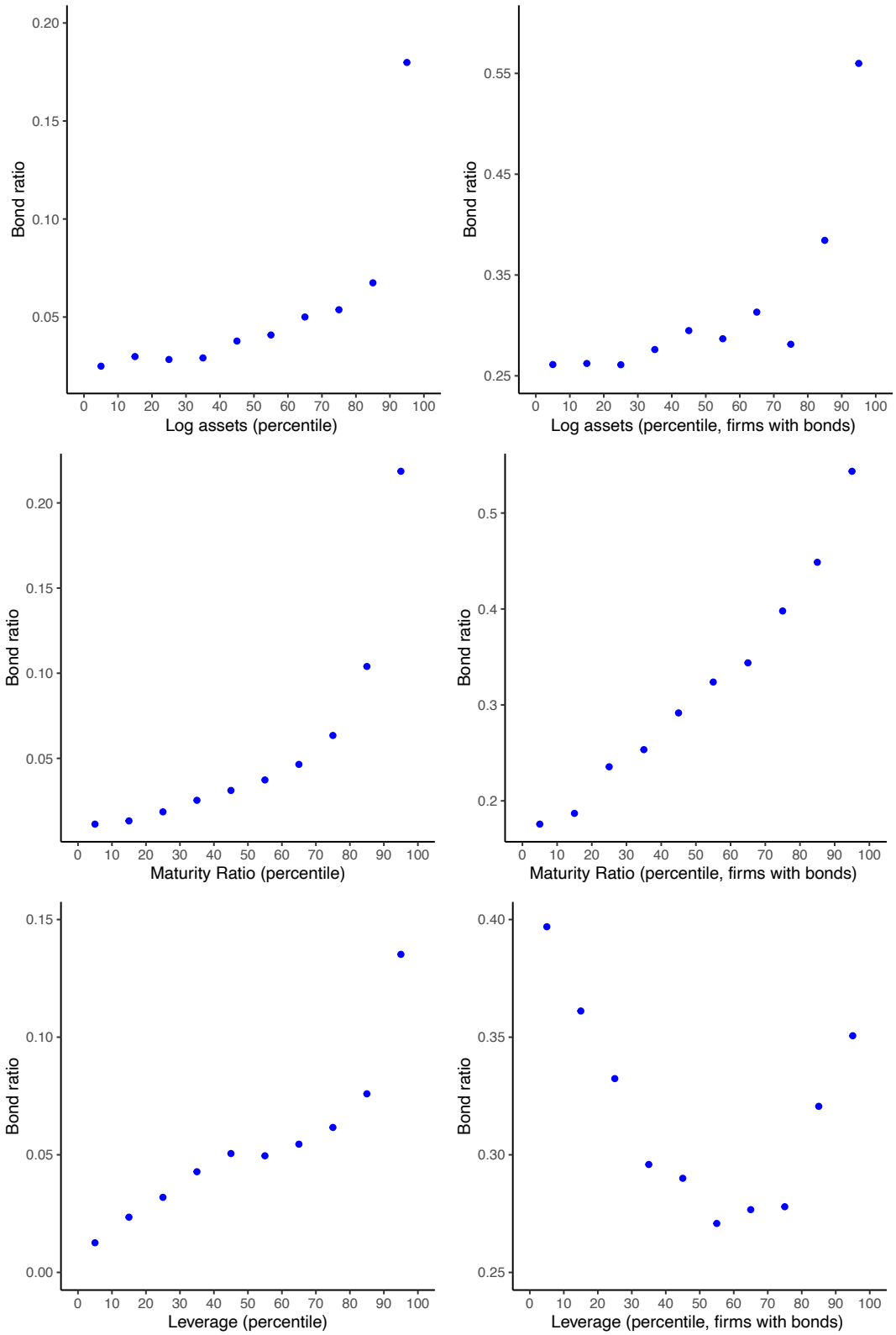
Each panel represents the average of the corresponding variable for three groups of firms: those with no bonds, those with bond ratios below the median (conditional on having bonds), and firms with bond ratios above the median.

Figure 3 shows binned scatter plots of the average bond ratio for different bins of, respectively, assets, maturity ratio and leverage.[21] Panels on the left are constructed using the full sample, while the ones on the right restrict the analysis to only firms with bonds. As can be seen in the top row, there is a positive relationship between bond ratio and asset size. The distribution has therefore significant skewness, with large firms having significantly larger bond ratios. In the top right panel, we see a similarly shaped distribution when we limit the sample to firms with non-zero bond debt, but with much higher values of bond ratios (and a bit more noise). In the middle row, we see that a similar pattern applies to the maturity ratio, defined as debt above 1 year maturity over total debt. Firms with higher maturity ratios tend therefore to have larger bond ratios.

In the bottom row, we highlight that doing a similar analysis by leverage reveals very different patterns in the full sample and the bond firms subsample. While we have a stable and monotonic positive relationship in the full sample, the subsample is U-shaped. The combination of the two panels shows that the low average bond ratios for firms in the lower leverage percentiles of the full sample are driven by firms with no bonds, but conditional on having bonds, low leverage firms actually have the highest share of bonds across the bins represented.

[21] Figure A2 in Appendix A also shows the equivalent charts for cash flow over assets.

Figure 3: Binned scatterplots of bond ratios by asset size, maturity and leverage



4. Monetary policy transmission to firm investment

In this section we first provide evidence of a strong link between BSP shocks and bond liquidity, as measured by bid-ask spreads. We then show that although both shocks are contractionary for firm investment, BSP shocks have a stronger impact on firms which are more reliant on market debt, while CMP shocks have a stronger impact on firms that are more reliant on bank loans.

4.1. Liquidity shocks and bid-ask spreads

In this sub-section, we show that BSP shocks have a strong link with bond market liquidity. We also show that the impact is felt at all levels of maturity, despite the BSP shocks being identified using only 10y French and German bonds.

We use smooth local projections (S-LP) as in Barnichon and Brownlees (2019). This penalization method can help deal with excess variability, without restricting ex-ante the shape of the impulse response function.[22] We use S-LP, as without the cross-sectional dimension of the panel, the number of observations is reduced considerably. Yet standard local projections are heavily parametrized and so estimates can be less precisely estimated and can be erratic in smaller samples (Ramey, 2016). On the other hand, more efficient VAR approaches might be too restrictive and lead to bias. To address these issues, S-LP make use of a shrinkage parameter that pins down the bias/variance trade-off of the estimator. When this parameter is set to 0, the method coincides with standard local projections estimated by least squares, whereas when it is large the impulse response converges to a polynomial distributed lag model (Almon, 1965). We follow Barnichon and Brownlees (2019), and let the data choose the shrinkage parameter using 5-fold cross-validation, picking the value that provides the best pseudo-out-of-sample fit.[23]

We run smooth local projections on daily bid-ask spread data for French sovereign bonds with maturities running from 1m to 50y.[24] Given that momentum is an important factor for asset prices at higher frequencies, we include 6 lags of the shocks. Our specification is as follows:

[22] See Li, Plagborg-Møller, and Wolf (2024).

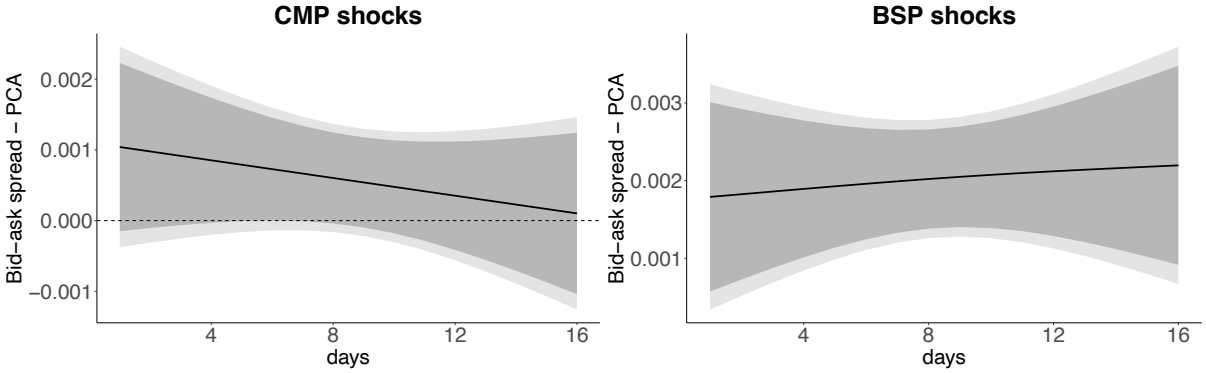
[23] For additional details on the method and its properties, see Barnichon and Brownlees (2019).

[24] The following maturities were considered: 1m, 3m, 6m, 1y, 2y, 3y, 5y, 10y, 15y, 20y, 30y and 50y.

$$BidAsk_{i,t+h} = \alpha_h S_t + \sum_{l=1}^{6} \Gamma_l^h S_{t-l} + \epsilon_{i,t+h} \tag{2}$$

where $BidAsk_{i,t+h}$ is the h-day forward bid-ask spread of bonds with maturity i (or their first principal component). S_t is either the CMP or the BSP shock at time t. In Figure 4, we show the impact of each shock on the first principal component of bid-ask spreads across all maturities. In Figure A4 of Appendix A, we also show the results for each individual maturity.

Figure 4: Response of first principal component of bid-ask spreads to CMP (left panel) and BSP (right panel) shocks



Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: days after the shock.

Figure 4 shows that the BSP shocks have a consistent positive impact on the first principal component of bid-ask spreads, pointing to the worsening of market liquidity of French sovereign bonds. After a contractionary BSP shock, the common component of bid-ask spreads across all maturities rises by about 0.2bp. The average bid-ask spread of 10y bonds is about 0.4bp, so it is not a trivial value.[25]

The same is not true for CMP shocks, for which there is no consistent liquidity impact, as can be seen on the left panel of Figure 4. Looking at specific maturities, there is some limited impact on intermediate maturities but overall the effect is much less clear. In particular, there is no positive impact of CMP on any of the bid-ask spreads in the set of maturities used to identify CMP shocks (1m, 3m, 9m and 1y). However, there is some impact on spreads at 10y maturity, highlighting the importance

[25] In Appendix A, Figure A5 shows results are robust to excluding data after 2019, while A6 shows this result is robust to including a Great Financial Crisis dummy that takes the value 1 during 2008, albeit it reinforces the result that CMP shocks are much less linked to liquidity than BSP ones.

of orthogonalizing movements in the spread from CMP shocks when identifying BSP shocks.

Although it is hard to claim that *all* high frequency movements in the French-German spread are due to relative liquidity across these markets, the evidence provided in this section clearly indicates that there is a stronger and more consistent link between BSP shocks and liquidity in bond markets, than there is between CMP shocks and liquidity. This is not claiming that CMP has no impact on liquidity, rather simply that BSP shocks can be a useful proxy for (monetary policy driven) bond market liquidity shocks.

4.2. Investment response to monetary policy

We now examine the aggregate effect of monetary policy shocks on French firms' investment rates. To capture the time profile of the response, we use a panel local projection approach proposed by Jordà (2005). We define net investment rate $I_{i,t}$ of firm i as the first difference of net tangible assets in year t, scaled by total assets in year *t-1*.[26]

To measure the effect of conventional and unconventional monetary policy shocks at time t on investment at horizons $h \in (0, 1, ..., 5)$, we estimate the following set of equations:

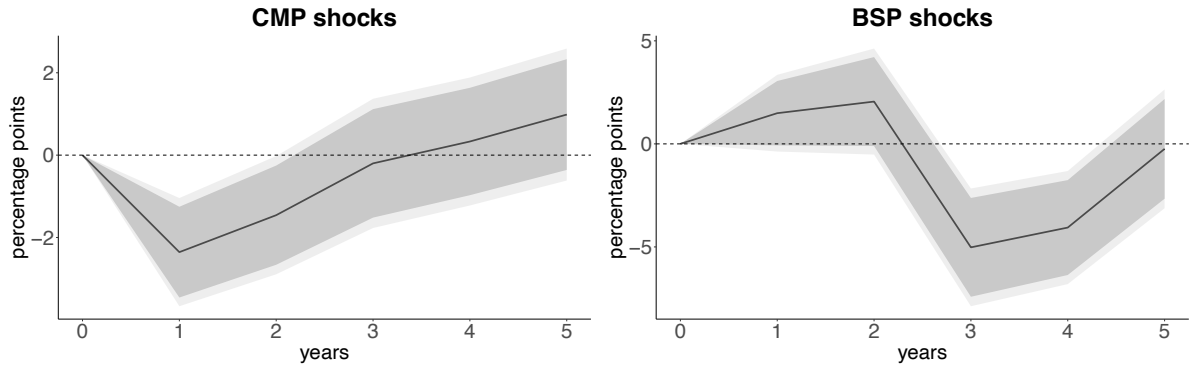
$$\Delta I_{i,t+h} = \alpha^h S_{i,t} + \Psi^h Z_{t-1} + \sum_{l=1}^{3} \Gamma_l^h X_{i,t-l} + \mu_i^h + \epsilon_{i,t+h} \tag{3}$$

where $\Delta I_{i,t+h}$ is the h-year forward difference in the net investment rate: $\Delta I_{i,t+h} = I_{i,t+h} - I_{i,t-1}$. $S_{i,t}$ is a vector of CMP and BSP shocks aligned to the reporting month of firm i. Z_{t-1} is the control vector of lagged aggregate controls: French output gap, French inflation, VIX, 10-year French sovereign rate, 3-month interbank rate. $X_{i,t-1}$ is the vector of lagged firm-specific controls: leverage, total assets, cash flows to total assets, bond ratio, maturity ratio and a bond dummy that is equal to 1 for firms that have non-zero share of bond financing. μ_i are firm fixed effects.

Figure 5 shows the average impulse response function of investment rate to a 100 basis point upward surprise for CMP (left panel) and BSP shock (right panel) at each horizon h (from 1 to 5 years). A CMP tightening of 100bp leads to a 2.4pp decline of investment with respect to firm's total assets, while a contractionary BSP shock of 100bp reduces it by close to 5pp. The CMP shock has an economically and statistically

[26]Our focus is on tangible investment, as research has indicated that fluctuations in debt financing have a more significant impact on physical investment, whereas equity financing fluctuations are more closely linked to R&D investment dynamics (Bianchi, Kung, and Morales, 2019)

Figure 5: Average response of investment to CMP (left panel) and BSP (right panel) shocks



Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

significant negative effect in the first two years after the shock, while the BSP shock decreases French firms' investment with a longer lag, starting only on the third year after the shock.

Our estimates for CMP shocks are consistent with the ones found in the literature.[27] The identified BSP shock has an impact at its peak that is larger per bp of the shock. We normalize the impulse response to 100bp shocks, which we find easier to interpret economically. Nevertheless, Figures A10 and A11 of Appendix A show the impulse responses of specifications (3) and (4) when each shock is rescaled by its standard deviation in the sample.

To better understand whether the investment response to BSP shocks is driven by sovereign French market developments or rather by movements in German yield (possibly reflecting some flight-to-safety reactions), we also add to our specification the surprises in 10-year German bond yields during the same announcement window.[28] Figure A19 in Appendix A displays the average response of investment to CMP and BSP shocks with this additional control, confirming that our results are not driven by surprises in 10-year German yields.

In Appendix A, we additionally show that results are robust to not using any

[27] For example, using US firm-level data, Cloyne et al. (2023) show that a 100bp rise in the interest rate leads to a fall in business investment between 2.4 and 3.2% on average after two years. Papers where investment is defined as the log change in net tangible assets tend to find larger values, such as 10% in Ferreira, Ostry, and Rogers (2023) or 31% in Durante, Ferrando, and Vermeulen (2022).

[28] As in case of the BSP shocks, we orthogonalise the surprises in 10-year German yield on the days of ECB announcements with CMP shock.

aggregate controls[29] (Figure A12) or using only output gap and inflation (Figure A13). Finally, results are also robust to aggregating shocks through a simple calendar year sum (instead of adjusting to each firm's reporting month), which is not surprising given most firms in our sample report their results in December (Figure A14).

4.3. Heterogeneous Transmission of Monetary Policy

In order to investigate possible heterogeneity in the transmission of the two types of monetary policy shocks, we explore the role of corporate debt structure. To do this, we again estimate LP, but we interact the shock with the (lagged) firm-specific *bond ratio* (defined as the share of bond liabilities in total firm debt). A value of 0 indicates that the firm has only bank loans, while a value of 1 implies the firm has no bank loans but only bond debt.

$$\Delta I_{i,t+h} = \alpha_b^h B_{i,t-1} S_{i,t} + \alpha_m^h M_{i,t-1} S_{i,t} + \sum_{l=1}^{3} \Gamma_l^h X_{i,t-l} + \mu_i^h + \theta_{s,t}^h + \epsilon_{i,t+h} \tag{4}$$

where $B_{i,t-1}$ is the lagged bond ratio and $M_{i,t-1}$ the maturity ratio, defined as long-term debt over total debt. We include a maturity interaction term since the literature has previously highlighted the role of maturity[30] and the average maturity of bond debt tends to be longer than for bank loans.[31] Although we do not have data on the full maturity structure of each firm, we can still construct a maturity ratio $M_{i,t}$, defined as the share of firm debt with maturity above 1 year. In our sample, the unconditional correlation of $M_{i,t}$ with the bond ratio $B_{i,t}$ is equal to 0.27.

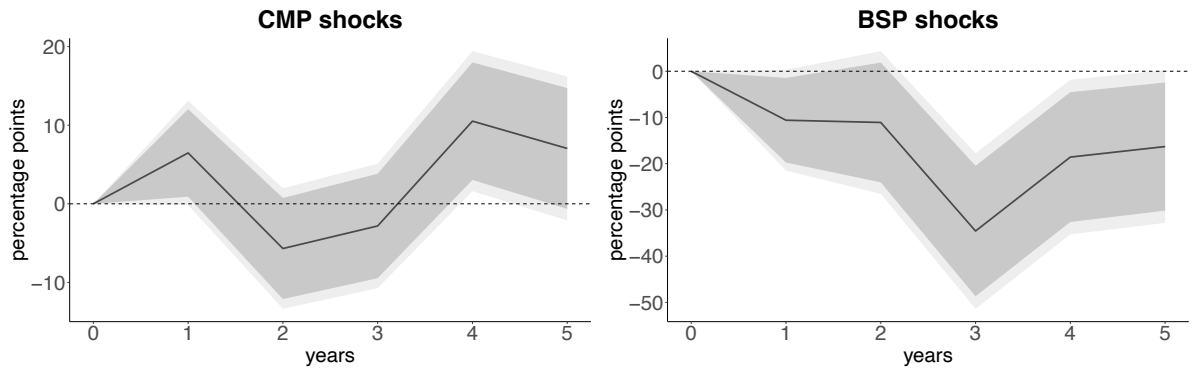
Since we are now interested only in the heterogeneity of responses, we can include sector-time fixed effects $\theta_{s,t}$ which will (among other things) absorb aggregate demand effects of monetary policy and any sector-specific responses to the shocks. All other variables, such as shocks, firm-specific controls and firm fixed effects, are as in Equation (3).

[29] Since monetary policy shocks are exogenous, in principle lagged aggregate controls are not strictly necessary.

[30] Deng and Fang (2022) show that firms who hold more long-term debt are less responsive to conventional monetary shocks. Using detailed bond-level date, Jungherr et al. (2022) show that firms with more maturing debt are more exposed to fluctuations in the real interest rate.

[31] It is also important to mention that 83% of debt of French companies is fixed-rate debt (Gueuder and Ray, 2022). See Gürkaynak, Karasoy-Can, and Lee (2022) for the impact of cash flow exposure on monetary policy transmission.

Figure 6: Heterogeneous response of investment to CMP (left panel) and BSP (right panel) shock depending on firms' bond share



Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure 6 shows the estimated coefficients for the interaction variable between monetary policy shocks and the lagged bond share in NFC debt. As the left panel of the graph indicates, after a contractionary 100bp CMP shock, firm investment falls less, the higher its share of market financing is. In particular, the contemporaneous decline in investment with respect to total assets of firms with no bonds is 6.4pp bigger on impact compared to fully bond reliant ones (i.e. comparing a firm with a bond ratio of 0 to one with a bond ratio equal to 1) and peaks at 10.5pp three years after the shock (year 4). On the other hand, after a contractionary 100bp BSP shock, firm investment falls more, the higher is its market financing share. On impact, the decline in investment of fully bond reliant firms is 10.6pp bigger compared to fully bank reliant ones, and this effect increases to 34.6pp at its peak, two years after the shock (year 3). In other words, a one standard deviation higher bond ratio (i.e by 0.16) would be associated with an extra 5.5 percentage point reduction in investment following a 100bp BSP shock.

In Figure A16 of Appendix A, we also provide the impulse response functions for the interaction terms between monetary policy shocks and maturity. We show in Figure A18 that the bond ratio results are also robust to excluding this maturity interaction from the specification. Figure A20 in Appendix A shows that results are also robust to adding the interaction of lagged bond share with the surprises in the German 10-year bond yields, using the same window on ECB announcement days. Finally, as in the previous section, we also provide figures that show that results are robust to aggregating shocks through a simple calendar year sum without accounting for the

month the firms report their results (Figure A15).

The transmission of monetary policy to firm investment is therefore contingent upon a firm's market debt structure and the specific monetary policy instrument employed. Conventional monetary policy has a stronger impact on firm investment when the firm is more reliant on bank loans, while unconventional policies that increase liquidity in bond markets (such as QE) have a stronger effect when firm financing is more market-based. To shed light on why this is the case, we explore in the next section the links between each type of credit supply and the two types of monetary policy shocks.

