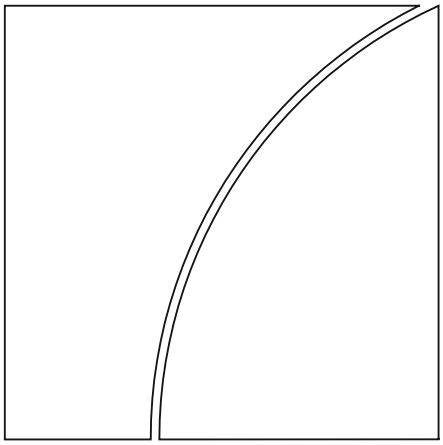
Irving Fisher Committee on Central Bank Statistics



IFC Bulletin

No 50

The use of big data analytics and artificial intelligence in central banking

Proceedings of the IFC – Bank Indonesia International Workshop and Seminar in Bali on 23-26 July 2018

May 2019

BANK FOR INTERNATIONAL SETTLEMENTS

Contributions in this volume were prepared for the IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"* in Bali, Indonesia, on 23-26 July 2018. The views expressed are those of the authors and do not necessarily reflect the views of the IFC, its members, the BIS and the institutions represented at the meeting.

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

ISSN 1991-7511 (online)

ISBN 978-92-9259-262-2 (online)

The use of big data analytics and artificial intelligence in central banking

IFC Bulletin No 50
May 2019

Proceedings of the IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Conference overview

Opening remarks

1. The big data revolution: new data sources and analytical techniques

2. Opportunities for central banks

Enhanced statistics

Macroeconomic forecasting

Financial market forecasting and monitoring

The use of big data analytics and artificial intelligence in central banking

Okiriza Wibisono, Hidayah Dhini Ari, Anggraini Widjanarti, Alvin Andhika Zulen and Bruno Tissot[1]

Executive summary

Information and internet technology has fostered new web-based services that affect every facet of today's economic and financial activity. This creates enormous quantities of "**big data**" – defined as "*the massive volume of data that is generated by the increasing use of digital tools and information systems*" (FSB (2017)). Such data are produced in real time, in differing formats, and by a wide range of institutions and individuals. For their part, central banks face a surge in "financial big data sets", reflecting the combination of new, rapidly developing electronic footprints as well as large and growing financial, administrative and commercial records.

This phenomenon has the potential to strengthen analysis for decision-making, by providing more complete, immediate and granular information as a complement to "traditional" macroeconomic indicators. To this end, a number of techniques are being developed, often referred to as "***big data analytics***" and "***artificial intelligence***" (AI). These promise faster, more holistic and more connected insights, as compared with traditional statistical techniques and analyses. An increasing number of central banks have launched specific big data initiatives to explore these issues. They are also sharing their expertise in collecting, working with, and using big data, especially in the context of the BIS's Irving Fisher Committee on Central Bank Statistics (IFC); see IFC (2017).

Getting the most out of these new developments is no trivial task for policymakers. Central banks, like other public authorities, face numerous challenges, especially in handling these new data and using them for policy purposes. In particular, significant resources are often required to handle large and complex data sets, while the benefits of such investments are not always clear-cut. For instance, to what extent should sophisticated techniques be used to deal with this type of information? What is the added value over more traditional approaches, and how should the results be interpreted? How can the associated insights be integrated into current decision-making processes and be communicated to the public? And, lastly, what are the best strategies for central banks seeking to realise the full potential of

[1] Respectively Big Data Analyst (okiriza_w@bi.go.id); Head of Digital Data Statistics and Big Data Analytics Development Division (dhini_ari@bi.go.id); Big Data Analyst (anggraini_widjanarti@bi.go.id); Big Data Analyst (alvin_az@bi.go.id); and Head of Statistics and Research Support, BIS, and Head of the IFC Secretariat (Bruno.Tissot@bis.org).

The views expressed here are those of the authors and do not necessarily reflect those of Bank Indonesia, the Bank for International Settlements (BIS), or the Irving Fisher Committee on Central Bank Statistics (IFC).

new big data information and analytical tools, considering in particular resource constraints and other priorities?

Against this backdrop, Bank Indonesia organised with support from the BIS and the IFC a workshop on "*Big data for central bank policies*" and a high-level policy-oriented seminar on "*Building pathways for policymaking with big data*". Convening in July 2018 at Bali, Indonesia, the events were attended by officials from central banks, international organisations and national statistical offices from more than 30 jurisdictions across the globe, as well as by representatives from other public agencies, the financial sector and academia. This proved a useful opportunity to take stock of the various big data pilots conducted by the central bank community and of the growing use of big data analytics and associated AI techniques to support public policy. The following points of interest were highlighted:

- Big data offers **new types of data source** that complement more traditional varieties of statistics. These sources include Google searches, real estate and consumer prices displayed on the internet, and indicators of economic agents' sentiment and expectations (eg social media).
- Thanks to IT innovation, **new techniques** can be used to collect data (eg web-scraping), process textual information (text-mining), match different data sources (eg fuzzy-matching), extract relevant information (eg machine learning) and communicate or display pertinent indicators (eg interactive dashboards).
- In particular, big data techniques such as decision trees may shed i**nteresting light** on the decision-making process of economic agents, eg how investors behave in financial markets. As another example, indicators of economic uncertainty extracted from news articles, could help explain movements of macroeconomic indicators. This illustrates big data's potential in providing insights not only into what happened, but also into what might happen and why.
- In turn, these new insights can usefully **support central bank policies** in a wide range of areas, such as market information (eg credit risk analysis), economic forecasting (eg nowcasting), financial stability assessments (eg network analysis) and external communication (eg measurement of agents' perceptions). Interestingly, the approach can be very granular, helping to target specific markets, institutions, instruments and locations (eg zip codes) and, in particular, to support macroprudential policies. Moreover, big data indicators are often more timely than "traditional" statistics – for instance labour indicators can be extracted from online job advertisements almost in real time.
- As a note of caution, feedback from central bank pilot projects consistently highlights the complex privacy implications of dealing with big data, and the associated **reputational risks**. Moreover, while big data applications such as machine learning algorithms can excel in terms of predictive performance, they can lend themselves more to explaining what is happening rather then why. As such, they may be exposed to public criticism when insights gained in this way are used to justify policy decisions.
- Another **concern** is that, as big data samples are often far from representative (eg not everyone is on Facebook, and even fewer are on Twitter), they may not be as reliable as they seem. Lastly, there is a risk that collecting and

processing big data will be hindered by privacy laws and/or change in market participants. Relevant authorities should coordinate their efforts so that they can utilise the advantages of big data analytics without compromising data privacy and confidentiality.

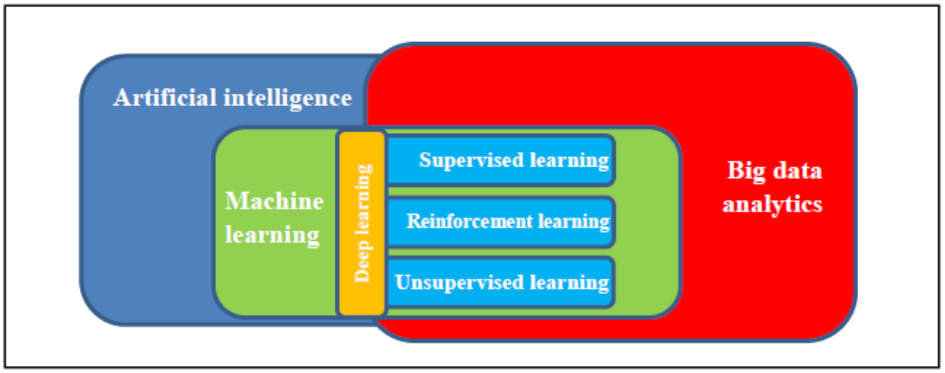
The related presentations, referred to in this overview and included in this *IFC Bulletin*, analysed the various aspects related to the use of big data and associated techniques by central banks. They cover three main aspects: (1) an assessment of the main big data sources and associated analytical techniques that are relevant for central banks; (2) the insights provided by big data for economic policy, with an overview of concrete central bank projects aiming at improving statistical information, macroeconomic analysis and forecasting, financial market monitoring and financial risk assessment; and (3) the use of big data in crafting central bank policies, including organisational aspects and related challenges.

1. The big data revolution: new data sources and analytical techniques

As emphasised in the opening remarks by Erwin Rijanto, Deputy Governor of Bank Indonesia, policymakers should not miss out on the opportunities provided by big data – described by some as the new oil of the 21st century (The Economist (2017)). Public institutions are not the main producers of big data sets, and some of this information may have little relevance for their daily work. Yet central banks are increasingly dealing with "financial big data" sources that impinge on a wide range of their activities, as noted by Claudia Buch, IFC Chair and Vice President of the Deutsche Bundesbank.

Data volumes have surged hand in hand with the development of specific techniques for their analysis, thanks to "*big data analytics*" – broadly referring to the general analysis of these data sets – and "*artificial intelligence*" (AI) – defined as "*the theory and development of computer systems able to perform tasks that traditionally have required human intelligence*" (FSB (2017)). Strictly speaking, these two concepts can differ somewhat (for instance, one can develop tools to analyse big data sets that are not based on AI techniques), as shown in Graph 1.

Graph 1: A schematic view of AI, machine learning and big data analytics



Source: FSB (2017).

In practice, and as underlined by Yati Kurniati (Bank Indonesia) in her welcoming remarks, big data analytics are not very different from traditional econometrical techniques, and indeed they borrow from many long-established methodologies and techniques developed for general statistics; for instance, principal component analysis, developed at the beginning of the last century. Yet one key characteristic is that they are applied to modern data sets that can be both very large and complex. Extracting relevant information from these sources is not straightforward, often requiring a distinct set of skills, depending on the type of information involved. As a result, big data analytics and AI techniques comprise a variety of statistical/modelling approaches, such as machine learning, text-mining techniques, network analysis, agent-based modelling[2] etc.

The seminar and workshop provided an opportunity to review, first, the main big data sources relevant for central banks, and, second, the principal techniques developed in recent years for analysing big data – focusing on the classification and clustering of information derived from large quantitative data sets, with machine learning, text-mining and network analysis all playing an important role.