5. Inspecting the transmission channel

5.1. Impact on aggregate debt flows and prices

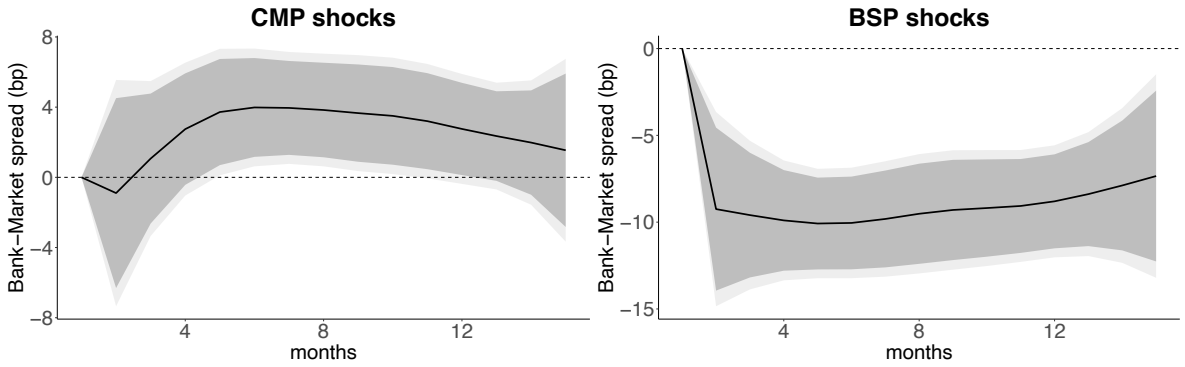
In Section 4.2., we established that the different types of monetary policy affect firms differently depending on their financing structure. In this section, we investigate the channels by looking at funding cost data. Unfortunately, we do not have data on firm-specific funding costs so we need to look at aggregate variables. On the other hand, this allows us to use monthly frequency which might be important when looking at financial variables.

Looking at aggregate variables reduces the sample size, which is why we use smooth local projections, as in Section 4.1. We explore the transmission channel in more detail by looking at the impact of monetary policy shocks on the cost of debt, as well as on the quantity dimension (flows and stocks).[32] In Figure 7, we show the response of the bank-market spread, defined as the rate of bank loans compared with the average yield of corporate bonds.

In the left panel, we see the response of the aggregate bank-market spread to a CMP shock. As monetary policy contracts, the spread seems to marginally and non-significantly fall in the short run, but quickly becomes persistently and significantly positive. As highlighted by Schnabel (2021), conventional shocks have indeed a *stronger* pass-through to bank loan rates relative to bond ones. After a BSP shock, market rates rise more than bank rates and therefore spreads are reduced. The impact of such unconventional shocks is then also *stronger* for bond markets. Firms facing these

[32]Monthly data on French NFC financing is published on Banque de France website: https://www.banque-france.fr/en/statistics/loans/financing-entreprises-2024-06.

Figure 7: Response of bank-market spreads to CMP (left panel) and BSP (right panel) shocks



Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: months after the shock.

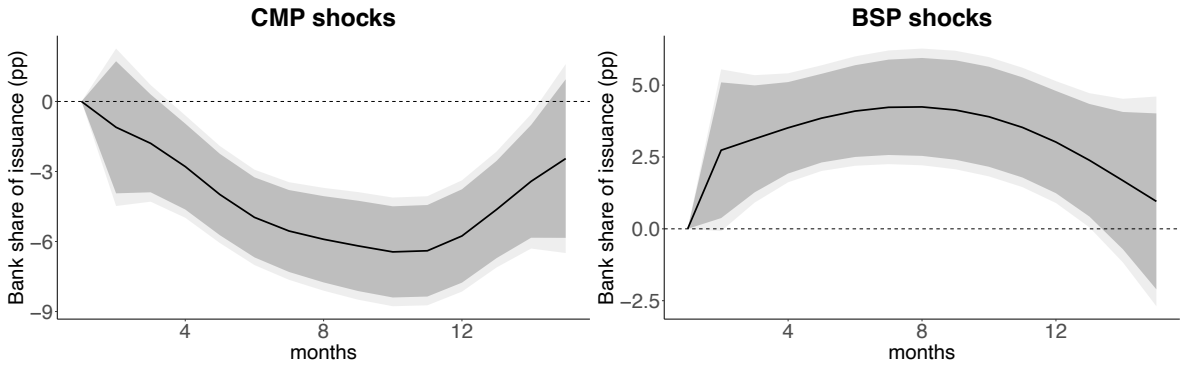
dynamics could then try to substitute bank debt with market debt after a CMP shock and conversely, they could substitute market debt with bank debt after a BSP shock.

The difference in speed of adjustment in spreads to each of the shocks is also consistent with Lane (2022) who argues that the pass-through is *faster* to bond market prices than to bank rates. After a contractionary CMP shock, a faster but smaller rise in bond yields could explain the delayed reaction in the left graph, while a slower and smaller rise in bank rates could also explain why the impact after the BSP shock peaks so soon. The impulse response can then shed light on how the two statements are not necessarily contradictory.

In Figure 8 we can see the impact on debt flows in response to the two shocks. In the left panel, we observe that the share of bank debt in new issuance falls after a CMP shock and in the right panel we see that it rises after a BSP shock. This is again consistent with the interpretation that there is segmented transmission and different pass-through of different shocks to different debt markets. The banking sector is more sensitive to CMP shocks and so interest rates hikes have a higher pass-through to bank loans. On the contrary, bond markets are more sensitive to BSP shocks, which have therefore a stronger pass-through to bond debt volumes than to bank loans. In Appendix A, we also show that the same effects can be observed in the relative stocks of debt (Figure A7) but also the absolute flows and not just the relative ones (Figures A8 and A9).

These results shed light on the channels explaining the results of our baseline

Figure 8: Response of bank share of issuance to CMP (left panel) and BSP (right panel) shocks



Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: months after the shock.

regressions. After a CMP tightening, firms that are more dependent on bank lending tend to contract investment significantly more than firms that have more access to bond markets, while BSP tightening affects bond-dependent firms relatively more. Given the reaction of quantities and prices, the two shocks act as relative supply shocks on each of the two markets: CMP for bank debt and BSP for bond debt. We also show that firms with high reliance on bond financing can use it as a "spare tyre" when faced with CMP shock. Yet, bond financing makes them more exposed to unconventional monetary policy tightening, in particular to shocks that impact the liquidity of bond markets, like our BSP shock.

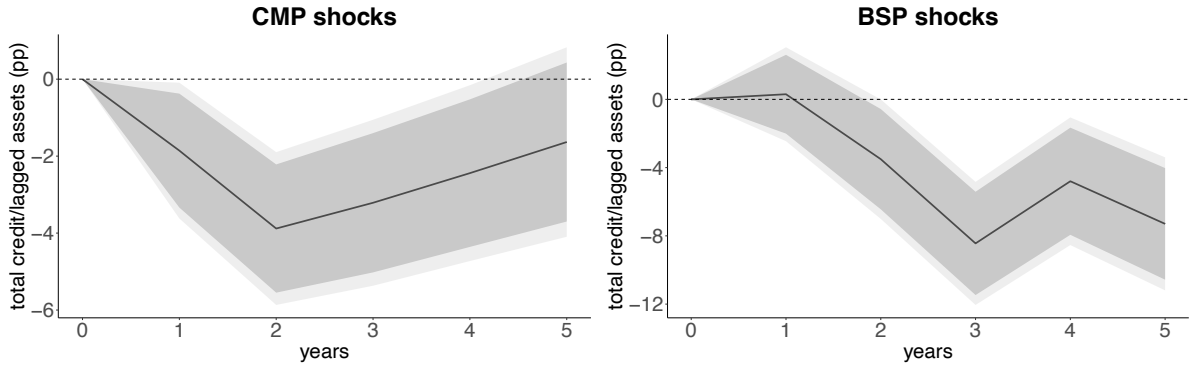
The two markets are not perfectly integrated and firms have difficulty substituting one for the other, irrespective of the direction required. As the left panel of Figure A8 and the right panel of Figure A9 in Appendix A show, there is some degree of substitutability since bond flows rise after a contractionary CMP shock and bank loan issuance grows after a contractionary BSP shock. However, this is not sufficient to stop the contractionary effects on aggregate investment, as highlighted in our baseline panel results.

5.2. Firm-level data: Total Credit

Although we do not have data on firm-level funding costs for each debt instrument, in this section we explore the panel data to shed additional light on the credit channel of monetary policy transmission. Consistent with our interpretation of this channel,

along with imperfect substitution across credit instruments, we expect total credit to fall across firms for all shocks, but CMP to have a stronger impact on firms that are more bank-based, while BSP to have a stronger impact on more market-based firms.

Figure 9: Average response of total credit to CMP (left panel) and BSP (right panel) shocks



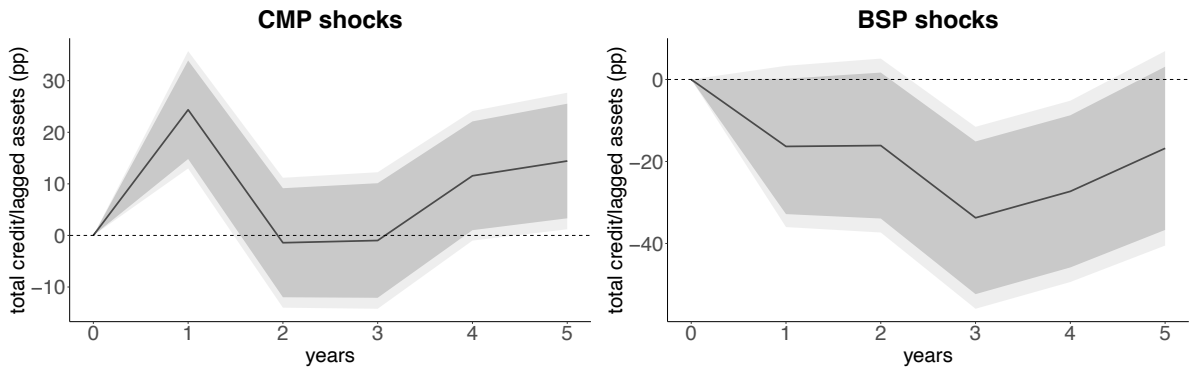
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

We first look at total firm credit, scaled by lagged assets, using the same controls and fixed effects as in Equation (3). Using LP to compute the impulse responses, we show in Figure 9 the estimated effect in percentage points of a 100 basis point upward surprise for CMP (left panel) and BSP (right panel) shocks. The left panel highlights that CMP shocks have an economically and statistically significant negative effect, with the total credit falling up to 3.9pp of lagged total assets 2 years after the shock. The BSP on the other hand leads to a fall in credit that reaches around 8.4pp of lagged assets 3 years after the shock. Unsurprisingly, both shocks are contractionary and lead to reductions in firm credit.

We then explore the cross-sectional heterogeneity, and let the panel results reveal the role of debt structure in the monetary policy transmission. To do so, we interact monetary policy shocks with the lagged bond ratio in NFC debt, including the same controls and fixed effects as in Equation (4). Figure 10 shows the estimated coefficients for the interaction variable between monetary policy shocks and the lagged bond ratio in NFC debt.

After a contractionary CMP shock (left panel), firms' total credit falls less, the higher the bond share is. On the other hand, after a contractionary BSP shock, firms' total credit falls more, the higher the bond share is. The transmission of conventional monetary policy to total credit is stronger for firms that are more dependent on bank financing,

Figure 10: Heterogeneous response of total credit to CMP (left panel) and BSP (right panel) shock depending on firms' bond ratio



Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

while it is weaker for those that have more market financing. On the other hand, those that are more market-based are more exposed to unconventional monetary policy shocks that affect liquidity in bond markets.[33]

6. Conclusion

In this paper, we identify significant heterogeneity in the transmission of monetary policy across firms. Using a novel approach to identify unconventional monetary policy shocks that are tightly linked to liquidity and a large panel of French firms, we show that while both conventional monetary policy and bond spread shocks reduce average firm investment, the strength of this effect depends on their debt structure.

Firms which are more reliant on bank credit contract investment relatively more after contractionary CMP shocks, but are affected less by contractionary bond spread shocks. This points to imperfect integration across the two debt markets. Using aggregate data, we show that there is substantial substitution between types of debt after each type of monetary policy shock. While there is some degree of substitutability, it is insufficient to stop the contractionary effect of monetary policy on NFC investment.

Heterogeneous monetary policy impact on firms' investment has important policy

[33] In Figure A17 of Appendix A, we also provide the impulse response functions of total credit for the interaction term between monetary policy shocks and maturity.

implications. Investment of large NFC with better access to capital markets could be more affected by quantitative tightening, while investment of smaller firms more reliant on bank loans would decrease more following a conventional tightening. In the absence of a coordinated approach, monetary policy can generate winners/losers depending on the tool used. On the other hand, policy can be more targeted when there are specific issues with one type of funding.

References

Adelino, M. and M. A. Ferreira (2016). "Bank Ratings and Lending Supply: Evidence from Sovereign Downgrades". *Review of Financial Studies* 29.7, pp. 1709–1746.

Adrian, T., P. Colla, and H. S. Shin (2013). "Which Financial Frictions? Parsing the Evidence from the Financial Crisis of 2007 to 2009". *NBER Macroeconomics Annual* 27, pp. 159–214.

Almeida, H. et al. (2017). "The Real Effects of Credit Ratings: The Sovereign Ceiling Channel". *Journal of Finance* 72.1, pp. 249–290.

Almon, S. (1965). "The Distributed Lag Between Capital Appropriations and Expenditures". *Econometrica* 33.1, pp. 178–196.

Altavilla, C., G. Carboni, and R. Motto (2021). "Asset Purchase Programs and Financial Markets: Lessons from the Euro Area". *International Journal of Central Banking* 17.70, pp. 1–48.

Altavilla, C. and D. Giannone (2017). "The Effectiveness of Non-Standard Monetary Policy Measures: Evidence from Survey Data". *Journal of Applied Econometrics* 32.5, pp. 952–964.

Altavilla, C. et al. (2019). "Measuring euro area monetary policy". *Journal of Monetary Economics* 108, pp. 162–179.

Arce, O., S. Mayordomo, and R. Gimeno (2020). "Making Room for the Needy: The Credit-Reallocation Effects of the ECB's Corporate QE". *Review of Finance* 25.1, pp. 43–84.

Ashcraft, A. B. and M. Campello (2007). "Firm balance sheets and monetary policy transmission". *Journal of Monetary Economics* 54.6, pp. 1515–1528.

Barclay, M. J. and C. W. Smith Jr (1995). "The Maturity Structure of Corporate Debt". *Journal of Finance* 50.2, pp. 609–631.

Barnichon, R. and C. Brownlees (2019). "Impulse Response Estimation by Smooth Local Projections". *The Review of Economics and Statistics* 101.3, pp. 522–530.

Becker, B. and V. Ivashina (2014). "Cyclicality of credit supply: Firm level evidence". *Journal of Monetary Economics* 62, pp. 76–93.

Bernanke, B. S. and A. S. Blinder (1988). "Credit, Money, and Aggregate Demand". *American Economic Review* 78.2, pp. 435–439.

Bernanke, B. S. and M. Gertler (1995). "Inside the Black Box: The Credit Channel of Monetary Policy Transmission". *Journal of Economic Perspectives* 9.4, pp. 27–48.

Bianchi, F., H. Kung, and G. Morales (2019). "Growth, slowdowns, and recoveries". *Journal of Monetary Economics* 101.C, pp. 47–63.

Bolton, P. and D. S. Scharfstein (1996). "Optimal Debt Structure and the Number of Creditors". *Journal of Political Economy* 104.1, pp. 1–25.

Borensztein, E., K. Cowan, and P. Valenzuela (2013). "Sovereign ceilings "lite"? The impact of sovereign ratings on corporate ratings". *Journal of Banking & Finance* 37.11, pp. 4014–4024.

Bräuning, F., J. Fillat, and J. C. Wang (2020). *Corporate Debt Maturity and Monetary Policy*. Current Policy Perspectives 88967. Federal Reserve Bank of Boston.

Chitu, L. et al. (2023). *Financial Shock Transmission to Heterogeneous Firms: The Earnings-Based Borrowing Constraint Channel*. Working Paper Series 2860. European Central Bank.

Cloyne, J. et al. (2023). "Monetary policy, corporate finance, and investment". *Journal of the European Economic Association* 21.6, pp. 2586–2634.

Crouzet, N. (2017). "Aggregate Implications of Corporate Debt Choices". *The Review of Economic Studies* 85.3, pp. 1635–1682.

— (2021). "Credit Disintermediation and Monetary Policy". *IMF Economic Review* 69.1, pp. 23–89.

Darmouni, O., O. Giesecke, and R. Rodnyansky (2020). *The Bond Lending Channel of Monetary Policy*. Cambridge Working Papers in Economics 2049. Faculty of Economics, University of Cambridge.

Darmouni, O. and M. Papoutsi (2022). *The rise of bond financing in Europe*. Working Paper Series 2663. European Central Bank.

De Fiore, F. and H. Uhlig (2015). "Corporate Debt Structure and the Financial Crisis". *Journal of Money, Credit and Banking* 47.8, pp. 1571–1598.

De Santis, R. A. and A. Zaghini (2021). "Unconventional monetary policy and corporate bond issuance". *European Economic Review* 135, p. 103727.

Dees, S. et al. (2022). *The Financing Structure of Non-Financial Corporations and Macro-Financial Implications in France*. Working papers 880. Banque de France.

Deng, M. and M. Fang (2022). "Debt maturity heterogeneity and investment responses to monetary policy". *European Economic Review* 144.C.

Diamond, D. W. and Z. He (2014). "A Theory of Debt Maturity: The Long and Short of Debt Overhang". *Journal of Finance* 69.2, pp. 719–762.

Durante, E., A. Ferrando, and P. Vermeulen (2022). "Monetary policy, investment and firm heterogeneity". *European Economic Review* 148, p. 104251.

ECB (2009). *ECB Monthly Bulletin*, September.

Ejsing, J. and J. Sihvonen (2009). *Liquidity premia in German government bonds*. Working Paper Series 1081. European Central Bank.

Ferreira, T., D. Ostry, and J. H. Rogers (2023). *Firm financial conditions and the transmission of monetary policy*. Finance and Economics Discussion Series 2023-037. Washington: Board of Governors of the Federal Reserve System.

Gerko, E. and H. Rey (2017). "Monetary Policy in the Capitals of Capital". *Journal of the European Economic Association* 15.4, pp. 721–745.

Gertler, M. and S. Gilchrist (1994). "Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms". *The Quarterly Journal of Economics* 109.2, pp. 309–340.

Gertler, M. and P. Karadi (2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity". *American Economic Journal: Macroeconomics* 7.1, pp. 44–76.

Giambona, E. et al. (2020). *Quantitative easing, investment, and safe assets: the corporate-bond lending channel*. Economics Working Papers 1722. Department of Economics and Business, Universitat Pompeu Fabra.

Greenspan, A. (1999). *Do efficient financial markets mitigate financial crises?* Remarks before the 1999 Financial Markets Conference of the Federal Reserve Bank of Atlanta, Sea Island, Georgia. October 19, 1999.

Grosse-Rueschkamp, B., S. Steffen, and D. Streitz (2019). "A Capital Structure Channel of Monetary Policy". *Journal of Financial Economics* 133.2, pp. 357–378.

Gueuder, M. and S. Ray (2022). "Rise in interest rates: European companies will not be affected at the same pace [Hausse des taux d'intérêt : les entreprises européennes ne seront pas affectées au même rythme]". *Bulletin de la Banque de France* 243.

Gürkaynak, R., H. G. Karasoy-Can, and S. S. Lee (2022). "Stock Market's Assessment of Monetary Policy Transmission: The Cash Flow Effect". *The Journal of Finance* 77.4, pp. 2375–2421.

Gürkaynak, R. S., B. Sack, and E. Swanson (2005). "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements". *International Journal of Central Banking* 1.1.

Holm-Hadulla, F. and C. Thürwächter (2021). "Heterogeneity in corporate debt structures and the transmission of monetary policy". *European Economic Review* 136.C.

Hubert, P. et al. (2024). "Same actions, different effects: The conditionality of monetary policy instruments". *Journal of Monetary Economics*, p. 103596.

Ippolito, F., A. K. Ozdagli, and A. Perez-Orive (2018). "The transmission of monetary policy through bank lending: The floating rate channel". *Journal of Monetary Economics* 95, pp. 49–71.

Jarociński, M. and P. Karadi (2020). "Deconstructing Monetary Policy Surprises—The Role of Information Shocks". *American Economic Journal: Macroeconomics* 12.2, pp. 1–43.

Jeenas, P. (2023). *Firm balance sheet liquidity, monetary policy shocks, and investment dynamics*. Universitat Pompeu Fabra, Department of Economics and Business.

Jiménez, G. et al. (2012). "Credit Supply and Monetary Policy: Identifying the Bank Balance-Sheet Channel with Loan Applications". *American Economic Review* 102, pp. 2301–26.

Jordà, Òscar (2005). "Estimation and Inference of Impulse Responses by Local Projections". *The American Economic Review* 95.1, pp. 161–182.

Jungherr, J. et al. (2022). "Corporate debt maturity matters for monetary policy". *Econometrica* 1.4, p. 345.

Kashyap, A. K., J. C. Stein, and D. W. Wilcox (1993). "Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance". *American Economic Review* 83.1, pp. 78–98.

Kishan, R. P. and T. P. Opiela (2000). "Bank Size, Bank Capital, and the Bank Lending Channel". *Journal of Money, Credit and Banking* 32.1, pp. 121–141.

Kuttner, K. N. (2001). "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market". *Journal of Monetary Economics* 47.3, pp. 523–544.

Lane, P. R. (2022). *The transmission of monetary policy*. Speech at the 7th SUERF - CGEG | Columbia | SIPA - EIB - Société Générale Conference on "EU and US Perspectives: New Directions for Economic Policy", New York City, 11 October 2022.

Lee, A. S. and C. Engel (2024). *U.S. Liquid Government Liabilities and Emerging Market Capital Flows*. Working Paper 32572. National Bureau of Economic Research.

Lhuissier, S. and U. Szczerbowicz (2022). "Monetary Policy and Corporate Debt Structure". *Oxford Bulletin of Economics and Statistics* 84.3, pp. 497–515.

Li, D., M. Plagborg-Møller, and C. K. Wolf (2024). "Local projections vs. vars: Lessons from thousands of dgps". *Journal of Econometrics*, p. 105722.

Lo Duca, M., G. Nicoletti, and A. Vidal Martínez (2016). "Global corporate bond issuance: What role for US quantitative easing?" *Journal of International Money and Finance* 60.C, pp. 114–150.

Nakamura, E. and J. Steinsson (2018). "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect". *The Quarterly Journal of Economics* 133.3, pp. 1283–1330.

Ottonello, P. and T. Winberry (2020). "Financial Heterogeneity and the Investment Channel of Monetary Policy". *Econometrica* 88.6, pp. 2473–2502.

Ramey, V. (2016). "Chapter 2 - Macroeconomic Shocks and Their Propagation". Ed. by J. B. Taylor and H. Uhlig. Vol. 2. Handbook of Macroeconomics. Elsevier, pp. 71–162.

Schnabel, I. (2021). *The rise of non-bank finance and its implications for monetary policy transmission.* Speech at the Annual Congress of the European Economic Association (EEA), Frankfurt am Main, 24 August 2021.

Schwarz, K. (2018). "Mind the Gap: Disentangling Credit and Liquidity in Risk Spreads". *Review of Finance* 23.3, pp. 557–597.

Stein, J. C. and A. K. Kashyap (2000). "What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?" *American Economic Review* 90.3, pp. 407–428.

Szczerbowicz, U. (2015). "The ECB Unconventional Monetary Policies: Have They Lowered Market Borrowing Costs for Banks and Governments?" *International Journal of Central Banking* 11.4.

Wright, J. H. (2012). "What does Monetary Policy do to Long-term Interest Rates at the Zero Lower Bound?" *The Economic Journal* 122.564, F447–F466.

Appendix A

Figure A1: Additional histograms

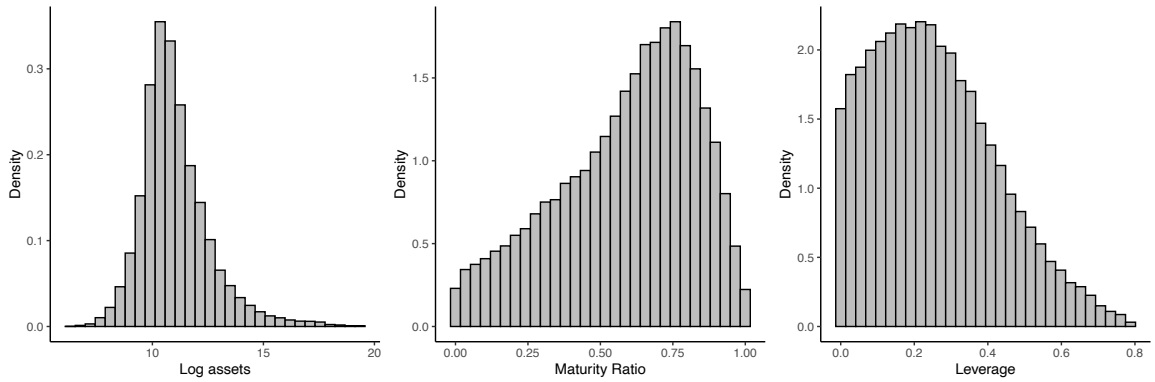


Figure A2: Binned scatterplots of bond ratios by cash flow over assets

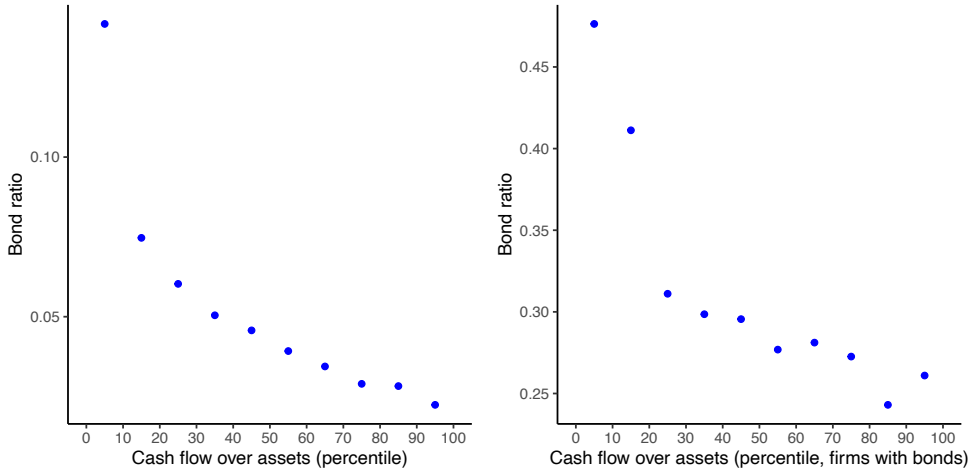


Figure A3: CMP shocks (left) and BSP shocks (right) on the days of ECB Governing Council meetings

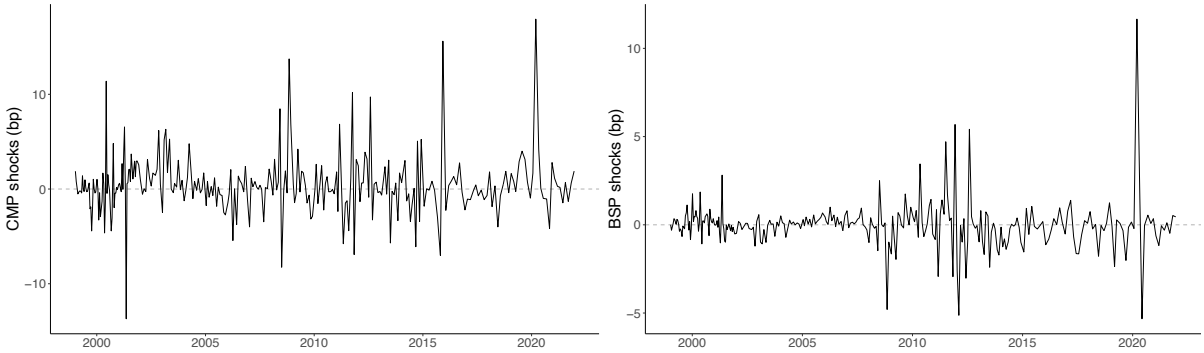
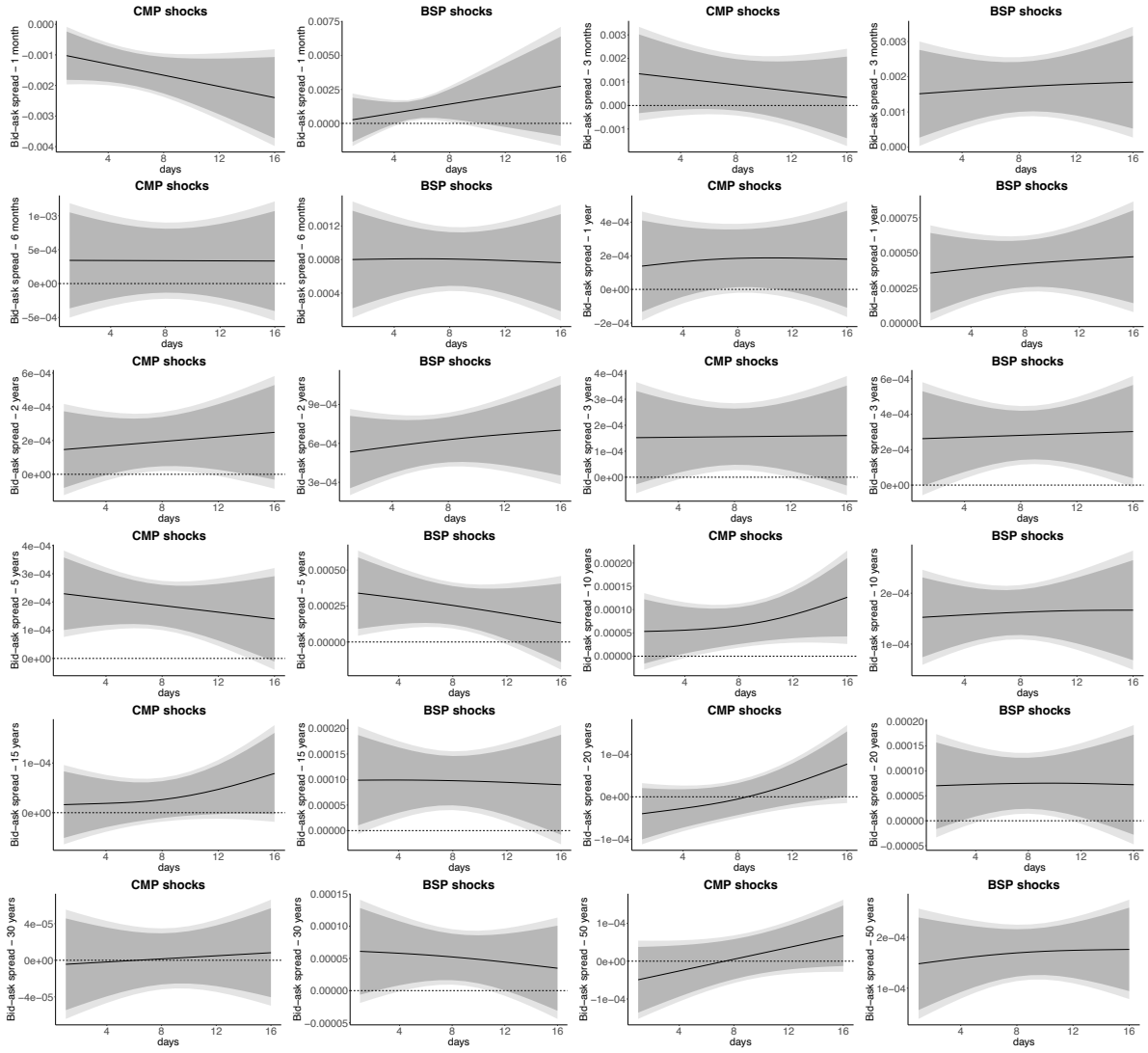
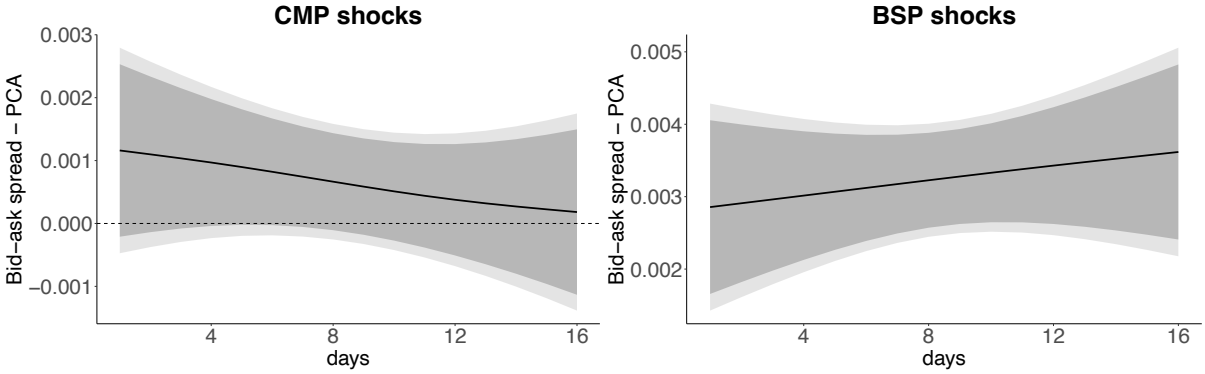


Figure A4: Response of bid-ask spreads at different maturities to to CMP and BSP shocks



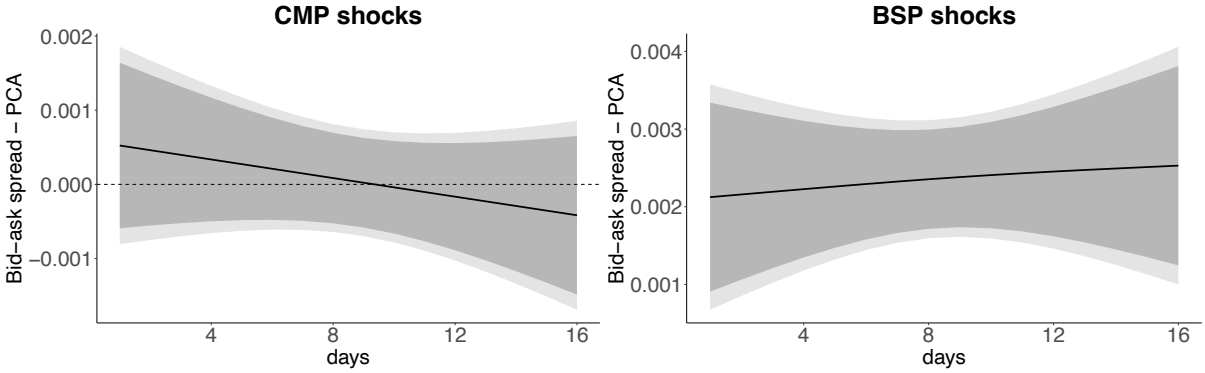
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: days after the shock.

Figure A5: Response of first principal component of bid-ask spreads to CMP (left panel) and BSP (right panel) shocks (excluding 2020 onwards)



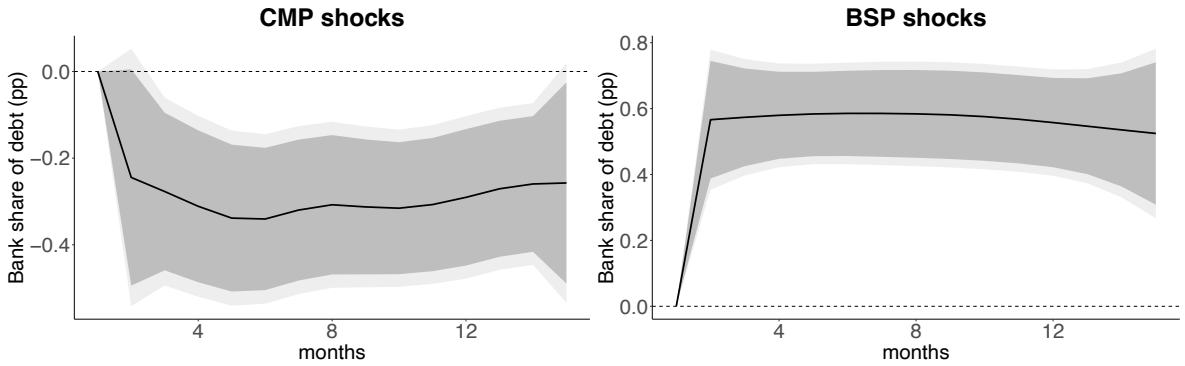
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: days after the shock.

Figure A6: Response of first principal component of bid-ask spreads to CMP (left panel) and BSP (right panel) shocks (with 2008 dummy)



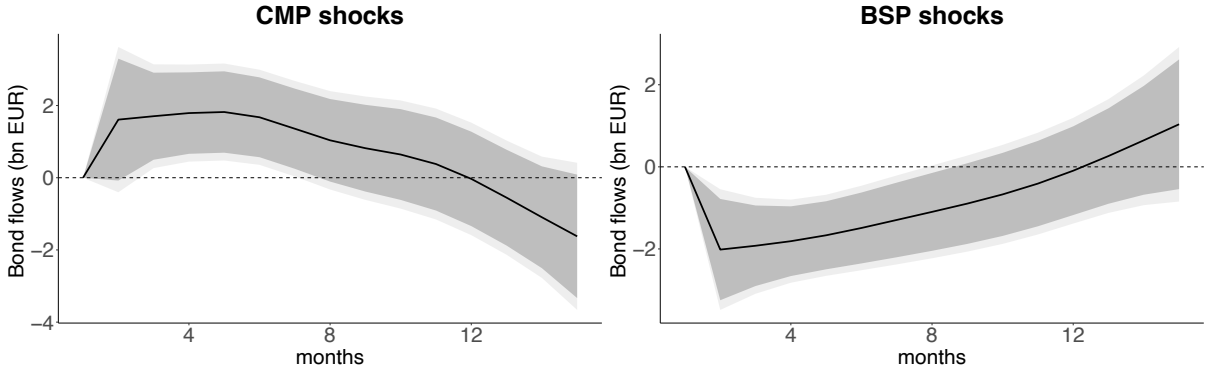
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: days after the shock.

Figure A7: Response of bank share of debt to CMP (left panel) and BSP (right panel) shocks



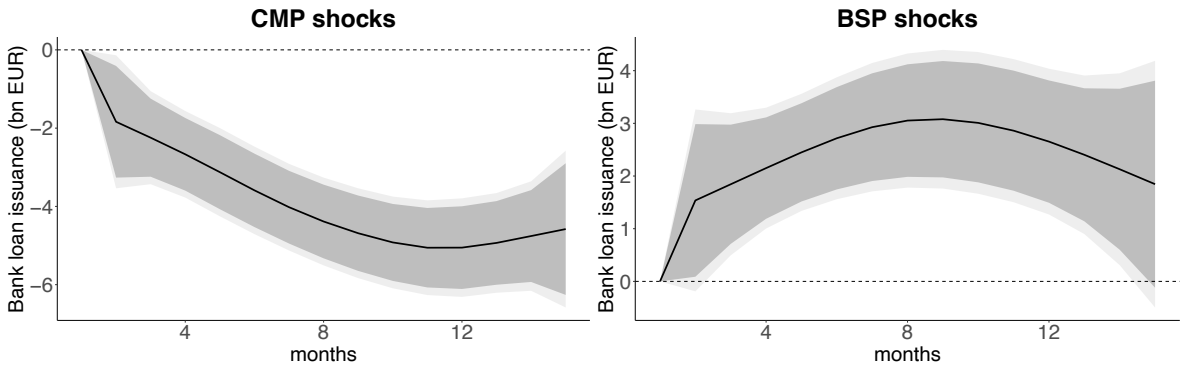
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: months after the shock.

Figure A8: Response of bond issuance CMP (left panel) and BSP (right panel) shocks



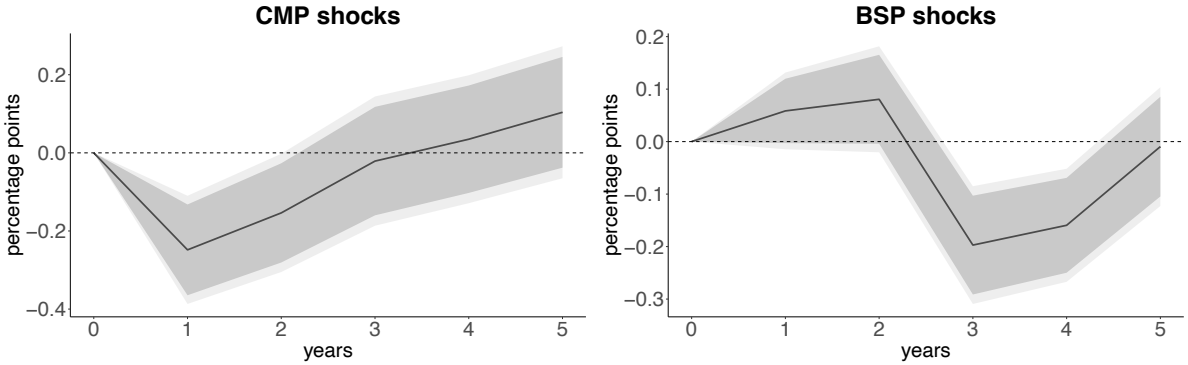
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: months after the shock.

Figure A9: Response of bank loan flows to CMP (left panel) and BSP (right panel) shocks



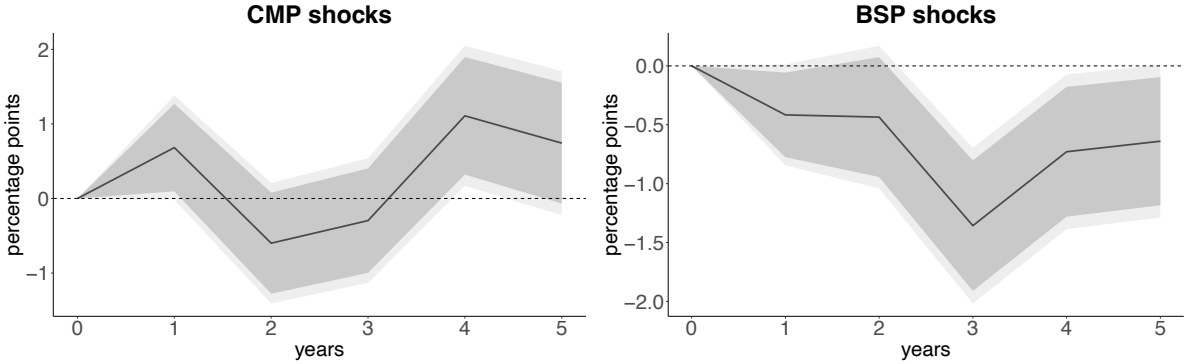
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: months after the shock.