Big data information for central banks

Three main sources of big data can be identified, as Paul Robinson (Bank of England) noted in his introductory presentation on fundamental concepts and frameworks.[3] These categories are related to (i) social networks (human-sourced information such as blogs and searches); (ii) traditional business systems (process-mediated data, such as files produced by commercial transactions, e-commerce, credit card operations); and (iii) the internet of things (machine-generated data, such as information produced by pollution/traffic sensors, mobile phones, computer logs etc). These are very generic categories and, in practice, big data will comprise multiple types of heterogeneous data set derived from these three main sources.

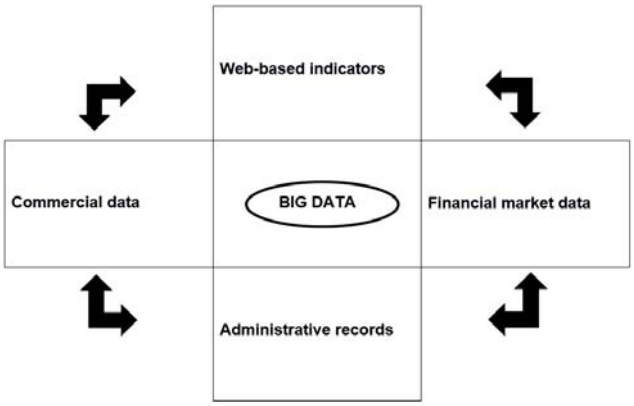
Focusing more specifically on central banks, the presentation by Bruno Tissot (BIS) identified **four types of data set** that would usually be described as financial big data (see Graph 2): internet-based indicators, commercial data sets, financial market indicators and administrative records.[4]

[2] See Haldane (2018), who argues that big data can facilitate policymakers' understanding of economic agents' reactions through the exploration of behaviours in a "virtual economy".

[3] Following the work conducted under the aegis of the United Nations (see Meeting of the Expert Group on International Statistical Classifications (2015)).

[4] For the use of administrative data sources for official statistics, see for instance Bean (2016) in the UK context.

Graph 2: Four main types of financial big data set



Compared with the private sector,[5] central banks' use of **web-based indicators** may be somewhat more limited, especially as regards unstructured data such as images. Even so, several projects are under way to make use of data collected on the internet to support monetary and financial policymaking (see Section 2). Moreover, an important aspect relates to the increased access to **digitalised information**, reflecting both the fact that more and more textual information is becoming available on the web (eg social media) and also that "traditional" printed documents can now be easily digitalised, searched and analysed in much the same way as web-based indicators.

In reality, however, the bulk of the financial big data sets relevant to central banks consists of the very granular information provided by large and growing records covering commercial transactions, financial market developments and administrative operations. This type of information has been spurred by the expansion of the **micro-level data sets** collected in the aftermath of the Great Financial Crisis (GFC) of 2007–09, especially in the context of the Data Gaps Initiative (DGI) endorsed by the G20 (FSB-IMF (2009)). For instance, significant efforts have been made globally to compile large and granular loan-by-loan and security-by-security databases as well as records of individual derivatives trades (IFC (2018)). As a result, central banks now have at their disposal very detailed information on the financial system, including at the level of specific institutions, transactions or instruments.

Extracting knowledge from large quantitative data sets: classification and clustering

The expansion of big data sources has gone hand in hand with the development of new analytical tools to deal with them. The first, and particularly important, category of these big data techniques aims at extracting summary information from large quantitative data sets. This is an area that is relatively close to "traditional statistics", as it does not involve the treatment of unstructured information (eg text, images). In fact, many big data sets are well structured, and can be appropriately dealt with using statistical algorithms developed for numerical data sets. The main goal is to obtain summary indicators by condensing the large amount of data points available,

[5] Especially the major US technology companies ("GAFAs"): Google, Apple, Facebook and Amazon.

basically by finding similarities between them (through classification) and regrouping them (through clustering).

Many of these techniques involve so-called ***machine learning***. This is a subset of AI techniques, that can be defined as "*a method of designing a sequence of actions to solve a problem that optimise automatically through experience and with limited or no human intervention*" (FSB (2017)). This approach is quite close to conventional econometrics, albeit with three distinct features. First, machine learning is typically focused on prediction rather than identifying a causal relationship. Second, the aim is to choose an algorithm that fits with the actual data observed, rather than with a theoretical model. Third, and linked to the previous point, the techniques are selected by looking at their goodness-to-fit, and less at the more traditional statistical tests used in econometrics.

The lecture by Sanjiv Das (Santa Clara University) recalled that there are several categories of machine learning, which can be split into two main groups. First, in **supervised machine learning**, "*an algorithm is fed a set of 'training' data that contain labels on the observations*" (FSB (2017)). The goal is to classify individual data points, by identifying, among several classes (ie categories of observations), the one to which a new observation belongs. This is inferred from the analysis of a sample of past observations, ie the training data set, for which their group (category) is known. The objective of the algorithm is to predict the category of a new observation, depending on its characteristics. For instance, to predict the approval of a new loan ("yes" or "no", depending on its features and in comparison with an observed historical data set of loans that have been approved or rejected); or whether a firm is likely to default in a few months. Various algorithms can be implemented for this purpose, including logistic regression techniques, linear discriminant analysis, Naïve Bayes classifier, support vector machines, k-nearest neighbours, decision trees, random forest etc.