Figure A10: Average response of investment to CMP (left panel) and BSP (right panel) shocks (shocks normalized by their standard deviation)



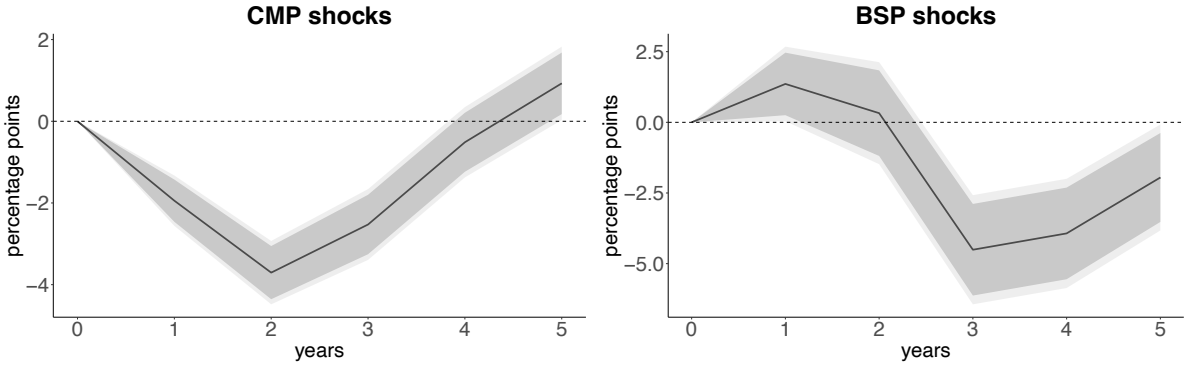
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure A11: Heterogeneous response of investment to CMP (left panel) and BSP (right panel) shock depending on firms' bond share
(shocks normalized by their standard deviation)



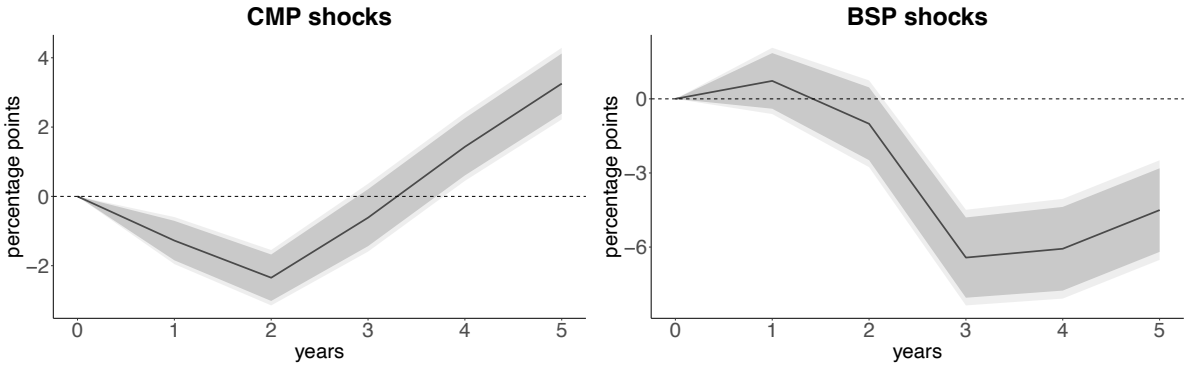
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure A12: Average response of investment to CMP (left panel) and BSP (right panel) shocks
(no aggregate controls)



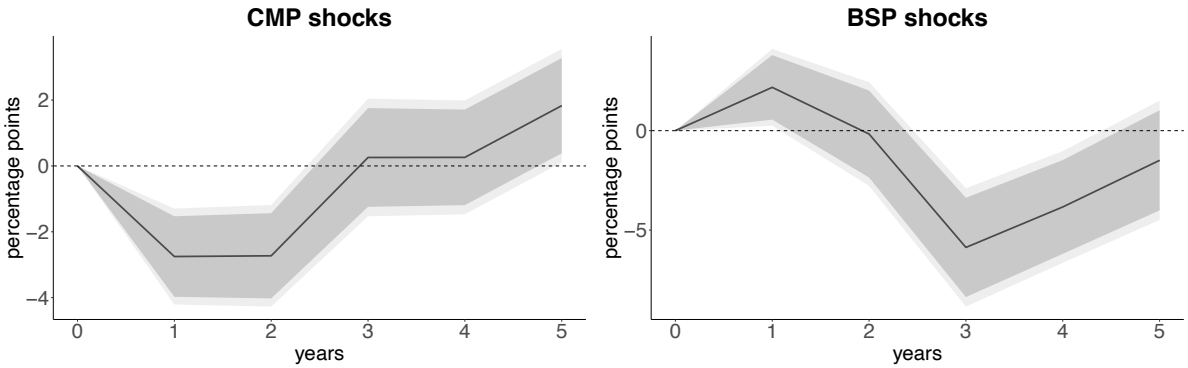
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure A13: Average response of investment to CMP (left panel)
and BSP (right panel) shocks
(only output gap and inflation as controls)



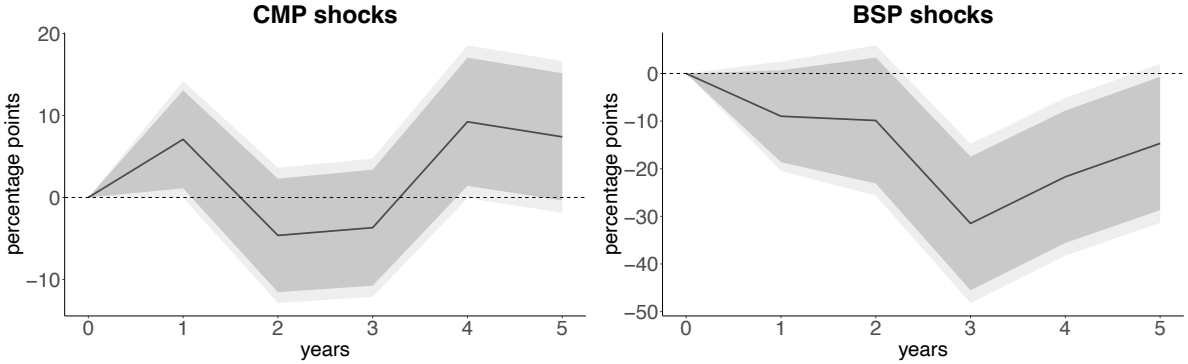
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals.
X-axis: years after the shock.

Figure A14: Average response of investment to CMP (left panel)
and BSP (right panel) shocks
(yearly non-firm-specific shocks)



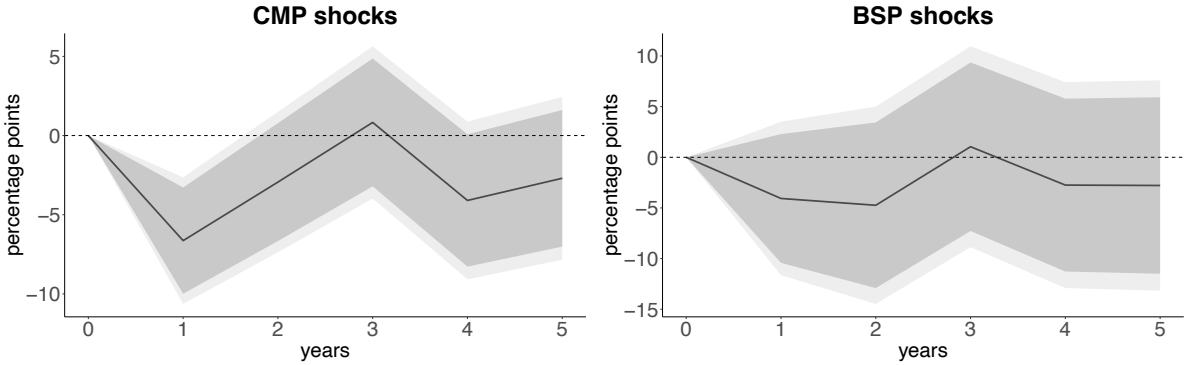
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals.
X-axis: years after the shock.

Figure A15: Heterogeneous response of investment to CMP (left panel) and BSP (right panel) shock depending on firms' bond share (yearly non-firm-specific shocks)



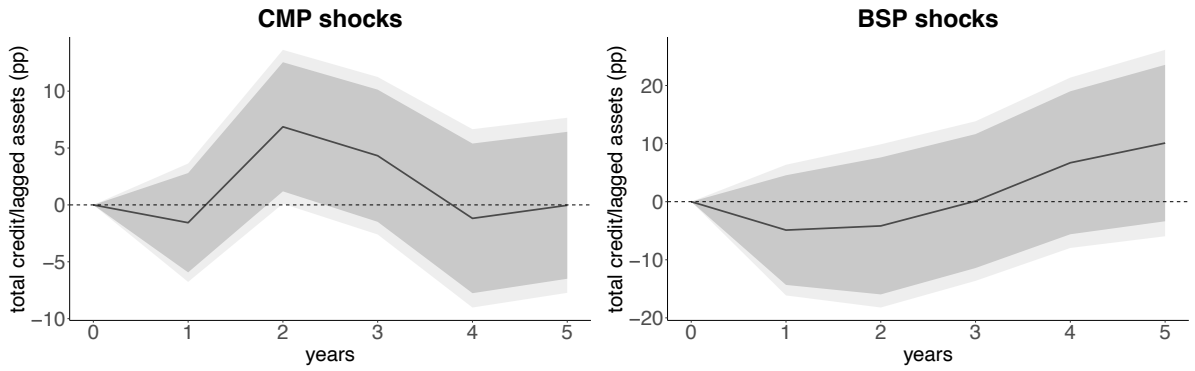
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure A16: Heterogeneous response of investment to CMP (left panel) and BSP (right panel) depending on firms' share of long-term debt $M_{i,t}$



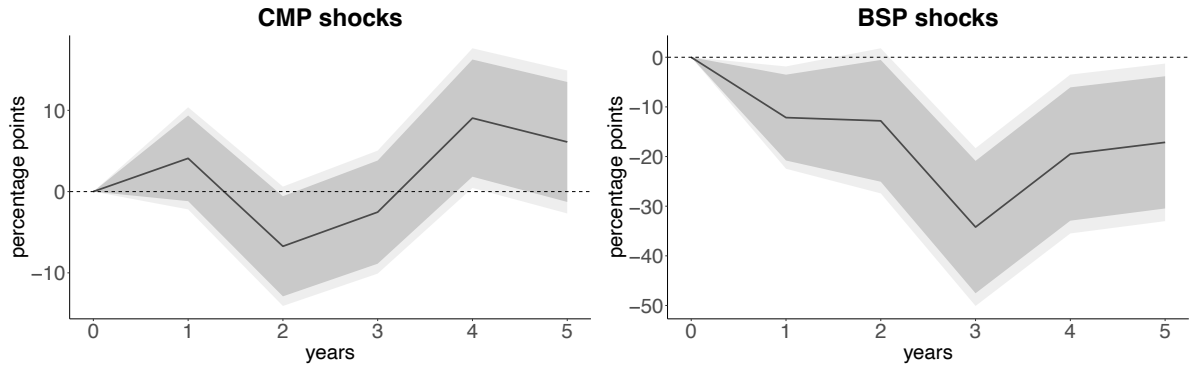
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure A17: Heterogeneous response of total credit to CMP (left panel) and BSP (right panel) depending on firms' share of long-term debt $M_{i,t}$



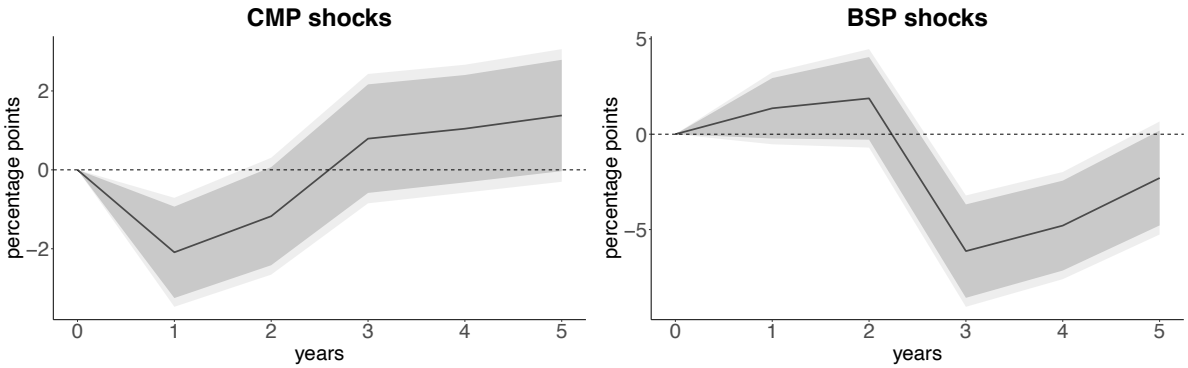
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure A18: Heterogeneous response of investment to CMP (left panel) and BSP (right panel) shock depending on firms' bond share (no maturity interaction)



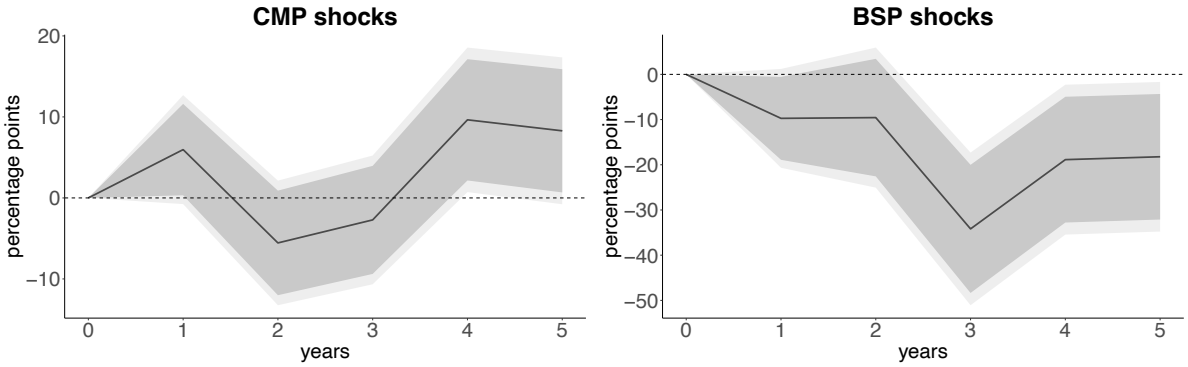
Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure A19: Average response of investment to CMP (left panel) and BSP (right panel) shocks (controlling for surprises in 10y German yield)



Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Figure A20: Heterogeneous response of investment to CMP (left panel) and BSP (right panel) shock depending on firms' bond share (controlling for surprises in 10y German yield)



Note: Light (dark) grey bands correspond to 95% (to 90%) confidence intervals. X-axis: years after the shock.

Corporate debt structure and heterogeneous monetary policy transmission

Marie Alder*, Nuno Coimbra**, Urszula Szczerbowicz***

* European University Institute
** Banque de France
*** SKEMA Business School

Workshop "New insights from financial statements"

The views expressed in this paper are solely those of the authors and do not represent the views of the Eurosystem or the Banque de France

Corporate bonds since the Global Financial Crisis

- Share of bond debt in non-financial corporations (NFCs) has increased since GFC
- Rise in bond share of corporate debt particularly high in the Euro Area (EA)
  - Almost doubled between 2007 and 2021 (9% to 17%)
  - France: from 19% to 30%
  - Spain: from 3% to 15% in 2021

Corporate bonds since the Global Financial Crisis

- Share of bond debt in non-financial corporations (NFCs) has increased since GFC
- Rise in bond share of corporate debt particularly high in the Euro Area (EA)
  - Almost doubled between 2007 and 2021 (9% to 17%)
  - France: from 19% to 30%
  - Spain: from 3% to 15% in 2021

$\Rightarrow$ How much does debt structure matter for MP transmission?

- Does it depend on the type of monetary policy?

▸ Literature ▸ Contribution

Corporate debt structure and monetary policy transmission

Debt structure is important in the view of policy-makers:

- **P. Lane (2022)**: "As a more bank-based system, the euro area might entail a more delayed reaction through the interest rate channel, as compared to countries where firms finance themselves predominantly with market-based debt."

Corporate debt structure and monetary policy transmission

Debt structure is important in the view of policy-makers:

- **P. Lane (2022)**: "As a more bank-based system, the euro area might entail a more delayed reaction through the interest rate channel, as compared to countries where firms finance themselves predominantly with market-based debt."

- **I. Schnabel (2021)**: "the overall cost of credit is more responsive to conventional monetary policy measures in these [bank-based] economies than in economies with a higher share of bond finance."

Corporate debt structure and monetary policy transmission

Debt structure is important in the view of policy-makers:

- **P. Lane (2022)**: "As a more bank-based system, the euro area might entail a more delayed reaction through the interest rate channel, as compared to countries where firms finance themselves predominantly with market-based debt."

- **I. Schnabel (2021)**: "the overall cost of credit is more responsive to conventional monetary policy measures in these [bank-based] economies than in economies with a higher share of bond finance."

$\Rightarrow$ Affects speed and responsiveness of MP transmission

▸ Literature ▸ Contribution

Findings

Does NFC debt structure matter for the transmisson of monetary policy to French firms' investment?

Findings

Does NFC debt structure matter for the transmisson of monetary policy to French firms' investment? Yes.

Findings

Does NFC debt structure matter for the transmisson of monetary policy to French firms' investment? Yes.

- Firms more reliant on bank credit:
  - Contract investment relatively **more** after contractionary conventional monetary policy,
  - yet reduce investment relatively **less** after contractionary bond spread (BSP) shocks.

Findings

Does NFC debt structure matter for the transmisson of monetary policy to French firms' investment? Yes.

- Firms more reliant on bank credit:
  - Contract investment relatively **more** after contractionary conventional monetary policy,
  - yet reduce investment relatively **less** after contractionary bond spread (BSP) shocks.

Inspecting the mechanism:

- CMP tightening $\Rightarrow$ bank loan rates increase relatively **more** compared with corporate bond yields, bank loan issuance **declines**.
- BSP tightening $\Rightarrow$ bond rates increase relatively **more**, bond issuance **declines**.

Findings

Does NFC debt structure matter for the transmisson of monetary policy to French firms' investment? Yes.

- Firms more reliant on bank credit:
  - Contract investment relatively **more** after contractionary conventional monetary policy,
  - yet reduce investment relatively **less** after contractionary bond spread (BSP) shocks.

Inspecting the mechanism:

- CMP tightening $\Rightarrow$ bank loan rates increase relatively **more** compared with corporate bond yields, bank loan issuance **declines**.
- BSP tightening $\Rightarrow$ bond rates increase relatively **more**, bond issuance **declines**.

$\Rightarrow$ *Akin to funding supply shocks in each market (CMP $\leftrightarrow$ bank loans, BSP $\leftrightarrow$ bonds)*

High-frequency Identification of Monetary Policy Shocks

Conventional Monetary Policy shocks

CMP shocks: updated Jarociński and Karadi (2020) shocks

- Unexpected changes in assets around the ECB announcements
  - First principal component of Overnight Index Swaps (OIS) with maturities up to 1 year
- Monetary Event-window as in Altavilla et al. (2019)
  - Median quote from 13:25-13:35 compared to median quote 15:40-15:50
- Cleaned from CB information effects

▸ Decomposition

Bond Spread shocks (BSP)

Overview

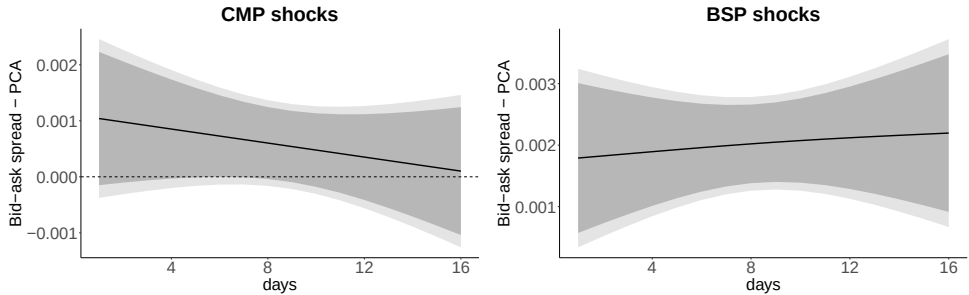
- Capture unconventional monetary policy shocks connected to FR bond markets
- High-frequency changes in 10-Year France-Germany sovereign spread
  - Around ECB announcements
  - Monetary Event-window from Altavilla et al. (2019)
- Further orthogonalized with respect to CMP surprises
  - Remove structural impact of CMP on spreads and liquidity

▸ Motivation ▸ Time Series ▸ Narrative description

Bond Spread shocks and bid-ask spreads

Strong link between BSP shocks and bid-ask spreads of FR sovereign bonds

- First Principal Component across all maturities (FR bonds): ▸ Individual maturities



▸ Details

Baseline results

FIBEN dataset

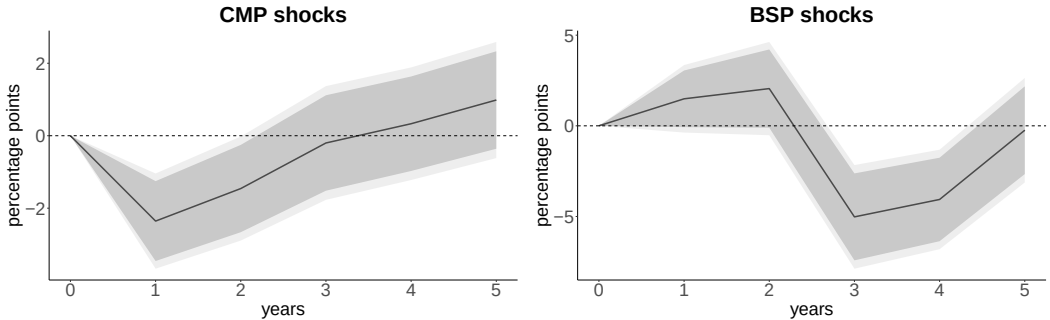
- Yearly firm-level data on French companies from the FIBEN consolidated database
  - Data collected from corporate tax declarations (*liasse fiscale*)
- >80k observations for >11k different firms
  - Unbalanced panel, firm entry and exit from dataset
- Yearly aggregation of shocks aligned to reporting month of each firm
- **Net investment rate**: first difference of net tangible assets (NTA) in year t, divided by total assets (A) in year $t-1$

$$I_{i,t} = \frac{NTA_{i,t} - NTA_{i,t-1}}{A_{i,t-1}}$$

▸ Hypotheses

MP impact on French NFC investment

Average response of investment to CMP and Bond Spread shocks ▸ Specification

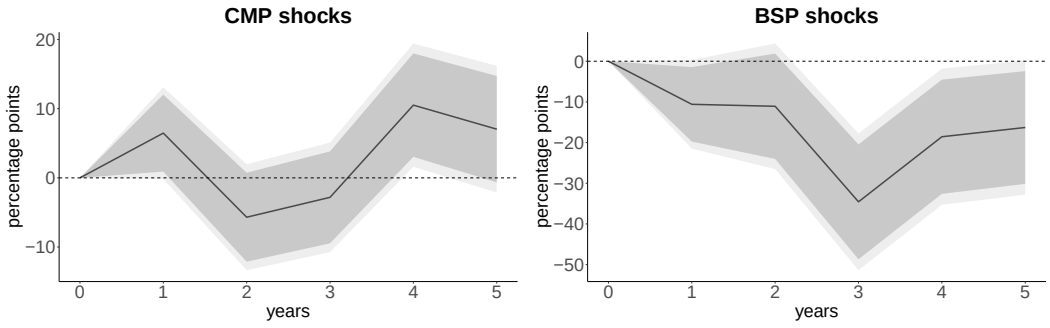


Note: estimated effect of a 100 bp upward surprise for CMP (left panel) and BSP (right panel) shocks on the net investment rate at the firm level, at each horizon h (up to 5 years).

- ↗ 100bp CMP ⇒ 2.4pp ↘ of investment wrt firm's total assets
- ↗ 100bp BSP ⇒ 5.0pp ↘ of investment wrt firm's total assets

MP and Investment: the role of debt structure

Interacting shocks with firms' bond ratios ▸ Specification



Note: IRFs for the interaction term of the MP shock with the lagged bond share. 100 bp upward surprise for CMP (left panel) and BSP (right panel) shocks, at each horizon h (up to 5 years).

MP and Investment: the role of bond share

- Contractionary CMP: investment rate falls less, the higher its share of bond financing
  - Contemporaneous decline of a fully bond reliant firm ($B_{i,t-1} = 1$) is 6.4pp less per 100bp than a fully bank reliant one ($B_{i,t-1} = 0$)
  - Effect peaks at 10.5pp during the 3rd year after the shock (1.11pp per stdev)

MP and Investment: the role of bond share

- Contractionary CMP: investment rate falls less, the higher its share of bond financing
  - Contemporaneous decline of a fully bond reliant firm ($B_{i,t-1} = 1$) is 6.4pp less per 100bp than a fully bank reliant one ($B_{i,t-1} = 0$)
  - Effect peaks at 10.5pp during the 3rd year after the shock (1.11pp per stdev)

  $\Rightarrow$ **Schnabel**: *bank-based economy more responsive to CMP*

MP and Investment: the role of bond share

- Contractionary CMP: investment rate falls less, the higher its share of bond financing
  - Contemporaneous decline of a fully bond reliant firm ($B_{i,t-1} = 1$) is 6.4pp less per 100bp than a fully bank reliant one ($B_{i,t-1} = 0$)
  - Effect peaks at 10.5pp during the 3rd year after the shock (1.11pp per stdev)

  $\Rightarrow$ **Schnabel**: *bank-based economy more responsive to CMP*

- Contractionary BSP: investment falls more, the higher its share of bond financing
  - Contemporaneous decline in investment rates of fully bond reliant firms is 10.5pp larger per 100bp than a fully bank reliant one
  - Effect peaks at 34.6pp during the 3rd year after the shock (1.36pp per stdev)

MP and Investment: the role of bond share

- Contractionary CMP: investment rate falls less, the higher its share of bond financing
  - Contemporaneous decline of a fully bond reliant firm ($B_{i,t-1} = 1$) is 6.4pp less per 100bp than a fully bank reliant one ($B_{i,t-1} = 0$)
  - Effect peaks at 10.5pp during the 3rd year after the shock (1.11pp per stdev)

  $\Rightarrow$ **Schnabel**: *bank-based economy more responsive to CMP*

- Contractionary BSP: investment falls more, the higher its share of bond financing
  - Contemporaneous decline in investment rates of fully bond reliant firms is 10.5pp larger per 100bp than a fully bank reliant one
  - Effect peaks at 34.6pp during the 3rd year after the shock (1.36pp per stdev)

Robustness: ▸ Calendar aggregation ▸ Other aggregate Controls ▸ No aggregate Controls ▸ Excluding Maturity

Inspecting the Transmission Channels

Inspecting the mechanism

Look at the impact of the two types on shocks on:

- **Total credit** of French firms (panel) ▸ Average impact ▸ Role of bond share
  - Similar results: funding $\leftrightarrow$ investment

Inspecting the mechanism

Look at the impact of the two types on shocks on:

- **Total credit** of French firms (panel) ▸ Average impact ▸ Role of bond share
  - Similar results: funding $\leftrightarrow$ investment
- **Relative cost** of bonds and bank loans (aggregate) ▸ Results
  - CMP tightening $\Rightarrow$ relative cost of bank loans rises
  - BSP tightening $\Rightarrow$ relative cost of bank loans falls

Inspecting the mechanism

Look at the impact of the two types on shocks on:

- **Total credit** of French firms (panel) ▸ Average impact ▸ Role of bond share
  - Similar results: funding $\leftrightarrow$ investment
- **Relative cost** of bonds and bank loans (aggregate) ▸ Results
  - CMP tightening $\Rightarrow$ relative cost of bank loans rises
  - BSP tightening $\Rightarrow$ relative cost of bank loans falls
- **Debt issuance** in bonds and bank loans (aggregate) ▸ Relative flows ▸ Absolute flows
  - CMP tightening $\Rightarrow$ bank loan issuance *falls*, bond issuance *rises*
  - BSP tightening $\Rightarrow$ bank loan issuance *rises*, bond issuance *falls*
  - Aggregate flows contract but some degree of **substitutability**

Conclusions

- Contractionary CMP and BSP shocks decrease investment of French firms
- MP impact on firm investment depends on firm debt structure:
  - Firms which are more reliant on bank credit contract investment relatively more after contractionary CMP shocks,
  - but contract investment less after contractionary BSP shocks
- Imperfect integration across the two markets
  - Some degree of substitutability,
  - but not enough to undo impact on NFC investment and credit

Conclusions

Key messages for statistical producers

- Data granularity and firm-level analysis
  - Importance of firm-level data to capture heterogeneous impact of MP
  - Lack of data on firm-specific funding costs for each debt instrument
- Data frequency
  - Monthly frequency can be important when looking at financial variables (firm credit...)
- Need for more loan-level data

Appendix

Related Literature

Go Back

Financial frictions and investment

- Cloyne et al. (2023) find that financial frictions account for about one third of the aggregate investment response to monetary policy.
- Firm-level response to MP shock depends on default risk (Ottonello and Winberry, 2020), the firms' age (Cloyne et al., 2023), firms' size (Gertler and Gilchrist, 1994) and their holdings of liquid assets (Jeenas, 2019).

Related Literature

Go Back

Financial frictions and investment

- Cloyne et al. (2023) find that financial frictions account for about one third of the aggregate investment response to monetary policy.
- Firm-level response to MP shock depends on default risk (Ottonello and Winberry, 2020), the firms' age (Cloyne et al., 2023), firms' size (Gertler and Gilchrist, 1994) and their holdings of liquid assets (Jeenas, 2019).

Debt composition and monetary transmission

- Share of floating-rate debt and the debt maturity were shown to affect the transmission of MP to firms' investment and stock prices (Ippolito et al., 2017, Gürkaynak et al. 2022, Jungherr et al., 2022).
- Bond/loan composition: higher bond share weakens CMP transmission (Crouzet 2021, Holm-Hadulla and Thurwachter 2021), but the opposite is true when bond market frictions dominate (Darmouni et al. 2020) or when LT rates are affected by the CB (Holm-Hadulla and Thurwachter 2021).

Related Literature

Go Back

Financial frictions and investment

- Cloyne et al. (2023) find that financial frictions account for about one third of the aggregate investment response to monetary policy.
- Firm-level response to MP shock depends on default risk (Ottonello and Winberry, 2020), the firms' age (Cloyne et al., 2023), firms' size (Gertler and Gilchrist, 1994) and their holdings of liquid assets (Jeenas, 2019).

Debt composition and monetary transmission

- Share of floating-rate debt and the debt maturity were shown to affect the transmission of MP to firms' investment and stock prices (Ippolito et al., 2017, Gürkaynak et al. 2022, Jungherr et al., 2022).
- Bond/loan composition: higher bond share weakens CMP transmission (Crouzet 2021, Holm-Hadulla and Thurwachter 2021), but the opposite is true when bond market frictions dominate (Darmouni et al. 2020) or when LT rates are affected by the CB (Holm-Hadulla and Thurwachter 2021).

MP identification

- MP shocks using high-frequency identification: Kuttner (2001), Gerko and Rey (2017), Jarocinski and Karadi (2020), Altavilla et al (2019), ...

Contribution

Go Back

1. We identify a **bond-liquidity channel of MP** and provide evidence on its impact on French firms' investment. In particular, we provide evidence on the impact of the ECB policies on the liquidity of the French bond market and its effect on corporate bond prices.

2. We study the **role of corporate debt structure** in the transmission of both types of monetary policy to investment.

3. We uncover the **relative importance of bond and bank credit supply shocks** induced by CMP and Bond Spread (BSP) shocks

   $\Rightarrow$ novel evidence on the bank lending and liquidity channels of both MP types.

Conventional Monetary Policy shocks

Go back

Variable	Shock	
	Monetary policy (negative co-movement)	CB information (positive co-movement)
m_t, high frequency		
Interest rate	+	+
Stock index	−	+

Source: Jarociński and Karadi (2020)

BSP shocks

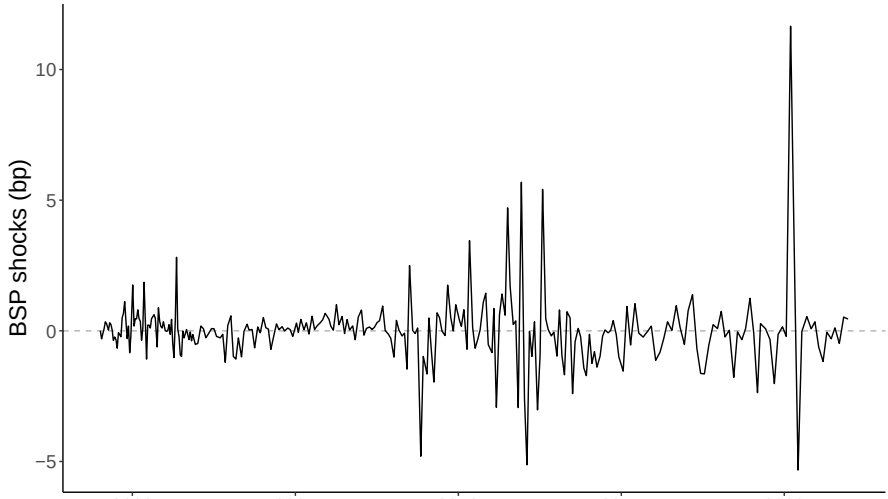
Go back

Motivation

- French and German sovereign bond markets have many similarities
  - Ejsing and Sihvonen, 2009
  - Same currency, similar credit rating, comparable amounts outstanding
- Evidence that movements in spread reflect mostly changes in **liquidity premia**
  - ECB Monthly Bulletin 09/2009
- They also affect French sovereign bond market liquidity

Bond Spread shocks: daily

Go back



Bond Spread shocks: narrative description

Go back

- **Largest BSP shocks and the events that triggered them:**

Date	Shock (bp)	Description
2012/02/09	-5.13	Eligibility rules eased for sovereigns
2012/12/08	5.69	Downplayed possible renewal of bond purchases.
2020/03/12	11.66	CL: "We are not here to close spreads"
2020/06/04	-5.33	PEPP increased, includes corporates

BSP shocks and bid-ask spreads

Go back

Smooth Local Projections (Barnichon and Brownlees, 2019)

We check the link between BSP shocks and liquidity

- Daily data on bid-ask spreads of FR sovereign bonds

BSP shocks and bid-ask spreads

Go back

Smooth Local Projections (Barnichon and Brownlees, 2019)

We check the link between BSP shocks and liquidity

- Daily data on bid-ask spreads of FR sovereign bonds
- Only a limited number of MP shocks to pin down responses
  - Aggregate data: cannot use information from cross-section

BSP shocks and bid-ask spreads

Go back

Smooth Local Projections (Barnichon and Brownlees, 2019)

We check the link between BSP shocks and liquidity

- Daily data on bid-ask spreads of FR sovereign bonds
- Only a limited number of MP shocks to pin down responses
  - Aggregate data: cannot use information from cross-section

Smooth **L**ocal **P**rojections (Barnichon and Brownlees, 2019)

BSP shocks and bid-ask spreads

Go back

Smooth Local Projections (Barnichon and Brownlees, 2019)

We check the link between BSP shocks and liquidity

- Daily data on bid-ask spreads of FR sovereign bonds
- Only a limited number of MP shocks to pin down responses
  - Aggregate data: cannot use information from cross-section

Smooth **L**ocal **P**rojections (Barnichon and Brownlees, 2019)

- Standard LP are heavily parameterized and data intensive
- VAR approaches are efficient but restrictive/biased

BSP shocks and bid-ask spreads

Go back

Smooth Local Projections (Barnichon and Brownlees, 2019)

We check the link between BSP shocks and liquidity

- Daily data on bid-ask spreads of FR sovereign bonds
- Only a limited number of MP shocks to pin down responses
  - Aggregate data: cannot use information from cross-section

Smooth **L**ocal **P**rojections (Barnichon and Brownlees, 2019)

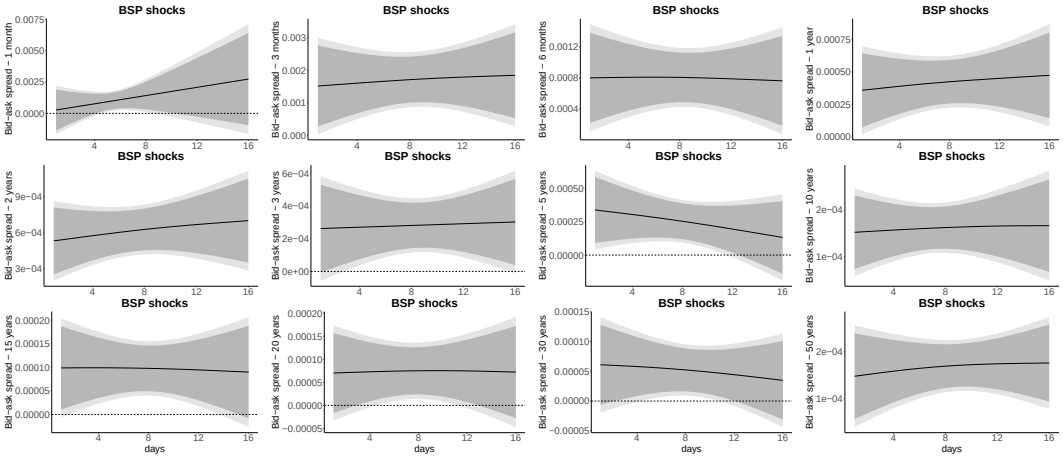
- Standard LP are heavily parameterized and data intensive
- VAR approaches are efficient but restrictive/biased

$\Rightarrow$ S-LP penalize variability w/o ex-ante restricting shape of IRFs

Detail

Bond Spread shocks and bid-ask spreads

Go Back



Smooth Local Projections (Barnichon and Brownlees, 2019)

Back

S-LP: IR estimation methodology based on B-spline smoothing

Smooth Local Projections (Barnichon and Brownlees, 2019)

Back

S-LP: IR estimation methodology based on B-spline smoothing

Shrinkage hyper-parameter λ:

- Pins down bias/variance trade-off of estimator

Smooth Local Projections (Barnichon and Brownlees, 2019)

Back

S-LP: IR estimation methodology based on B-spline smoothing

Shrinkage hyper-parameter λ:

- Pins down bias/variance trade-off of estimator
- $\lambda = 0$: standard LP estimated by least squares
- $\lambda \to \infty$: polynomial distributed lag model (Almon, 1965)

Smooth Local Projections (Barnichon and Brownlees, 2019)

Back

S-LP: IR estimation methodology based on B-spline smoothing

Shrinkage hyper-parameter λ:

- Pins down bias/variance trade-off of estimator
- $\lambda = 0$: standard LP estimated by least squares
- $\lambda \to \infty$: polynomial distributed lag model (Almon, 1965)

Following Barnichon and Brownlees (2019):

- 5-fold cross-validation to select λ
- Pick λ with best mean pseudo-out-of-sample fit

Main hypotheses (1/2) Go back

1. CMP and the bank lending channel

- CMP tightening ⇒ firms switch from bank loans to bonds
- In line with the bank lending channel
- Kashyap et al. (1992), Becker and Ivashina (2014)

▶ Rate hike ⇒ investment should fall **more** for NFCs with higher shares of **bank** debt

Main hypotheses (2/2)

Go back

2. Quantitative Easing and the bond liquidity channel

- QE reduces risk premia on debt securities $\Rightarrow$ stimulates corporate bond issuance
- Altavilla & Giannone (2017), Lhuissier & Szczerbowicz (2021), Grosse-Rueschkamp et al. (2019)

▶ BSP rise $\Rightarrow$ investment should fall **more** for NFCs with higher shares of **bond** debt

MP and Investment: average effects

Go back

The average effects of the ECB MP on French firms' investment are evaluated using panel local projections (Jordá, 2005).

$$\Delta I_{i,t+h} = \beta^h S_{i,t} + \Psi^h Z_{t-1} + \sum_{l=1}^{L} \Gamma_l^h X_{i,t-\ell} + \mu_i^h + \epsilon_{i,t+h} \tag{1}$$

$\mathbf{\Delta I_{i,t+h}} = I_{i,t+h} - I_{i,t-1}$: h-year forward difference in the net investment rate

$\mathbf{S_{i,t}}$: vector of CMP and BSP shocks

MP and Investment: average effects

Go back

The average effects of the ECB MP on French firms' investment are evaluated using panel local projections (Jordá, 2005).

$$\Delta I_{i,t+h} = \beta^h S_{i,t} + \Psi^h Z_{t-1} + \sum_{l=1}^{L} \Gamma_l^h X_{i,t-\ell} + \mu_i^h + \epsilon_{i,t+h} \quad (1)$$

$\mathbf{\Delta I_{i,t+h}} = I_{i,t+h} - I_{i,t-1}$: h-year forward difference in the net investment rate

$\mathbf{S_{i,t}}$: vector of CMP and BSP shocks

$\mathbf{Z_{t-1}}$: vector of lagged aggregate controls: French output gap, French inflation, VIX, 10y French sovereign rate, 3m interbank rate

MP and Investment: average effects

Go back

The average effects of the ECB MP on French firms' investment are evaluated using panel local projections (Jordá, 2005).

$$\Delta I_{i,t+h} = \beta^h S_{i,t} + \Psi^h Z_{t-1} + \sum_{l=1}^{L} \Gamma_l^h X_{i,t-\ell} + \mu_i^h + \epsilon_{i,t+h} \tag{1}$$

$\mathbf{\Delta I_{i,t+h}} = I_{i,t+h} - I_{i,t-1}$: h-year forward difference in the net investment rate

$\mathbf{S_{i,t}}$: vector of CMP and BSP shocks

$\mathbf{Z_{t-1}}$: vector of lagged aggregate controls: French output gap, French inflation, VIX, 10y French sovereign rate, 3m interbank rate

$\mathbf{X_{i,t-\ell}}$: vector of lagged firm-specific controls: leverage, total assets, cash flows to total assets, bond dummy

MP and Investment: average effects

Go back

The average effects of the ECB MP on French firms' investment are evaluated using panel local projections (Jordá, 2005).

$$\Delta I_{i,t+h} = \beta^h S_{i,t} + \Psi^h Z_{t-1} + \sum_{l=1}^{L} \Gamma_l^h X_{i,t-\ell} + \mu_i^h + \epsilon_{i,t+h} \quad (1)$$

$\mathbf{\Delta I_{i,t+h}} = I_{i,t+h} - I_{i,t-1}$: h-year forward difference in the net investment rate

$\mathbf{S_{i,t}}$: vector of CMP and BSP shocks

$\mathbf{Z_{t-1}}$: vector of lagged aggregate controls: French output gap, French inflation, VIX, 10y French sovereign rate, 3m interbank rate

$\mathbf{X_{i,t-\ell}}$: vector of lagged firm-specific controls: leverage, total assets, cash flows to total assets, bond dummy

$\mu_\mathbf{i}$: firm fixed effects

MP and Investment: average effects

Go back

The average effects of the ECB MP on French firms' investment are evaluated using panel local projections (Jordá, 2005).

$$\Delta I_{i,t+h} = \beta^h S_{i,t} + \Psi^h Z_{t-1} + \sum_{l=1}^{L} \Gamma_l^h X_{i,t-\ell} + \mu_i^h + \epsilon_{i,t+h} \tag{1}$$

$\mathbf{\Delta I_{i,t+h}} = I_{i,t+h} - I_{i,t-1}$: h-year forward difference in the net investment rate

$\mathbf{S_{i,t}}$: vector of CMP and BSP shocks

$\mathbf{Z_{t-1}}$: vector of lagged aggregate controls: French output gap, French inflation, VIX, 10y French sovereign rate, 3m interbank rate

$\mathbf{X_{i,t-\ell}}$: vector of lagged firm-specific controls: leverage, total assets, cash flows to total assets, bond dummy

μ_i: firm fixed effects

$\mathbf{L = 3}$ for all results shown today

MP and Investment: the role of debt structure

Go back

- Investigate cross-sectional differences wrt corporate debt structure
  - Interact MP shocks with the firm's share of bond debt (bond ratio)