The second group is **unsupervised machine learning**, for which "*the data provided to the algorithm do not contain labels*". This means that categories have not been identified ex ante for a specific set of observations, so that the algorithm has to identify the clusters, regrouping observations for which it detects similar characteristics or "patterns". Two prominent examples are clustering and dimensionality reduction algorithms. In **clustering**, the aim is to detect the underlying groups that exist in the granular data set – for example, to identify groups of customers or firms that have similar characteristics – by putting the most similar observations in the same cluster in an agglomerative way (bottom-up approach).[6] **Dimensionality reduction** relates to the rearrangement of the original information in a smaller number of pockets, in a divisive (top-down) way; the objective is that the number of independent variables becomes (significantly) smaller, without too much compromise in terms of information loss.

There are, of course, additional types of algorithm. One is **reinforcement learning**, which complements unsupervised learning with additional information feedback, for instance through human intervention. Another is **deep learning** (or artificial neural networks), based on data representations inspired by the function of neurons in the brain. Recent evaluations suggest that deep learning can perform better than traditional classification algorithms when dealing with unstructured data

[6] More precisely, cluster analysis can be defined as "*a statistical technique whereby data or objects are classified into groups (clusters) that are similar to one another but different from data or objects in other clusters*" (FSB (2017)).

such as texts and images – one reason being that applying traditional quantitative algorithms is problematic, as it requires unstructured information to be converted into a numerical format. In contrast, deep learning techniques can be used to deal directly with the original raw data.

In view of this diversity, the choice of a specific algorithm will depend on the assumptions made regarding the features of the data set of interest – for instance, a Naïve Bayes classifier would be appropriate when the variables are assumed to be independent and follow a Gaussian distribution. In practice, data scientists will have to identify which algorithm works best for the problem at hand, often requiring a rigorous and repetitive process of trial and error; this is often as much art as it is science.

In choosing the right "model", it is important to define an evaluation metric. The aim is to measure how well a specific algorithm fits, and to compare the performance of alternative algorithms. The most straightforward metric for classification is **prediction accuracy**, which is simply the percentage of observations for which the algorithm predicts the class variable correctly (this will usually be done by comparing the result of the algorithm with what a human evaluator would conclude on a specific data sample). But an accuracy metric may not be suitable for all exercises, particularly in the case of an unbalanced distribution of classes. For example, when looking at whether a transaction is legitimate or fraudulent, a very simplistic model could be adopted that assumes that all transactions are legitimate: its accuracy will look very high, because *a priori* most transactions are not fraudulent; but the usefulness of such a simple model would be quite limited. Hence, other metrics have to be found for evaluating algorithms when the distribution of classes is highly unbalanced.[7] Another possible approach is to address the class imbalance issue at the observation level, for instance, by duplicating (over-sampling) elements from the minority class or, conversely, by discarding those (under-sampling) from the dominant class.

Text mining

Another rapidly developing area of big data analytics is text-mining, ie analysis of semantic information – through the automated analysis of large quantities of natural language text and the detection of lexical or linguistic patterns with the aim of extracting useful insights. While most empirical work in economics deals with numerical indicators, such as prices or sales data, a large and increasing amount of textual information is also generated by economic and financial activities – including internet-based activities (eg social media posts), but also the wider range of textual information provided by, say, company financial reports, media articles, public authorities' deliberations etc. Analysing this unstructured information has become of key interest to policymakers, not least in view of the important role played by "soft" indicators such as confidence and expectations during the GFC. As illustrated in the lecture delivered by Stephen Hansen (University of Oxford), text-mining techniques can usefully be applied to dealing with these data in a structured, quantitative way.

[7] Such other metrics include, for instance, precision, recall and F1-score. For binary (two-class) classification, precision is defined as the percentage of times an algorithm makes a correct prediction for the positive class; recall is defined as the percent of positive class that the algorithm discovers from a given data set; and the F1-score is the harmonic average of precision and recall.

Text analysis typically starts with some standard **pre-processing steps**, such as tokenisation (splitting text into words), stopword removal (discarding very frequent/non-topical words eg "a", "the", "to"), stemming or lemmatising (converting words into their root forms, for instance "prediction" and "predicted" into "predict"), and merging words within a common message (eg "Bank" and "Indonesia" grouped into "Bank Indonesia"). Once this is done, the initial document can be transformed into a document-term matrix, which indicates for each specific text a term's degree of appearance (or non-appearance). This vectoral text representation is made of numerical values that can then be analysed by quantitative algorithms; for example, to measure the degree of similarity between documents by comparing the related matrixes (Graph 3).

One popular algorithm for working on textual information is the **Latent Dirichlet Allocation** (LDA).[8] This assumes that documents are distributed by topics, which in turn are distributed by keywords. For example, one document may combine, for a respective 20% and 80%, a "monetary" and an "employment" topic, based on the number of words reflecting this topic distribution (ie 20% of them related to words such as "inflation" or "interest rate", and the remaining 80% related to words such as "jobs" and "labour"). Based on these calculations, one can build an indicator measuring how frequently a specific topic appears over time, for instance, to gauge the frequency of the messages related to "recession" – providing useful insights when monitoring the state of the economy.