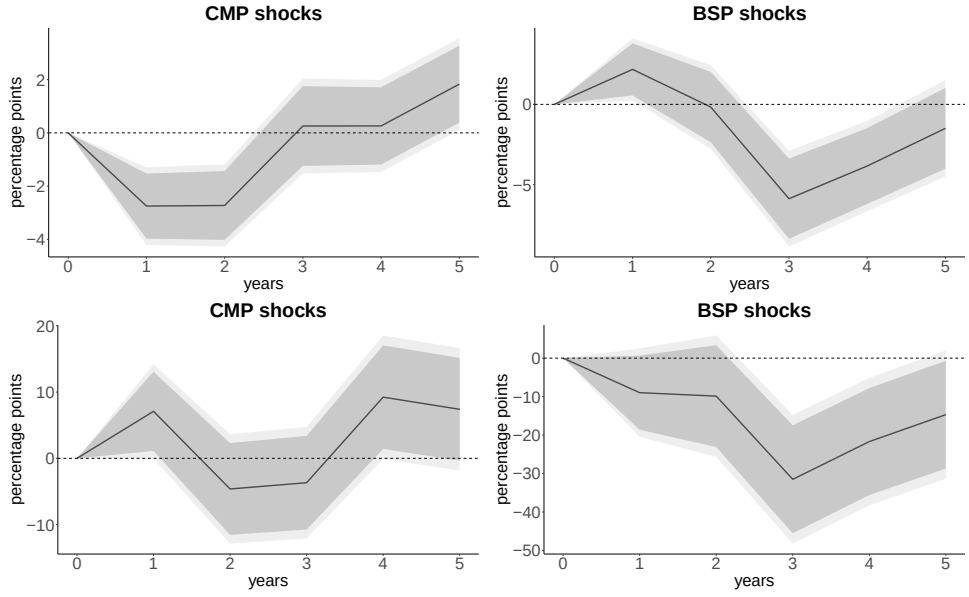
$$\Delta I_{i,t+h} = \alpha^h B_{i,t-1} \times S_{i,t} + \beta^h Mat_{i,t-1} \times S_{i,t} + \sum_{\ell=1}^{3} \Gamma_\ell^h X_{i,t-\ell} + \mu_i^h + \theta_{s,t}^h + \epsilon_{i,t+h} \quad (2)$$

- $B_{i,t-1}$: lagged bond ratio
- $Mat_{i,t-1}$: lagged maturity ratio
- $\theta_{s,t}^h$: sector-time fixed effects

$\Rightarrow$ Positive (negative) α^h means firms with higher bond ratios are less (more) responsive

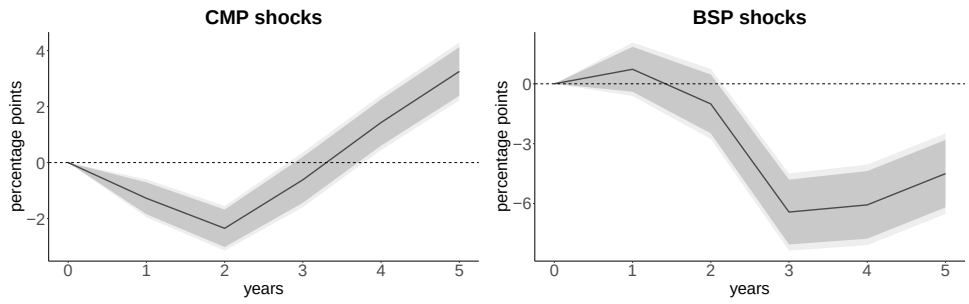
Simple calendar year aggregation

Go Back



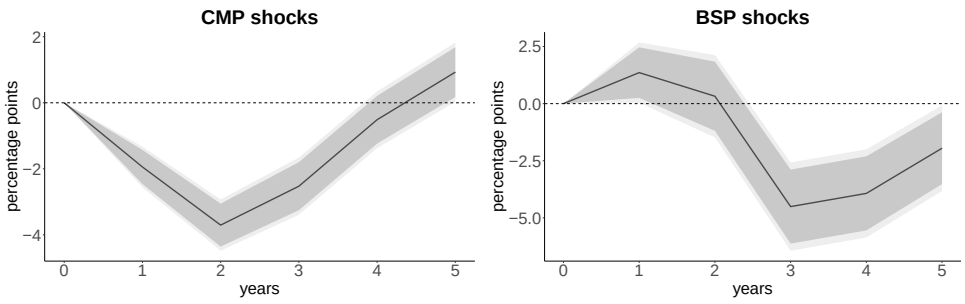
Only output gap and inflation as controls

Go Back



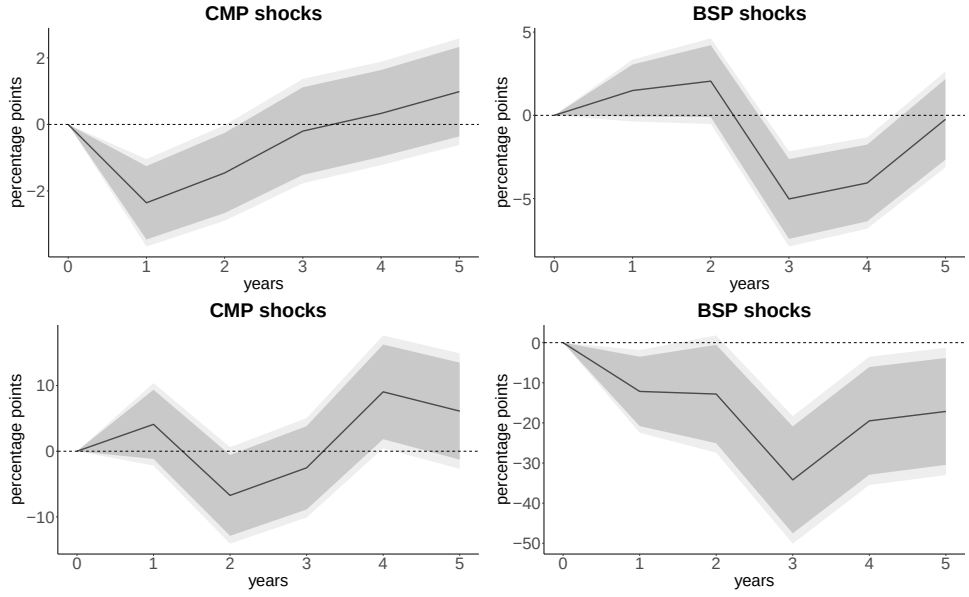
No aggregate controls

Go Back



Excluding maturity interaction

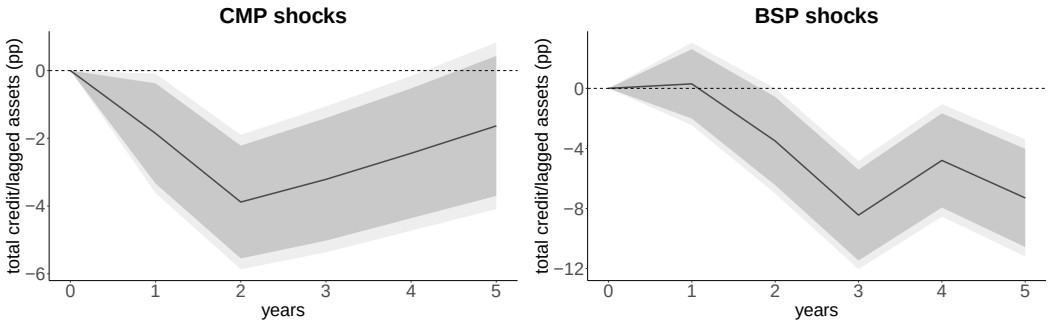
Go Back



Total Credit

Go Back

Average response

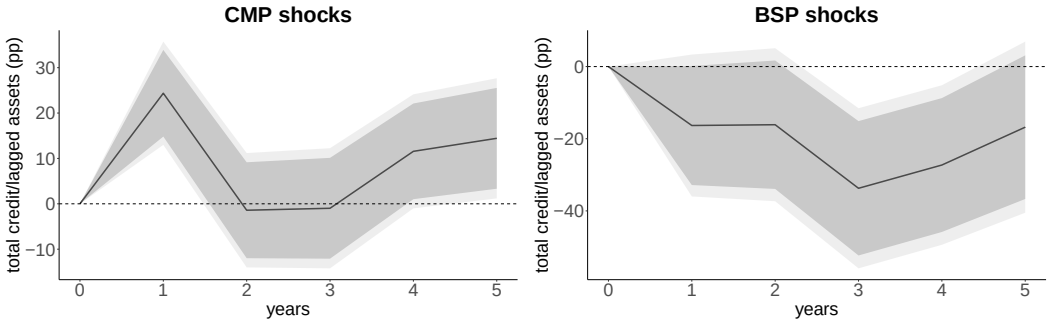


- Total credit falls across firms for all shocks

Total Credit

Go Back

Heterogeneous response

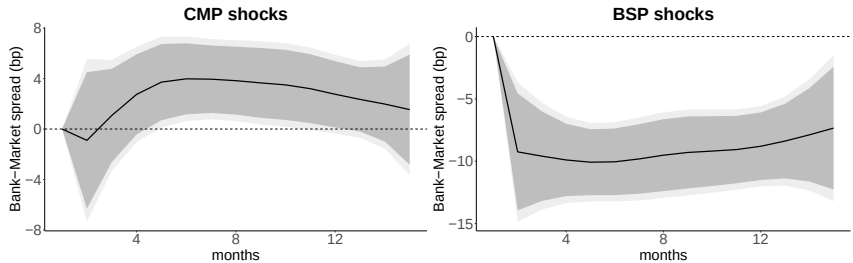


- CMP has a stronger impact on firms that are more bank-based
- BSP shocks have a stronger impact on more market-based firms

Relative Cost

Go Back

Bank-market spreads i.e. rate of bank loans compared with average yield on corporate bonds

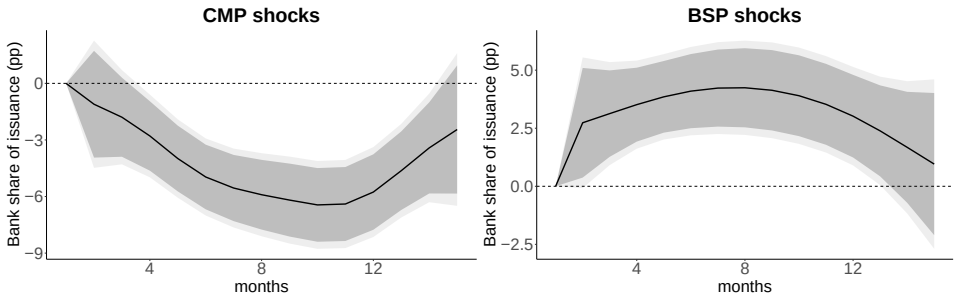


- CMP has stronger pass-through to bank loan rates (Schnabel, 2021),
- but slower transmission after shock (Lane, 2022)
- BSP shock $\Rightarrow$ market rates rise more than bank rates

Debt Issuance

Bank share of debt issuance

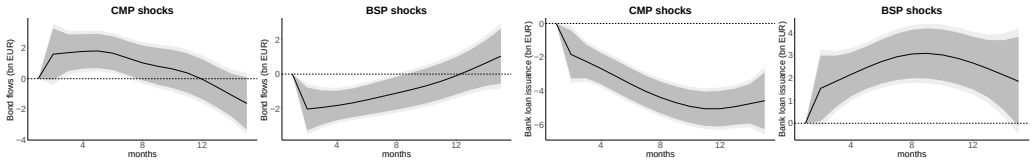
Go Back



- CMP shock ⇒ share of bank debt in new issuance falls
- BSP shock ⇒ share of bank debt in new issuance rises

Bond and loan issuance

Go Back



- CMP shock $\Rightarrow$ bond flows increase
- BSP shock $\Rightarrow$ bank loans increase
- Some degree of **substitutability**,
- ...but not sufficient to stop the contractionary effects on aggregate investment

IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

EU Regulation on combating late payment in commercial transactions[1]

Markus Federau and Bernd Schwendinger,
Vorarlberg University of Applied Sciences

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

EU Regulation on combating late payment in commercial transactions

Working capital impact assessment for European firms

Dr. Markus Federau
Senior Lecturer, Vorarlberg University of Applied Sciences
Hochschulstr. 1, 6850 Dornbirn, Austria
markus.federau@fhv.at

Bernd Schwendinger, Ph.D.
Program Director Master of Business, Vorarlberg University of Applied Sciences
Hochschulstr. 1, 6850 Dornbirn, Austria
bernd.schwendinger@fhv.at

Abstract

This paper examines the potential impact of the EU's late payment regulations on the working capital of European firms, with a particular focus on the differences between companies of different sizes. The study employs data from the ERICA and Osiris databases to examine pivotal metrics, such as days sales outstanding (DSO) and days payables outstanding (DPO), with the objective of evaluating the impact of payment terms on corporate liquidity and financing. The findings indicate that, despite a reduction in overall collection periods over recent years, payment terms remain substantially above the target of 30 days. Furthermore, larger companies tend to benefit from shorter collection periods and more favourable financing conditions through trade payables and therefore should be more negatively impacted by the regulation.

Keywords: payment terms, working capital, financing, regulation, ERICA

JEL classification: G31 (Capital Budgeting; Fixed Investment and Inventory Studies; Capacity), G38 (Government Policy and Regulation)

Contents

1. Introduction 3

2. Research Question & Methodology 3

3. ERICA sample 4

   3.1 Sample description 4

   3.2 Key figures 5

   3.3 Longitudinal analysis 6

   3.4 Size analysis 7

   3.5 Summary & Limitations 8

4. Osiris sample 8

   4.1 Sample description 9

   4.2 Key figures 9

   4.3 Longitudinal analysis 10

   4.4 Size analysis 13

   4.5 Summary & Limitations 14

5. Conclusion 14

References 16

1. Introduction

On 12 September 2023 the European Commission presented a proposal for a new regulation "on combating late payment in commercial transactions." The proposal represents a revision of the existing EU Late Payment Directive and is intended to replace it. The revision included amendments, particularly the introduction of a uniform EU-wide payment term upper limit of 30 days for business-to-business and business-to-government transactions ("B2B" and "B2G"), without exception. In contrast, the current version of the EU Late Payment Directive sets an upper limit of 60 days for payment terms in B2B transactions. However, this may be adjusted upwards by the contracting parties, provided this is not 'grossly unfair' for the creditor (Directive 2011/7/EU, Art. 3, Nr. 5). From the perspective of the European Commission the current directive lacked sufficient preventive and deterrent effect, and the existing enforcement mechanisms were inadequate. Small suppliers/creditors would face asymmetric bargaining power towards large customers/debtors (European Commission, 2023).

Given that the issue of late payment and long payment terms affects all EU member states, a uniform EU-wide regulation is being sought. Accordingly, the current directive shall be replaced by a regulation which would be applicable across the Union without the necessity of an additional adoption by the national parliaments. Business relationships between companies and consumers (B2C) are not subject to regulation. The proposal was adopted by the European Parliament in the first reading on 23 April 2024 (European Parliament, 2024a).

The original proposal put forth by the Commission underwent significant alterations. Firstly, while a basic payment deadline upper limit of 30 days is applicable in B2B/B2G transactions, it can be extended to 60 days in B2B, provided that this is explicitly stipulated in the contract. For certain retail products, such as slow-moving and seasonal goods, an upper limit of even 120 days may apply. New is also that interest on arrears accrues automatically. It is not possible for public bodies or large corporations, in their capacity as debtors, to contractually exclude default interest. Furthermore, a fixed-rate compensation is applied for each late transaction, with the amount varying between 50 and 150 EUR, depending on the volume. A contractual prohibition or restriction of the right to assign receivables for the use of financing services (e.g. factoring) is generally deemed invalid. The EU member states have to designate bodies responsible for enforcing the law, conducting investigations and authorised to impose administrative sanctions, as well as publishing the names of offenders. A central monitoring centre is to be established at the EU level. The implementation of the revised regulation is contingent upon the European Council's approval, which remains pending (European Parliament, 2024b).

2. Research Question & Methodology

The timing of cash receipts and disbursements affects a company's working capital management, which in turn impacts its internal financing capacity, profitability, and ultimately, shareholder value (Högerle et al., 2020; Le, 2019; Singh et al., 2017). A recent study by Federau (2024) for the German DAX indicates that the impact of the

regulation of payment terms could be significant, particularly for large corporations that rely heavily on the financing from suppliers.

The objective of this study is to expand the scope of analysis and examine the evolution of net working capital in Europe. This will enable an assessment of the current status and an evaluation of the potential impact of the proposed new regulations on payment terms. Additionally, the study seeks to address the hypothesis put forth by the European Commission that larger companies rely more on longer payment terms than smaller companies.

The analysis focuses on the key figures "Days Sales Outstanding" (DSO) for trade receivables and "Days Payables Outstanding" (DPO) for trade payables. The sample data, comprising approximately 1,000 European IFRS statements of accounts, has been derived from the European ERICA (European Records of IFRS Consolidated Accounts) database, the characteristics of which are described in greater detail below. In addition to the interpretation of the results, the limitations of the approach and the ERICA data are described. Recommendations are provided on how the published data can be enhanced for an increased usefulness for external parties. Finally, the results are compared with the analysis of a panel of approximately 1,000 European companies from the Osiris database hosted by Moody's.

3. ERICA sample

The ERICA database was created by the ERICA working group of the ECCBSO (European Committee of Central Balance Sheet Data Offices) and contains information on the annual accounts of consolidated non-financial listed groups from nine participating countries: Austria, Belgium, France, Germany, Greece, Italy, Portugal, Spain and Turkey (as the only country not part of the EU).

The data is not merely collected for the purpose of statistical investigations but also for risk assessment analysis. The comprehensive ERICA data at the company level is accessible solely to the ECCBSO members, ERICA working group members, and analysts affiliated with the European Central Bank (ECB) and the contributing national central banks (European Committee of Central Balance Sheet Data Offices (ECCBSO), 2024). Consequently, since 2019, the general public has only had access to aggregated data, which can be obtained via a web portal hosted by the French Central Bank (Banque de France, n.d.).

3.1 Sample description

The aggregated data from the ERICA database is available for download in a single Excel file from the aforementioned source. The data is categorised according to four distinct criteria: 1.) aggregation by size, 2.) aggregation by detailed sector, 3.) aggregation by country and sector and 4.) total aggregation (European Committee of Central Balance Sheet Data Offices (ECCBSO), 2024). The size classes are based on the companies' revenues with the following thresholds: less than €250 million (small), between €250 million and below €1,500 million (medium) and from €1,500 million and higher (large). The detailed sector aggregation employs a proprietary classification system based on the NACE classification of the European Union

resulting in a total of 14 so-called WGRA sectors such as "food products", "chemicals", "metals, electronic and electrical equipment".[1] The aggregation by sector (so-called "ERICA sectors") employs only the four categories of "construction", "energy", "industry" and "services". The country refers to the country of incorporation of the reporting entity and comprises the nine participating nations mentioned above. The database comprises the years 2005 to the most current year of 2022 (as of September 2024). An annual update shall be published around February (European Committee of Central Balance Sheet Data Offices (ECCBSO), 2023).

The composition of the database is subject to change; consequently, no aggregated consistent panel data is available. Table 1 provides a detailed overview of the number of companies included in the aggregated numbers for the figures of DSO and DPO over the years of analysis.

Number of companies in ERICA database included in DSO/DPO aggregation

Table 1

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
DSO	1,070	1,187	1,072	1,071	1,085	1,075	1,107	1,009	983	975	962	1,053	829
DPO	1,070	1,184	1,072	1,073	1,086	1,076	1,107	1,010	982	972	963	1,054	831

The number for the calculation of the quartiles/averages of the DSOs/DPOs in case of the total sample ranges between 829/831 (year 2022) and 1,187/1,184 (year 2011). The reduction in the sample size in 2022 is likely attributable to the missing Belgic companies from the dataset. Over time, there has been a minimal discrepancy between the DSO and DPO sample sizes, with a range of -0.2% to +0.3%.

A review of the sub-samples by size for the most recent year 2022, reveals a relatively even distribution with approximately one-third of the companies falling into each of the categories "small", "medium" and "large" (see Table 2). Additionally, the discrepancy number of companies for the calculation of DSO/DPO figures is nearly identical.

Number of companies in ERICA database 2022 included in DSO/DPO aggregation by company size

Table 2

	Small	Medium	Large	Total
DSO	282	260	287	829
DPO	282	261	288	831

3.2 Key figures

To analyse the turnover time of trade receivables and payables and to ascertain their relevance with regard to the requirement of working capital, we employ the use of DSOs/DPOs as proxies. As the data necessary for calculating these key figures is not sufficiently available in the public ERICA database, we must rely on the pre-calculated

[1] For more information on the NACE classification, refer to Eurostat (2008).

ratios from, which are designated "r19" for the DSO and "r20a" for the DPO (European Committee of Central Balance Sheet Data Offices (ECCBSO), 2023). Unfortunately, the publicly available documentation does not provide any information regarding the calculation of these ratios. Upon requesting such information from the ECCBSO secretariat, the following response was received:

$$\text{DSO} = \frac{\text{Trade Receivables}}{\text{Daily Average Revenues}} \quad \text{and} \quad \text{DPO} = \frac{\text{Trade Payables}}{\text{Daily Average Revenues}}$$

While the calculated DSO can be regarded as a proxy for the average time between invoicing and payment receipt, this cannot be assumed for the trade payables/DPO side. This is because trade payables are not the result of revenues in the denominator; rather they are driven by purchasing, which could be proxied by material expenses and expenses for other external products and services. However, as previously outlined, such calculations cannot be performed based on the aggregated ERICA data.

Moreover, an additional examination of working capital ratios, such as the ratios of trade receivables or trade payables as a percentage of the balance sheet total, is not possible due to the varying size and composition of the ERICA sample for the different input parameters. Consequently, we must rely on the quartile analysis of the pre-defined DSOs/DPOs.

It is also important to note that the data analysed here pertains to IFRS group financials. Accordingly, the figures do not solely reflect the customer/supplier relationships within Europe; rather, they encompass a global perspective. Therefore, the introduction of a further regulatory measure in Europe would only have a partial impact on the figures presented below.

3.3 Longitudinal analysis

On the trade receivables side, the median DSO decreased between 2010 and 2022 by 8.7 days (see Table 3).

DSOs [days] of ERICA database

Table 3

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Q1	45.4	42.5	41.4	42.1	42.7	42.5	42.7	43.0	40.2	40.6	41.5	41.0	35.6
Q2	68.3	66.3	63.1	63.8	66.6	64.4	66.0	64.7	64.4	62.2	63.9	64.6	59.6
Q3	103.3	104.0	95.8	93.2	95.9	96.9	100.2	98.0	95.2	93.2	95.0	96.8	84.5
Mean	83.3	80.9	77.9	78.2	81.4	88.6	2,525.9	79.7	780.5	85.0	91.5	81.7	68.4
IQR	57.9	61.4	54.5	51.1	53.3	54.5	57.4	55.0	55.0	52.7	53.5	55.8	49.0

Q1...Q3: quartile 1...3 / IQR: interquartile range

The statistical significance of this development cannot be tested based on the aggregated data. A decrease can also be observed for the development of the first

and third quartiles. Although this is a positive development and it appears that the sample companies are receiving their funds more promptly than in the past, fifty percent of the companies wait for their payments 60 days or longer. Moreover, the first quartile is still well above the envisaged policy target of a maximum payment period of 30 days.