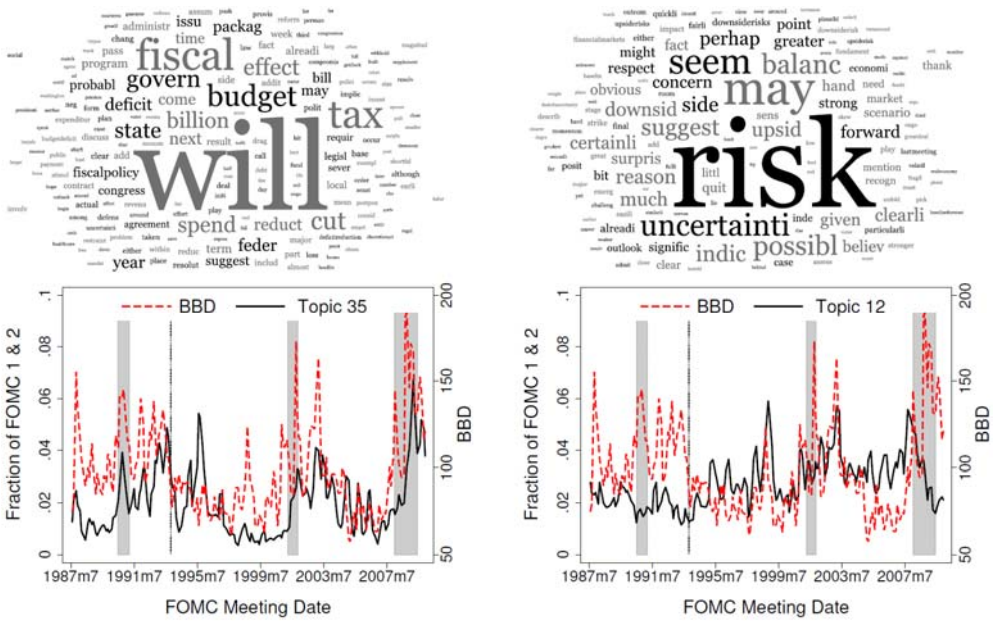
Besides quantitative algorithms, simpler **dictionary-based methods** can be also employed for analysing text data. A set of keywords can be selected that are relevant to the topic of interest – for example, a keyword related to "business confidence". Then an index can be constructed based on how frequently these selected keywords appear in a given document, allowing the subject indicator to be assessed (eg the evolution of business sentiment).[9] A prominent example is the Economic Policy Uncertainty (EPU), which quantifies the degree of uncertainty based on the appearance of a set of economic-, policy-, and uncertainty-related keywords in news articles; by the end of 2018, more than 20 country-specific EPU indexes had been compiled.[10]

[8] See Blei et al (2003).

[9] See for example Tetlock (2007) and Loughran and McDonald (2011).

[10] See Baker et al (2016) and www.policyuncertainty.com/.

Graph 3: Topic distributions obtained from text-mining techniques[11]



Source: Hansen (2018).

Network analysis

A third important area of big data analytics refers to financial network analysis, which can be seen as the analysis of the relations between the elements constituting the financial system. Insights into the functioning of this "network" are derived from graphical techniques and representations. The lecture by Kimmo Soramäki (Financial Network Analytics (FNA)) showed how this approach can measure how data is connected to other data, clarify how these connections matter and show how complex systems move in time. The approach can be particularly effective for big data sets, allowing for the description of complex systems characterised by rich interactions between their components.

The main **modes of analysis** comprise top-down approaches (eg analysis of system-wide risk), bottom-up analyses (eg analysis of connections between specific nodes of the system), network features analyses (eg transmission channels) and agent-based modelling – eg analysis of specific agents involved in the network, for instance, the role of central counterparties (CCPs) in the financial system. Typically, the work will involve three phases, ie analysis (data visualisation and identification of potential risks), monitoring (eg detection of anomalies in real time) and simulation (eg scenarios and stress tests).

In practice, a network is made of elements (nodes), linked to each other either directly or indirectly, and this can be represented by several types of graph. An important concept is **centrality**, which relates to the importance of nodes (or links) in the network, and which can be measured by specific metrics. Another is **community detection**, which aims at simplifying the visualisation of a large and

[11] Distributions obtained from LDA (black, solid line) and EPU dictionary-based index (BBD; red, dashed line). The word-clouds represent word distributions within each topic, with more frequent words shown in larger fonts.

complex network by regrouping nodes in clusters and filtering noise, through the use of specific machine learning algorithms (see above).

This sort of analysis appears particularly well suited to representing **interconnectedness** within a system, for instance, by mapping the global value chain across countries and sectors or the types of exposure incurred by financial institutions.[12] One example is recent work to assess the role of CCPs in the financial system by looking at the connections between them as well as with other financial institutions such as banking groups, in particular, by considering subsidiary-parent relationships (CPMI-IOSCO (2018)). This can help to reveal how a disruption originating in one single CCP would affect that CCP's clearing members and, in turn, other CCPs.

2. Opportunities for central banks

Big data can play an important role in improving the quality of economic analysis and research, as increasingly recognised by policymakers. This was the starting point for the presentation by Gabriel Quirós-Romero (IMF). The IMF is researching big data as a new way of measuring economic indicators, such as prices, labour market conditions, the housing market, business sentiment etc (Hammer et al (2017)).

Many central banks are now working on how to make use of the characteristics of financial big data sets in pursuing their mandates (Cœuré (2017)). As recalled in the introduction by Paul Robinson (Bank of England), big data has many advantages in terms of details, flexibility, timeliness and efficiency, as summarised in the list of their so-called "Vs" – eg volumes, variety, velocity, veracity and value; see Laney (2001) and Nymand-Andersen (2016). Central banks are interested in developing various pilot projects to better understand the new data sets and techniques, assess their value added in comparison with traditional approaches, and develop concrete "use cases" (IFC (2015)).

This experience has highlighted the opportunities that big data analytics can provide in key areas of interest to central banks, namely (i) the production of statistical information; (ii) macroeconomic analysis and forecasting; (iii) financial market monitoring; and (iv) financial risk assessment.

More and enhanced statistical information

Big data can be a useful means of improving the official statistical apparatus. First, it can be an **innovative source of support for the current production of official statistics**, offering access to a wider set of data, in particular to those that are available in an "*organic*" way. Unlike statistical surveys and censuses, these data are usually not collected ("*designed*") for a specific statistical purpose, being the by-product of other activities (Groves (2011)). Their range is quite large, covering transaction data (eg prices recorded online), aspirational data (eg social media posts, product reviews displayed on the internet), but also various commercial, financial and administrative indicators. In addition, they present various advantages for statistical

[12] For a recent example of the monitoring of network effects for global systemic institutions in the context of the DGI, see FSB (2011) and Bese Goksu and Tissot (2018).

compilers, such as their rapid availability and the relative ease of collection and processing with modern computing techniques (see Graph 4) – always noting, however, that actual access to such sources, private or public, can be restricted by commercial and/or confidentiality considerations.

Graph 4: Relative advantages of designed versus organic data

	Designed data	Organic data
Structure	Geographic and socio-economic	Behaviour
Representative	Yes	No
Sample selection	Response rates deteriorating	Extreme
Intrusive	Extremely intrusive	Non-intrusive
Cost	Large	Small
Curation	Well studied	Unclear
Privacy	Well protected	Large violations of privacy

Source: Rigobon (2018).

Organic data can be used to **enhance existing statistical exercises**, especially in improving coverage when this is incomplete. In some advanced economies, the direct web-scraping[13] of online retailers' prices data can, for instance, be used to better measure some specific components of inflation, such as fresh food prices.[14] At the extreme, these data can replace traditional indicators in countries where the official statistical system is underdeveloped. As noted by Roberto Rigobon (MIT Sloan School of Management), one example is the Billion Prices Project,[15] which allows inflation indices to be constructed for countries that lack an official and/or comprehensive index. Similarly, a number of central banks in emerging market economies have compiled quick price estimates for selected goods and properties, by directly scraping the information displayed on the web, instead of setting up specific surveys that can be quite time- and resource-intensive.

Second, big data can support **a timelier publication of official data**, by bridging the time lags before these statistics become available. In particular, the information generated instantaneously by the wide range of web and electronic devices – eg search queries – provides high-frequency indicators that can help current economic developments to be tracked more promptly (ie through the compilation of advance estimates). Indeed, another objective of the Billion Prices Project is to provide advance information on inflation in a large number of countries, including advanced economies, and with greater frequency – eg daily instead of monthly, as with a consumer price index (CPI). Turning to the real economy side, the real-time evolution of some "hard" indicators, such as GDP, can now be estimated in advance ("nowcast") by using web-based information combined with machine learning algorithms, as presented by Tugrul Vehbi (Reserve Bank of New Zealand) in the case of New Zealand.

[13] Web-scraping can be defined as the automated capturing of online information.

[14] Hill (2018) reports that 15% of the US CPI is now collected online.

[15] See www.thebillionpricesproject.com, and Cavallo and Rigobon (2016).

The high velocity of big data sources helps to provide more timely information, which can be particularly important during a crisis.

A third benefit is to provide **new types of statistics that complement "traditional" statistical data sets**, as emphasised in the presentation by Naruki Mori (Bank of Japan). Two important developments should be noted here. One is the increased availability of digitalised textual information, which allows the measurement of "soft" indicators such as economic agents' sentiment and expectations – derived, for instance, from social media posts. Traditional statistical surveys can also provide this kind of information, but they typically focus on specific items eg firms' production expectations and consumer confidence. In contrast, internet-based sources can cover a much wider range of topics. In addition, they are less intrusive than face-to-face surveys, and may therefore better reflect true behaviours and thoughts. A second important element has been the increased use of large granular data sets to improve the compilation of macroeconomic aggregates, allowing for a better understanding of their dispersion (IFC (2016a)) – this type of distribution information is generally missing in the System of National Accounts (SNA; European Commission et al (2009)) framework.[16]