An inverse trend can be observed for trade payables (see Table 4). The median DPO increased by 3.7 days between 2010 and 2022. However, between 2011 and 2013, DPOs decreased, specifically the median by -1.9 days and the third quartile by -4.3 days. Overall, the DPO distribution seems to be more influenced by outliers, as the mean DPO deviates noticeably more from the median compared to the DSO figures.

DPOs [days] of ERICA database

Table 4

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Q1	31.9	30.0	28.7	29.5	30.9	30.0	31.2	32.1	34.2	33.1	34.9	38.3	32.6
Q2	48.4	47.9	45.2	46.0	47.8	47.3	47.2	50.4	52.4	52.1	54.3	60.2	52.1
Q3	74.9	75.7	70.5	71.4	73.6	72.8	74.1	76.2	79.8	77.8	84.8	92.0	79.5
Mean	111.7	73.0	63.0	65.5	77.2	1,048	73,692	95.9	99,245	77,951	452.7	100.7	82.0
IQR	43.0	45.8	41.8	41.9	42.7	42.8	43.0	44.1	45.5	44.7	50.0	53.8	46.9

Q1...Q3: quartile 1...3 / IQR: interquartile range

Although the median DPO is at around 52 days in 2022, it can be assumed that the actual payment period for payables in the sample is much higher. This is due to method of calculating DPOs in ERICA, which is based on revenues and not, for example, on material expenses as described above.

3.4 Size analysis

A key argument in favour of stricter regulation of payment terms is that smaller companies have to accept significantly longer payment terms due to their lower bargaining power (European Commission, 2023). As explained in the description of the sample, we have to rely on the thresholds defined in the ERICA database, i.e. €250 million and €1,500 million, to distinguish between small, medium and large groups. The results based on the year 2022 are shown in Table 5.

There is a notable difference between the quartiles of the DSOs of small and large companies, which amounts to 21.8 days for the median, resulting in a more than 30% shorter collection period for large companies. On the one hand, this could be due to a greater bargaining power of large corporations vis-à-vis their customers compared to smaller corporations, or to a more effective receivables management, e.g. through more consequent dunning procedures or the utilisation of factoring (Brealey et al., 2022). On the other hand, it could also be the result of different customer segments, e.g. the split between B2B and B2C customers of large vs. small companies and different payment practices in the different segments.

DSOs and DPOs 2022 of ERICA database by size

Table 5

DSOs [days]				DPOs [days]			
	Small	**Medium**	**Large**		**Small**	**Medium**	**Large**
Q1	43.7	37.4	32.4	Q1	31.4	31.3	34.8
Q2	71.2	56.8	49.4	Q2	56.4	48.3	53.2
Q3	103.9	80.1	72.8	Q3	86.7	76.3	75.9
Mean	85.6	63.3	56.3	Mean	122.4	59.4	62.9
IQR	60.1	42.8	40.5	IQR	55.3	45.0	41.1

Q1...Q3: quartile 1...3 / IQR: interquartile range

On the DPO side there is only a small difference in medians, i.e. small companies have a median KPI that is 3.2 days longer than large companies. This does not indicate that smaller companies have less bargaining power with their suppliers compared to their larger peers.

3.5 Summary & Limitations

Overall, the ERICA sample provides a broad basis for analysis with around 1,000 data points per year. The data for the receivables (DSO) of the companies show a level of collection periods well above the targeted 30 days. A regulation could therefore be expected to have an impact at least on the part of B2B sales in Europe. A breakdown of the sample by size reveals a large difference in DSOs between large and small companies , which may be due to different factors and should be analysed in further research.

The main limitations of the analysis that need to be taken into account are the following: The analysed ERICA data is only available in aggregated form. Due to the lack of individual cases, a different level of aggregation as offered by the ECCBSO (e.g. by different size classes or different industry clusters) is not possible. Furthermore, the analysis can only be based on descriptive statistics. Hypothesis testing is not possible due to missing distribution parameters such as standard deviation or rank sums of the sub-samples. The inability to construct stable panels limits the interpretation of time series data. Timely analysis of actual data is also constrained by the relatively long update cycles. The transparency of the calculation of ratios, such as DSO/DPO, should also be improved. For the analysis and benchmarking of balance sheet data, we propose to include not only absolute figures, but also relative values expressed as percentages of the balance sheet total. Similarly, the absolute figures of the income statement could be extended to include ratios for each line item expressed as a percentage of revenues.

4. Osiris sample

To overcome some of the limitations and validate the findings, we conduct a separate analysis based on financial data from the Osiris database. Osiris was previously part

of Bureau van Dijk, which was acquired by Moody's Analytics in 2017. It is sometimes referred to as Orbis Global. It is a comprehensive resource covering balance sheet data of publicly listed companies worldwide (Moody's Analytics, n.d.). The advantage over the aggregated ERICA data is that information is available at the individual company level. This means that the data can be filtered individually, outliers can be assessed, and stable panels can be analysed over longer periods. In addition, KPIs can be defined as needed. They can be calculated at the company level and aggregated and tested as required.

4.1 Sample description

The sample was selected as follows: From the entire database of publicly listed companies (69,327), only those based in the European Union were filtered (5,759 remaining). In a next step financial companies were excluded based on their GICS codes (4,076 remaining). To ensure comparability, companies not applying IFRS were also eliminated (2,913 remaining).

To ensure comparability, we included the same period (2010 to 2022) in our analysis, although more recent figures are available in the database. We used a panel approach to ensure to track the development of the receivables and payables of the same set of companies over time. For the analysis of DSOs and trade receivables we included only those companies which reported meaningful receivables and revenues of over one million Euro in all years (1,011 remaining). For DPOs and trade payables, we applied the same procedure to the values of trade payables and costs of goods sold (1,081 remaining).

In order to compare groups of companies of different sizes, we also analysed the single year 2022. We used the same filter criteria as above, i.e. only companies with receivables/revenues or trade payables/costs of goods sold of more than one million Euro were included. This resulted in subsamples of 2,171 companies for the assessment of DSOs/receivables and 2,170 companies for DPOs/trade payables. For reasons of comparability, we applied the same size thresholds for the categorisation of small, medium and large groups as defined in the ERICA database.

4.2 Key figures

We base our analysis on a total of four different indicators. First, we analyse the receivables side by calculating the DSOs, sometimes also called "average collection period", based on the following formula (Sagner, 2014):

$$\text{DSO} = \frac{\text{Accounts Receivable (current)}}{\text{Total Revenues}} \times 360 \text{ days}$$

The analysis cannot be performed for trade receivables only, as the Osiris format does not differentiate current receivables further. To measure the amount of capital tied in receivables we also calculate the receivables in relation to total assets.

$$\text{AR in B/S total} = \frac{\text{Accounts Receivable (current)}}{\text{Total Assets}}$$

A similar approach is used for payables to suppliers. We calculate the DPOs as follows:

$$\text{DPO} = \frac{\text{Trade Payables}}{\text{Costs of Goods Sold (COGS)}} \times 360 \text{ days}$$

Unlike receivables, there is a separate item in the Osiris database for trade payables. We use the costs of goods sold in the denominator as this results in a closer proxy for payment periods towards suppliers than using turnover. However, it should be noted that ideally only material expenses and other third-party expenses should be included if such data are available (Federau, 2024).

As with receivables, we measure the relative amount of financing provided by suppliers by calculating trade payables in relation to total liabilities and equity:

$$\text{AP in B/S total} = \frac{\text{Trade Payables}}{\text{Total Liabilities \& Equity}}$$

The higher such amount, the more funding is provided by a company's suppliers and the less it needs to be covered by other, more costly financing forms such as financial debt or equity.

4.3 Longitudinal analysis

Looking at the evolution of the mean DSOs of the sample (see Table 6), the amounts decreased from 2010 to 2022 substantially by 12.0 days. However, due to the presence of substantial outliers, we focus on the interpretation of robust quartiles instead.

DSOs [days] of Osiris panel

Table 6

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Q1	43.2	41.7	40.8	40.8	40.6	40.0	41.4	40.5	39.9	37.9	38.3	39.0	35.0
Q2	62.2	60.9	59.3	59.1	59.0	56.5	59.9	59.6	58.9	56.6	55.1	56.4	52.7
Q3	89.8	88.1	85.0	82.8	84.1	80.8	83.6	81.1	81.7	77.7	77.3	76.9	74.7
Mean	74.2	73.2	70.4	69.4	70.4	68.0	70.3	67.9	66.9	63.2	64.1	65.8	62.2
IQR	46.6	46.4	44.1	42.0	43.4	40.8	42.3	40.6	41.8	39.7	39.0	37.8	39.6

Q1…Q3: quartile 1…3 / IQR: interquartile range

As such median decreased as well significantly ($p < 0.001$[2]) by 9.5 days from 62.2 to 52.7 days, representing an annual average (CAGR) of -1.4%. This figure is still well above 30 days. Interestingly, the largest decrease in the third quartile occurred

[2] Two-sided Mann-Whitney-Test, paired.

between 2011 and 2013, the period in which the existing EU Directive on late payment was agreed and came into force. This may indicate that the Directive had indeed an impact by reducing very long payment periods.

A reduction can also be observed when comparing the share of the accounts receivables in relation to the balance sheet total (see Table 7), where the median has decreased significantly by 3.2 percentage points ($p < 0.001$[3]). Overall, the European companies in the sample seem to have been able to manage their receivables efficiently over the period observed, freeing up working capital from receivables.

Accounts receivables of B/S total

Table 7

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Q1	10.2%	10.2%	10.0%	9.7%	9.3%	9.0%	9.3%	9.4%	9.1%	7.7%	7.1%	7.7%	7.7%
Q2	16.3%	16.4%	15.7%	15.3%	15.3%	14.9%	15.0%	15.2%	14.7%	13.4%	12.0%	12.5%	13.1%
Q3	24.1%	24.0%	23.1%	23.0%	22.9%	22.4%	22.5%	22.7%	21.9%	19.8%	18.1%	18.3%	19.2%
Mean	18.5%	18.6%	18.1%	17.7%	17.5%	17.2%	17.2%	17.2%	16.8%	15.2%	13.9%	14.3%	14.8%
IQR	13.9%	13.8%	13.1%	13.3%	13.6%	13.4%	13.2%	13.3%	12.8%	12.1%	10.9%	10.6%	11.5%

B/S: balance sheet / Q1…Q3: quartile 1…3 / IQR: interquartile range

The analysis of the payables side, measured by DPOs, shows a somewhat different picture (see Table 8). Over the analysis period median DPOs increased slightly, but still significantly, by 1.1 days to 75.3 days in 2022 ($p < 0.001$[4]).

DPOs [days] of Osiris panel

Table 8

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Q1	49.0	46.4	44.2	46.0	48.3	44.7	47.5	48.8	49.5	48.3	49.5	53.9	48.7
Q2	74.2	71.2	69.5	68.5	70.6	71.6	74.0	75.7	77.1	74.5	76.7	85.9	75.3
Q3	116.2	115.4	112.2	111.4	115.6	118.0	127.2	126.8	133.6	130.1	135.3	139.1	123.4
Mean	106.9	108.8	109.8	115.1	110.8	112.3	155.9	181.7	143.3	168.3	147.5	166.2	143.7
IQR	67.2	69.0	67.9	65.5	67.4	73.3	79.7	78.0	84.1	81.8	85.8	85.2	74.7

Q1…Q3: quartile 1…3 / IQR: interquartile range

The averages increased dramatically to 144 days in 2022, driven by a steep increase in the fourth quartile as can also be seen in the boxplots (see Figure 1).[5] This means that at least the 25% companies with longest DPOs increased the average payment periods towards their suppliers sharply.

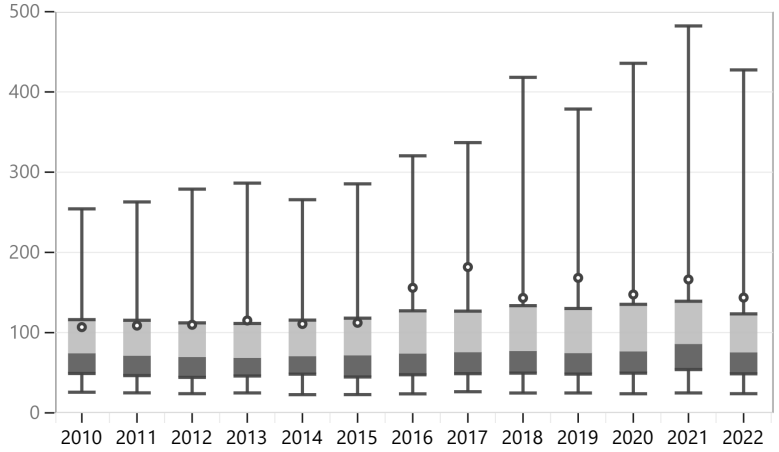
[3] Two-sided Mann-Whitney-Test, paired.

[4] Two-sided Mann-Whitney-Test, paired.

[5] We are setting the whiskers at 0.05 and 0.95 quantiles to give an overview of the overall distribution of the data without inclusion of the most severe outliers.

Boxplot DPOs [days] of Osiris panel

Figure 1



Whiskers representing the 0.05 and 0.95 quantiles

The DPO figures can only be interpreted relatively over time. This is because trade payables are set in relation to COGS. However, under IFRS, the COGS do not only include material expenses from acquired goods but also expenses related to a company's value creation, such as labour costs or depreciation, which are not incurring trade payables (IAS 2: Inventories, 2003). On the other hand, trade payables may result as well from expenses in other functional areas such as selling or general administration. Therefore, the discussed DPO figures give an indication of the trend but cannot be used as an accurate estimate of the actual payment periods or payment terms to suppliers.

The evolution of the trade payables as a percentage of the balance sheet total is rather stable, with a slight decrease in the median from 9.5% to 9.3% (see Table 9).

Trade payables of B/S total

Table 9

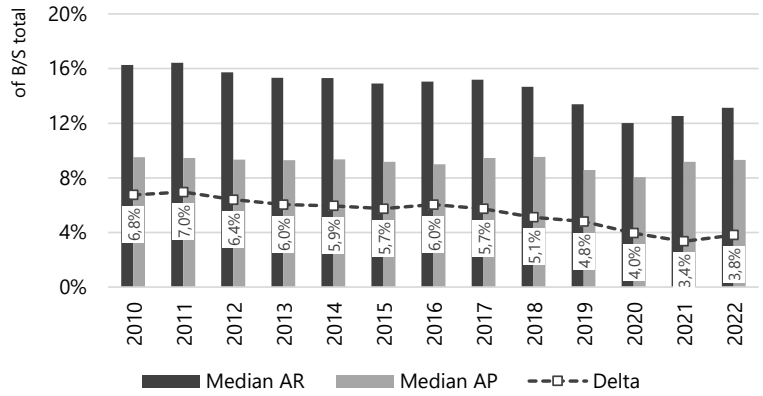
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Q1	5.8%	5.8%	5.7%	5.5%	5.4%	5.2%	5.3%	5.5%	5.5%	4.9%	4.4%	5.0%	5.3%
Q2	9.5%	9.5%	9.3%	9.3%	9.4%	9.2%	9.0%	9.4%	9.5%	8.6%	8.1%	9.2%	9.3%
Q3	15.6%	15.5%	14.9%	15.3%	15.2%	15.2%	15.1%	15.2%	15.7%	14.1%	13.1%	14.8%	14.7%
Mean	12.1%	12.1%	11.9%	11.9%	11.8%	11.7%	11.8%	11.8%	11.9%	10.9%	10.3%	11.3%	11.6%
IQR	9.8%	9.8%	9.3%	9.7%	9.9%	10.0%	9.8%	9.8%	10.2%	9.1%	8.7%	9.8%	9.5%

B/S: balance sheet / Q1…Q3: quartile 1…3 / IQR: interquartile range

Figure 2 shows the comparison between receivables and trade payables in relation to the balance sheet total over time. The aforementioned decline in relative receivables and the rather stable relative trade payables since 2010 imply a decreasing gap between working capital tied up in receivables and supplier financing through trade payables. This gap narrowed from 6.8% in 2010 to only 3.8% in 2022.

Medians of AR and AP in percent of B/S

Figure 2



4.4 Size analysis

Like the ERICA analysis, the next step was to analyse the data clustered by size. To ensure comparability, we applied the same thresholds as in the ERICA database, i.e. revenues less than €250 million (small), between €250 million and below €1,500 million (medium) and from €1,500 million and higher (large). The results for the year 2022 are shown in Table 10.

DSOs/receivables and DPOs/payables 2022 of Osiris panel by size

Table 10

DSOs [days]

	Small	Medium	Large
Q1	41.6	30.2	29.6
Q2	63.6	49.9	45.1
Q3	93.1	68.5	65.1
Mean	103.4	52.8	51.1
IQR	51.4	38.3	35.5

DPOs [days]

	Small	Medium	Large
Q1	52.6	43.1	50.9
Q2	86.1	70.6	80.2
Q3	157.3	122.7	130.9
Mean	190.7	253.9	164.6
IQR	104.8	79.6	80.0

Accounts receivables of B/S total

	Small	Medium	Large
Q1	8.0%	6.7%	6.5%
Q2	14.5%	13.2%	10.3%
Q3	23.0%	18.4%	15.0%
Mean	16.9%	13.8%	11.9%
IQR	15.0%	11.7%	8.4%

Trade payables of B/S total

	Small	Medium	Large
Q1	4.9%	5.3%	6.5%
Q2	9.1%	9.8%	10.6%
Q3	15.5%	15.7%	16.4%
Mean	12.3%	12.0%	12.6%
IQR	10.6%	10.5%	9.9%

Q1...Q3: quartile 1...3 / IQR: interquartile range

When comparing small and large companies, the following findings can be observed. Firstly, the median DSOs are significantly higher ($p < 0.001$[6]) for small companies, with a difference of 18.5 days. The median small company in the sample waits for almost 64 days for payment of their goods and services. A quarter of small companies have even collection periods of more than 93 days, compared with 65 days for large companies. Overall, this also translates into a median share of working capital tied up in trade receivables that is 4.2 percentage points higher for small enterprises. There may be several reasons for this difference as already described in section 3.4.

The picture is slightly different for trade payables. The difference in median DPOs is relatively smaller and insignificant with a delta of 5.9 days ($p = .052$[7]). Also, the group of small companies shows a higher median DPO (86.1 days) compared to the group of large companies (80.2 days). Therefore, the data do not support the hypothesis that large enterprises can obtain significantly longer payment terms from their suppliers. However, in relation to their balance sheet total, large enterprises rely more on trade payables financing than their smaller counterparts (10.6% vs. 9.1%, $p < 0.01$[8]).

4.5 Summary & Limitations

Overall, the analysis of the Osiris data mainly confirms the previous findings based on the ERICA data. Since 2010, there has been a significant reduction in DSOs, leading to a reduction in working capital tied up in receivables. However, the median collection period is still well above the regulator's 30-day general limit. The development of trade payables on the other hand measured by DPO or as percentage of balance sheet total is relatively stable over time. Comparing small and large companies again reveals a significant difference in DSOs and in the share of receivables and payables as well. Large companies have shorter collection periods, less working capital tied up in receivables and more financing through trade payables.

Although the analysis of the Osiris data has several advantages over ERICA, such as a stable panel, there are still some shortcomings regarding the interpretation. First, the analysis of DSOs needs to be based on total receivables as opposed to trade receivables only. Second, the analysis relies on financial data for publicly listed firms only. This means that the sample actually includes very few small- and medium-sized enterprises with a turnover of less than €50 million, which is the formal EU threshold for SMEs (European Commission, 2020).

5. Conclusion

In conclusion, the data examined in this paper, drawn from both the ERICA and Osiris databases, demonstrates that although overall collection periods have been reduced over the last years, the payment terms remain well above the general aim of the

[6] Two-sided Mann-Whitney-Test, unpaired.
[7] Two-sided Mann-Whitney-Test, unpaired.
[8] Two-sided Mann-Whitney-Test, unpaired.

European Commission of 30 days. The results highlight the discrepancies in payment practices between large and small firms. Large companies tend to benefit from shorter collection periods and more favourable financing conditions through trade payables, while smaller businesses face extended payment terms, leading to increased pressure on their working capital.

The question of whether the new EU regulation will come into force in the form decided by the EU Parliament is currently still unclear and depends on the approval of the European Council. In light of the potentially considerable implications, it would be advisable for companies to examine their own payment terms on both the customer and supplier sides in order to anticipate the potential consequences of the more rigorous regulations and to identify potential alternative financing strategies if necessary.

References

Banque de France. (n.d.). *ERICA (European Records of IFRS Consolidated Accounts) database*. https://www.bach.banque-france.fr/

Brealey, R., Myers, S., Allen, F., & Edmans, A. (2022). *Principles of Corporate Finance* (14th ed.). McGraw Hill.