Macroeconomic forecasting with big data

Many central banks are already using big data sets for macroeconomic forecasting. Indeed, nowcasting applications as described above can be seen as a specific type of forecasting exercise. For instance, the presentation by Per Nymand-Andersen (ECB) showed how Google Trends data can be used to compile short-term projections of estimates of car sales in the euro area, with a lead time of several weeks over actual publication dates. Moreover, and as argued in the presentation by Alberto Urtasun (Bank of Spain), big data allows a wider range of indicators to be used for forecasting headline indicators – for instance Google Trends,[17] uncertainty measures such as the EPU index (see above), or credit card operations as well as more traditional indicators. The devil is in the details, though, and statisticians need to try several approaches. For instance, some indicators may work well in **nowcasting** GDP (ie its growth rate over the current quarter) but less so in forecasting its future evolution (say, GDP growth one year ahead). Another point is that the internet is not the sole source of indicators that can be used in this context; in fact, some web-based indicators may work less well in nowcasting/forecasting exercises than do traditional business confidence surveys.[18]

In view of these caveats, and considering the vast amount of data potentially available, it may be useful to follow a **structured process** when conducting such exercises. The presentation of Paphatsorn Sawaengsuksant (Bank of Thailand)

[16] Indeed the SNA highlights the importance of considering the skewed distribution of income and wealth across households but recognises that getting this information is "*not straightforward and not a standard part of the SNA*" (2008 SNA, #24.69). It also emphasises that "*there would be considerable analytical advantages in having microdatabases that are fully compatible with the corresponding macroeconomic accounts*" (2008 SNA, #1.59). An important recommendation of the second phase of the DGI aims at addressing these issues (FSB-IMF (2015)).

[17] See https://trends.google.com/trends/. Google Trends provide indexes of the number of Google searches of given keywords. The indexes can be further segregated based on countries and provinces.

[18] For the use of nowcasting in forecasting "bridge models" using traditional statistics and confidence surveys, see Carnot et al (2011).

recommended a systematic approach when selecting the indicators of interest such as internet search queries. For instance, key words in Google Trends data could be selected if they satisfied several criteria, depending on their degree of generality, their popularity (ie number of searches recorded), their robustness (ie sensitivity to small semantic changes), their predictive value (ie correlation with macro indicators), and whether the relationship being tested makes sense from an economic perspective.

Financial market forecasting and monitoring

As in the macroeconomic arena, big data analytics have also proved useful in monitoring and forecasting financial market developments, a key area for central banks. A number of projects in this area facilitate the processing of huge volume of **quantitative information** in large financial data sets. For instance, the presentation by Tom Fong (Hong Kong Monetary Authority) showed that returns in a number of emerging sovereign bond markets can be predicted using various technical trading rules and machine learning techniques, to assess their robustness as well as the relative contributions of specific foreign (eg US monetary policy) and domestic factors.

Other types of project are looking at **less structured data**. For instance, the presentation by Okiriza Wibisono (Bank Indonesia) described how a text-mining algorithm could be used to measure public expectations for the direction of interest rates in Indonesia.[19] Specifically, a classification algorithm is trained to predict whether a given piece of text indicates an expectation for the future tightening, loosening or stability of the central bank policy rate. All the newspaper articles discussing potential developments in the policy rate from two weeks prior to monthly policy meetings are collected, and an index of policy rate expectation is produced. This index has facilitated the analysis of the formation of policy rate expectations, usefully complementing other sources (eg Bloomberg surveys of market participants). Similarly, the presentation by Stephen Hansen (University of Oxford) looked at the information content of news articles grouped into several categories using an "emotion dictionary sample", to predict equity returns. Other types of textual information, such as social media posts and official public statements, could also be usefully considered.

Experience reported by several central banks shows that new big data sources can also help to **elucidate** developments in financial markets, and shed light on their potential future direction. As regards the Bank of Japan, the use of high-frequency "tick data" has facilitated the assessment of market liquidity in the government bond market, and hence the risk of potential abrupt price changes. Similarly, the Bank of England has set up specific projects to identify forex market dynamics and liquidity at times of large market movements – eg when the Swiss National Bank decided to remove the EUR/CHF floor in January 2015 (Cielinska et al (2017)).

Financial risk assessment

Big data sources and techniques can also facilitate financial risk assessment and surveillance exercises that sit at the core of central banks' mandates – for both those in charge of micro financial supervision and those focusing mainly on financial

[19] See Zulen and Wibisono (2018).

stability issues and macro financial supervision (Tissot (2019)). In particular, the development of big data analytics has opened promising avenues for using the vast amounts of information entailed in granular financial data sets to assess financial risks.

To start with, they allow **new types of indicator** to be derived, as highlighted by the work presented by Vasilis Siakoulis (Bank of Greece) on the analysis of the financial strength of individual firms. Based on the granular information collected in the central bank's supervisory database (covering around 200,000 borrowers over a decade), a deep learning technique with a specific classification algorithm[20] was used to forecast the default for each loan outstanding. To facilitate policy monitoring work, this approach was complemented by a dimensionality reduction algorithm to reduce the number of variables to be considered.