European Commission. (2020). *User guide to the SME Definition*. https://ec.europa.eu/docsroom/documents/42921/attachments/1/translations/en/renditions/native

European Commission. (2023). *Proposal for a regulation of the European Parliament and of the council on combating late payment in commercial transactions, COM(2023) 533 final*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52023PC0533

European Committee of Central Balance Sheet Data Offices (ECCBSO). (2023). *ERICA User Guide Summary*. European Committee of Central Balance Sheet Data Offices (ECCBSO). http://www.bach.banque-france.fr

European Committee of Central Balance Sheet Data Offices (ECCBSO). (2024, February). *ERICA User's Guide*. ECCBSO. https://www.eccbso.org/system/files/2024-02/ERICA_users_guide_2022_data_F.pdf

European Parliament. (2024a). *P9_TA(2024)0299: European Parliament legislative resolution of 23 April 2024 on the proposal for a regulation of the European Parliament and of the Council on combating late payment in commercial transactions (COM(2023)0533 – C9-0338/2023 – 2023/0323(COD))*. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0299_EN.pdf

European Parliament. (2024b). *Procedure file 2023/0323(COD) - Commercial transactions: Combating late payment*. https://oeil.secure.europarl.europa.eu/oeil/popups/printficheglobal.pdf?id=746469&l=en

European Union (2011). *Directive 2011/7/EU of the European Parliament and of the Council of 16 February 2011 on Combating Late Payment in Commercial Transactions*. Official Journal of the European Union, L48, 1-10. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0007

Eurostat. (2008). *NACE Rev. 2: Statistical classification of economic activities in the European Community*. https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.pdf

Federau, M. (2024). Zahlungsziele und Net Working Capital. *Corporate Finance*, *15*(05–06), 109–115.

Högerle, B., Charifzadeh, M., Ferencz, M., & Kostin, K. (2020). The development of working capital management and its impact on profitability and shareholder value: Evidence from Germany. *Strategic Management*, *25*(2), 27–39. https://doi.org/10.5937/StraMan2002027H

IAS 2: Inventories (2003). https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2021/issued/part-a/ias-2-inventories.pdf

Le, B. (2019). Working capital management and firm's valuation, profitability and risk: Evidence from a developing market. *International Journal of Managerial Finance*, *15*(2), 191–204. https://doi.org/10.1108/IJMF-01-2018-0012

Moody's Analytics. (n.d.). *Orbis*. Retrieved 1 September 2024, from https://www.moodys.com/web/en/us/capabilities/company-reference-data/orbis.html

Sagner, J. S. (2014). *Working Capital Management: Applications and Case Studies*. John Wiley & Sons.

Singh, H. P., Kumar, S., & Colombage, S. (2017). Working capital management and firm profitability: A meta-analysis. *Qualitative Research in Financial Markets*, *9*(1), 34–47. https://doi.org/10.1108/QRFM-06-2016-0018

FHV
Vorarlberg University
of Applied Sciences

EU Regulation on combating late payment in commercial transactions

Working capital impact assessment for European firms

BIS/ECCBSO/Banco de España Workshop
New Insights from financial statements
Madrid, 17 Oct 2024

Markus Federau, Bernd Schwendinger

Agenda

1. EU Regulation
2. Research Objectives & Methodology
3. Sample & KPIs
4. Results
5. Conclusion
6. Remarks regarding statistical producers

EU Regulation



Status Quo

> Directive 2011/7/EU
> Applicable from 03/2013

Main Content

> Standard payment term (PT) of 30 days
> Max. PT 60 days
> PT over 60 days possible if…
…expressly agreed; and
…not 'grossly unfair'

Proposed Regulation



> Proposal by EU Commission in 09/2023
> Replacement of EU Directive with EU Regulation
> Adapted and adopted by EU Parliament in 04/2024

Main Content

> General PT limit of 30 days in B2B/B2G
> Increase to max. 60 days in B2B if mutually agreed
> Increase to max. 60 days for book industry and other slow-moving & seasonal goods
> No contractual restriction of assignment of receivables for use of financing services possible

Research Objectives & Methodology

Objectives

(1) Exploration of historical development
(2) Analysis of status quo
(3) Assessment of potential regulatory impact
(4) Recommendations regarding statistical procedures

Methodology

> Group financial statement analysis
> BACH / OSIRIS data
> Receivables/payables
> Days Sales Outstanding (DSO) / Days Payables Outstanding (DPO)
> Descriptives / U-Tests



Sample

ERICA

BANQUE DE FRANCE EUROSYSTÈME | ECCBSO

- Free aggregated consolidated balance sheet data of non-financial groups (mean, Q1-Q3, min/max)
- 8 EU countries + Turkey
- Since 2019 publicly available
- Pre-defined (proprietary) aggregation:
  Size, detailed sector, country + sector, total
- Size: S <250 m€ / M <1,500 m€ / L >=1,500 m€
- Changing sample composition
- Analysed period: 2010-2022
- Varying sample size: 829-1,187 p.a.

Osiris
MOODY'S

- Individual financial data of publicly listed companies
- Global coverage
- Applied filters:
  EU based + non-financial + IFRS application + rec./payables/sales/cogs >1m€
- Same size classes selected
- Stable panel approach
- Analysed period: 2010-2022
- Stable panels: 1,011/1,081
- Year 2022: 2,171/2,170



KPIs

	ERICA / BANQUE DE FRANCE EUROSYSTÈME / ECCBSO	Osiris / MOODY'S
DSO	$\frac{\text{Trade Receivables}}{\text{Daily Average Revenues}}$	$\frac{\text{Accounts Receivable (current)}}{\text{Total Revenues}} \times 360 \text{ days}$
AR in B/S Total	n/a	$\frac{\text{Accounts Receivable (current)}}{\text{Total Assets}}$
DPO	$\frac{\text{Trade Payables}}{\text{Daily Average Revenues}}$	$\frac{\text{Trade Payables}}{\text{Costs of Goods Sold (COGS)}} \times 360 \text{ days}$
AP in B/S Total	n/a	$\frac{\text{Trade Payables}}{\text{Total Liabilities \& Equity}}$

AR: Accounts Receivables / AP: Accounts Payables / B/S: Balance Sheet / DSO: Days Sales Outstanding / DPO: Days Payables Outstanding / n/a: not available

ERICA | BANQUE DE FRANCE EUROSYSTÈME | ECCBSO

Results – Longitudinal

DSO

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Q1	**45.4**	42.5	41.4	42.1	42.7	42.5	42.7	43.0	40.2	40.6	41.5	41.0	**35.6**
Q2	**68.3**	66.3	63.1	63.8	66.6	64.4	66.0	64.7	64.4	62.2	63.9	64.6	**59.6**
Q3	**103.3**	104.0	95.8	93.2	95.9	96.9	100.2	98.0	95.2	93.2	95.0	96.8	**84.5**
Mean	83.3	80.9	77.9	78.2	81.4	88.6	2,525.9	79.7	780.5	85.0	91.5	81.7	68.4
IQR	57.9	61.4	54.5	51.1	53.3	54.5	57.4	55.0	55.0	52.7	53.5	55.8	49.0

- Reduction in median DSOs
- Still well above 30/60 days
- No significance testing possible

DPO

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Q1	**31.9**	30.0	28.7	29.5	30.9	30.0	31.2	32.1	34.2	33.1	34.9	38.3	**32.6**
Q2	**48.4**	47.9	45.2	46.0	47.8	47.3	47.2	50.4	52.4	52.1	54.3	60.2	**52.1**
Q3	**74.9**	75.7	70.5	71.4	73.6	72.8	74.1	76.2	79.8	77.8	84.8	92.0	**79.5**
Mean	111.7	73.0	63.0	65.5	77.2	1,048	73,692	95.9	99,245	77,951	452.7	100.7	82.0
IQR	43.0	45.8	41.8	41.9	42.7	42.8	43.0	44.1	45.5	44.7	50.0	53.8	46.9

Q1…Q3: quartile 1…3 / IQR: interquartile range

- Increase in DPOs
- High distortion of data by outliers
- Impossibility to proxy payment periods

Results – Size 2022

DSOs [days]	Small	Medium	Large
Q1	43.7	37.4	32.4
Q2	**71.2**	56.8	**49.4**
Q3	103.9	80.1	72.8
Mean	85.6	63.3	56.3
IQR	60.1	42.8	40.5

DPOs [days]	Small	Medium	Large
Q1	31.4	31.3	34.8
Q2	**56.4**	48.3	**53.2**
Q3	86.7	76.3	75.9
Mean	122.4	59.4	62.9
IQR	55.3	45.0	41.1

Q1…Q3: quartile 1…3 / IQR: interquartile range

- Notable difference in DSOs small vs. large
- Potentially through bargaining power or different customer structures
- DPOs only smaller differences

Osiris
MOODY'S

Results – Longitudinal

DSO

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Q1	**43.2**	41.7	40.8	40.8	40.6	40.0	41.4	40.5	39.9	37.9	38.3	39.0	**35.0**
Q2	**62.2**	60.9	59.3	59.1	59.0	56.5	59.9	59.6	58.9	56.6	55.1	56.4	**52.7**
Q3	**89.8**	88.1	85.0	82.8	84.1	80.8	83.6	81.1	81.7	77.7	77.3	76.9	**74.7**
Mean	74.2	73.2	70.4	69.4	70.4	68.0	70.3	67.9	66.9	63.2	64.1	65.8	62.2
IQR	46.6	46.4	44.1	42.0	43.4	40.8	42.3	40.6	41.8	39.7	39.0	37.8	39.6

> Reduction in median DSOs by 9.5 days ($p < 0.001$*)

> Highest decrease of Q3 between 2011 and 2013

AR in B/S Total

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Q1	**10.2%**	10.2%	10.0%	9.7%	9.3%	9.0%	9.3%	9.4%	9.1%	7.7%	7.1%	7.7%	**7.7%**
Q2	**16.3%**	16.4%	15.7%	15.3%	15.3%	14.9%	15.0%	15.2%	14.7%	13.4%	12.0%	12.5%	**13.1%**
Q3	**24.1%**	24.0%	23.1%	23.0%	22.9%	22.4%	22.5%	22.7%	21.9%	19.8%	18.1%	18.3%	**19.2%**
Mean	18.5%	18.6%	18.1%	17.7%	17.5%	17.2%	17.2%	17.2%	16.8%	15.2%	13.9%	14.3%	14.8%
IQR	13.9%	13.8%	13.1%	13.3%	13.6%	13.4%	13.2%	13.3%	12.8%	12.1%	10.9%	10.6%	11.5%

B/S: balance sheet / Q1…Q3: quartile 1…3 / IQR: interquartile range

> Reduction in median AR ratios by 3.2 ppts ($p < 0.001$*)

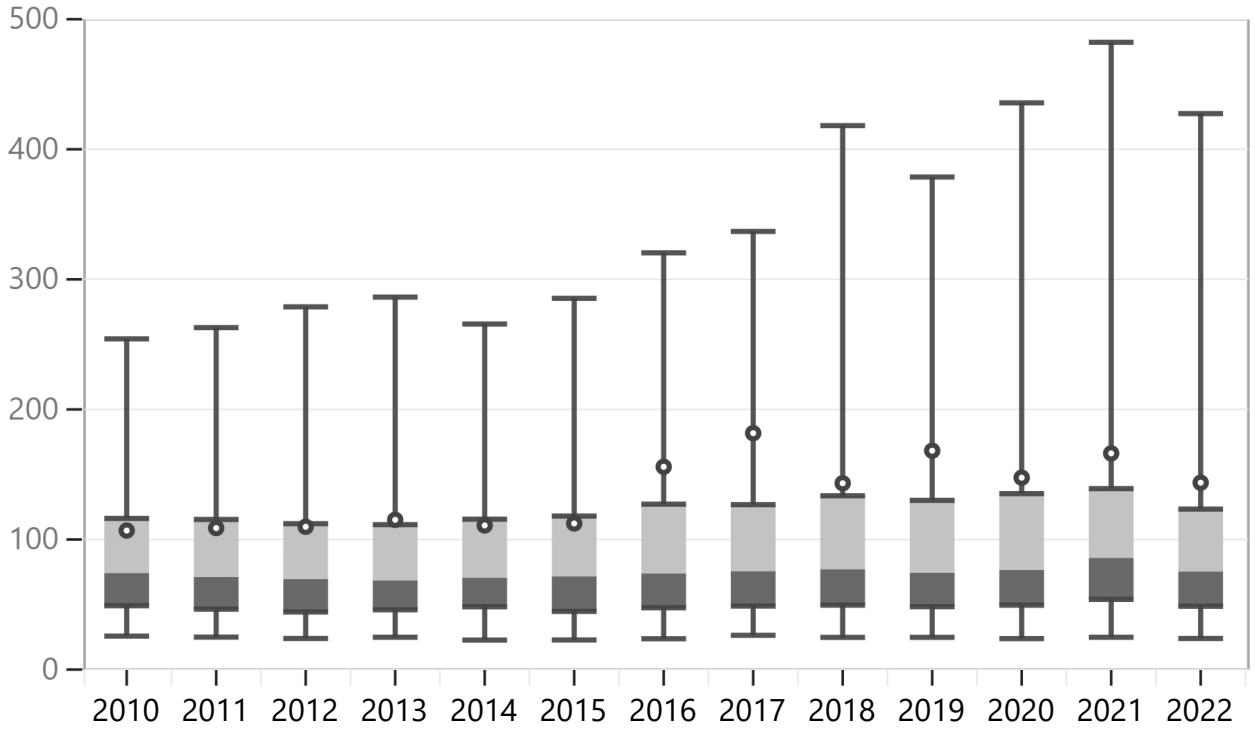
> Efficient receivables mgmt. freeing up WC

* Mann-Whitney-U-Test, two-sided, paired

Osiris
MOODY'S

Results – Longitudinal

DPO



Whiskers representing the 0.05 and 0.95 quantiles

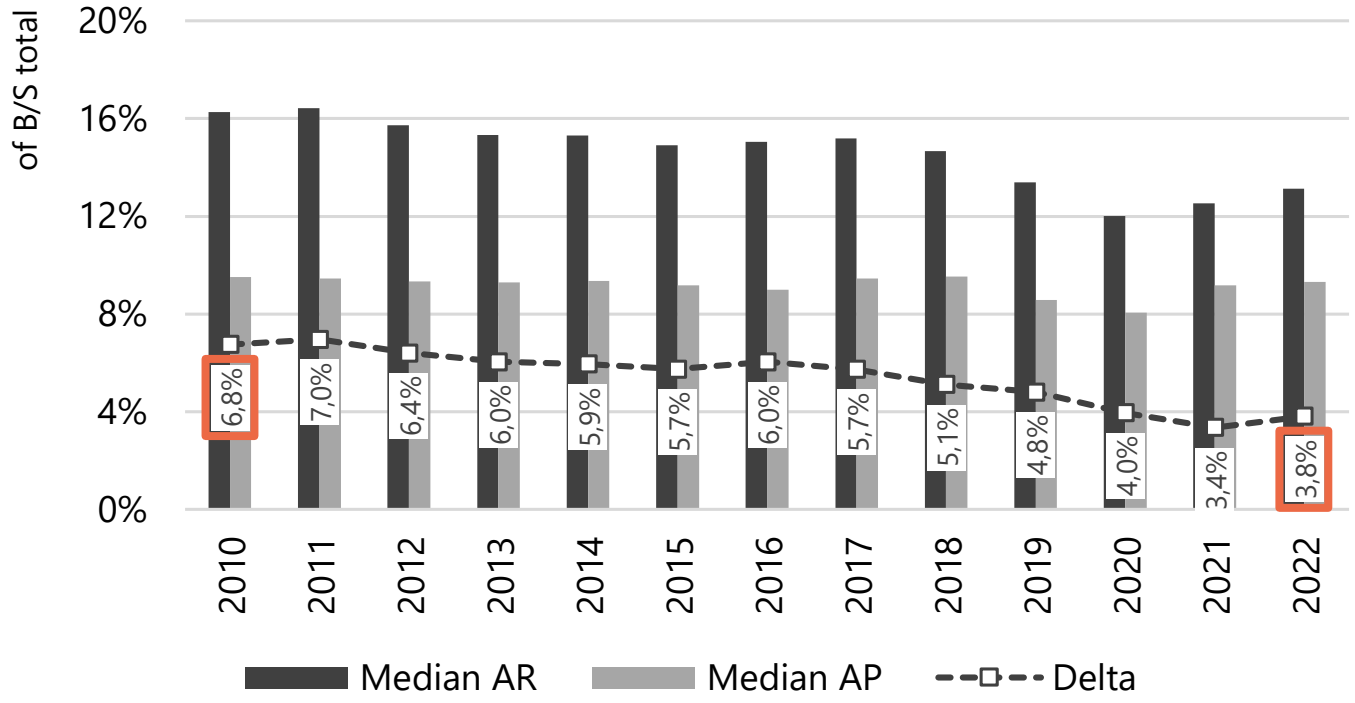
- Slight but significant increase in median DPOs +1.1 days ($p < 0.001$*)
- Steep increase of DPOs of Q4

* Mann-Whitney-U-Test, two-sided, paired

Osiris
MOODY'S

Results – Longitudinal

AR vs. AP in B/S Total



- Overall freeing up of NWC over time
- Reduction of gap between AR/AP by 3 ppts

Osiris
MOODY'S

Results – Size 2022

DSOs [days]	Small	Medium	Large
Q1	41.6	30.2	29.6
Q2	**63.6**	49.9	**45.1**
Q3	93.1	68.5	65.1
Mean	103.4	52.8	51.1
IQR	51.4	38.3	35.5

DPOs [days]	Small	Medium	Large
Q1	52.6	43.1	50.9
Q2	**86.1**	70.6	**80.2**
Q3	157.3	122.7	130.9
Mean	190.7	253.9	164.6
IQR	104.8	79.6	80.0

Accounts receivables of B/S total	Small	Medium	Large
Q1	8.0%	6.7%	6.5%
Q2	**14.5%**	13.2%	**10.3%**
Q3	23.0%	18.4%	15.0%
Mean	16.9%	13.8%	11.9%
IQR	15.0%	11.7%	8.4%

Trade payables of B/S total	Small	Medium	Large
Q1	4.9%	5.3%	6.5%
Q2	**9.1%**	9.8%	**10.6%**
Q3	15.5%	15.7%	16.4%
Mean	12.3%	12.0%	12.6%
IQR	10.6%	10.5%	9.9%

Q1…Q3: quartile 1…3 / IQR: interquartile range

- DSOs large vs. small -18.5 days ($p < 0.001$*)
- 4.2 ppts less WC tied in receivables ($p < 0.001$*)
- Smaller difference in DPOs of -5.9 days ($p = 0.052$*)

* Mann-Whitney-U-Test, two-sided, unpaired

Conclusion

1. EU regulation upcoming but still pending
2. Similar results based on ERICA / OSIRIS samples
3. Collection periods decreased significantly but still well above 30-day target
4. Larger companies benefiting from shorter collection periods
5. No longer DPOs for large vs. small companies
6. Review of payment terms for companies and preparation for upcoming regulation recommended

Remarks regarding statistical producers

Limitations of ERICA usage in benchmarking and scientific research

1. Long update cycle; some 2022 data still missing
2. No panel data available
3. Definition of pre-calculated ratios not public
   → *Possible Solution:*
   Include ratio definitions in ERICA documentation
4. Only pre-defined aggregations (e.g. size, industry) possible
   → *Possible Solution:*
   Continuing of aggregated data only but with dynamic filtering capabilities
5. No possibility of hypothesis testing
   → *Possible Solution:*
   Add std. dev. or rank sums to sub-samples
6. Only absolute B/S & P+L figures available
   → Possible Solution:
   Add ratios (% of B/S total or % of sales)

Resources

- Banque de France. (n.d.). *ERICA (European Records of IFRS Consolidated Accounts) database*. https://www.bach.banque-france.fr/
- Brealey, R., Myers, S., Allen, F., & Edmans, A. (2022). *Principles of Corporate Finance* (14th ed.). McGraw Hill.
- European Commission. (2020). *User guide to the SME Definition*. https://ec.europa.eu/docsroom/documents/42921/attachments/1/translations/en/renditions/native
- European Commission. (2023). *Proposal for a regulation of the European Parliament and of the council on combating late payment in commercial transactions, COM(2023) 533 final*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52023PC0533
- European Committee of Central Balance Sheet Data Offices (ECCBSO). (2023). *ERICA User Guide Summary*. European Committee of Central Balance Sheet Data Offices (ECCBSO). http://www.bach.banque-france.fr
- European Committee of Central Balance Sheet Data Offices (ECCBSO). (2024, February). *ERICA User's Guide*. ECCBSO. https://www.eccbso.org/system/files/2024-02/ERICA_users_guide_2022_data_F.pdf
- European Parliament. (2024a). *P9_TA(2024)0299: European Parliament legislative resolution of 23 April 2024 on the proposal for a regulation of the European Parliament and of the Council on combating late payment in commercial transactions (COM(2023)0533 – C9-0338/2023 – 2023/0323(COD))*. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0299_EN.pdf
- European Parliament. (2024b). *Procedure file 2023/0323(COD) - Commercial transactions: Combating late payment*. https://oeil.secure.europarl.europa.eu/oeil/popups/printficheglobal.pdf?id=746469&l=en
- European Union (2011). *Directive 2011/7/EU of the European Parliament and of the Council of 16 February 2011 on Combating Late Payment in Commercial Transactions*. Official Journal of the European Union, L48, 1-10. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0007
- Eurostat. (2008). *NACE Rev. 2: Statistical classification of economic activities in the European Community*. https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.pdf
- Federau, M. (2024). Zahlungsziele und Net Working Capital. *Corporate Finance*, *15*(05–06), 109–115.
- Högerle, B., Charifzadeh, M., Ferencz, M., & Kostin, K. (2020). The development of working capital management and its impact on profitability and shareholder value: Evidence from Germany. *Strategic Management*, *25*(2), 27–39. https://doi.org/10.5937/StraMan2002027H
- IAS 2: Inventories (2003). https://www.ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2021/issued/part-a/ias-2-inventories.pdf
- Le, B. (2019). Working capital management and firm's valuation, profitability and risk: Evidence from a developing market. *International Journal of Managerial Finance*, *15*(2), 191–204. https://doi.org/10.1108/IJMF-01-2018-0012
- Moody's Analytics. (n.d.). *Orbis*. Retrieved 1 September 2024, from https://www.moodys.com/web/en/us/capabilities/company-reference-data/orbis.html
- Sagner, J. S. (2014). *Working Capital Management: Applications and Case Studies*. John Wiley & Sons.
- Singh, H. P., Kumar, S., & Colombage, S. (2017). Working capital management and firm profitability: A meta-analysis. *Qualitative Research in Financial Markets*, *9*(1), 34–47. https://doi.org/10.1108/QRFM-06-2016-0018

Thank you

Q&A