Moreover, big data analytics can help to **enhance existing financial sector assessment processes**, by extending conventional methodologies and providing additional insights – in terms of eg financial sentiment analysis, early warning systems, stress-test exercises and network analysis. For instance, Sanjiv Das (Santa Clara University) and Kimmo Soramäki (Financial Network Analytics) presented a number of cases, including the use of network analytics for systemic risk measurement; the application of text analysis techniques to corporate e-mails and news for risk assessment; the measuring of interconnectedness to identify risk concentrations in CCPs and contagion effects; the identification of liquidity and solvency problems in payment systems; and the simulation of a financial institution's operational failure. These various experiences underlined the importance of having a sound theoretical framework to interpret the signals provided by disparate sources as well as to detect unusual, odd patterns in the data. They also highlighted the role played by model simplicity and transparency in the success of these initiatives, the benefit of a multidisciplinary approach, and the high IT and staff costs involved.

3. The use of big data in crafting central bank policy: organisation and challenges

Central bank experience suggests that the opportunities provided by big data sources and related analytical techniques can be significant, supporting a wide range of areas of policy interest. But how should central banks organise themselves to make the most of these opportunities? And what are the key challenges?

Organisational issues

Central banks' tasks cover a **wide range of topics** that can greatly benefit from big data. For instance, and as noted by Per Nymand-Andersen (ECB), central bankers need near-real-time and higher-frequency snapshots of the macro economy's state, its potential evolution (central scenario), and the risks associated with this outlook (eg early warning indicators and assessment of turning points). At the same time, their interest in financial stability issues calls for the ability to zoom in and get insights at

[20] eXtreme Gradient Boosting (XGBoost), which is commonly used in decision tree-based algorithms; see Chen and Guestrin (2016) and https://xgboost.readthedocs.io/en/latest/.

the micro level – see the ongoing initiatives among European central banks to develop very granular data sets on security-by-security issuance and holdings as well as on loan-by-loan transactions (the AnaCredit project; Schubert (2016)).

This puts a premium on **information systems** that can support this diversity of approaches. The presentation by Renaud Lacroix (Bank of France) argued that this requires a multidisciplinary and granular data platform, to supply flexible and innovative services to a wide range of internal users. The data lake platform project being developed at the French central bank will provide key data management services underlying multiple activities, covering data collection, supply (access), quality management, storage, sharing, analytics and dissemination. As presented by Robert Kirchner, the Deutsche Bundesbank has set up an integrated microdata-based information and analysis system (IMIDIAS) to facilitate the handling of granular data used to support its activities. It has also worked on fostering internal as well as external research on this information to gain new insights and facilitate policy analysis. Moreover, the Bundesbank actively supports the International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) (Bender et al (2018)).

A key takeaway was that the development of an adequate information system is only one element of a more **comprehensive strategy** to make the most of big data at central banks. As presented by David Roi Hardoon (Monetary Authority of Singapore), the "MAS Digital Supervision" initiative relies heavily on the use of machine learning, text-mining, natural language processing and visualisation techniques; and it is also backed by an extensive staff training programme on data analytics. Another example provided by Iman van Lelyveld relates to data science at the Netherlands Bank. Several use cases have been developed there – eg in the areas of credit risk, contagion risk, CCP risk, and stress testing in specific market segments. An important outcome has been the recognition of the important role played not only by the techniques used but also by the staff, organisation and culture.

Challenges and limitations

In practice, important challenges remain, especially in handling and using big data sources and techniques.

Handling big data can be resource-intensive, especially in collecting and accessing the information, which can require new, expensive IT equipment as well as state-of-the-art data security. Staff costs should not be underestimated too, as suggested by the experience reported for the Bank of Japan. First, large micro data sets on financial transactions often have to be corrected for false attributes, missing points, outliers etc (Cagala (2017)); this cleaning work may often require the bulk of the time of the statisticians working with these data. Second, a much wider set of profiles – eg statisticians, IT specialists, data scientists and also lawyers – are needed to work in big data multidisciplinary teams; ensuring a balanced skillset and working culture may be challenging. Third, there is a risk of a "war for talent" when attracting the right candidates, especially vis-à-vis private sector firms that are heavily investing in big data; public compensation and career systems may be less than ideally calibrated for this competition. In addition, and as seen above, a key organisational consideration is how to integrate the data collected into a coherent and comprehensive information model. The challenge for central bank statisticians is thus to make the best use of available data that were not originally designed for specific

statistical purposes and can be overwhelming (with the risk of too much "fat data", and too little valuable information). In most cases, this requires significant preparatory work and sound data governance principles, covering data quality management processes (eg deletion of redundant information), the setup of adequate documentation (eg metadata), and the allocation of clear responsibilities (eg "who does what for what purpose") and controls.

Using big data is also challenging for public authorities. One key limitation relates to the underlying quality of the information as noted above. This challenge can be reinforced by the large variety of big data formats, especially when the information collected is not well structured. Moreover, big data analytics rely frequently on correlation analysis, which can reflect coincidence as well as causality patterns.[21] Furthermore, the veracity of the information collected may prove insufficient. Big data sets may often cover entire populations, so by construction there is little sampling error to correct for, unlike with traditional statistical surveys. But a common public misperception is that, because big data sets are extremely large, they are automatically representative of the true population of interest. Yet this is not guaranteed, and in fact the composition bias can be quite significant, in particular as compared with much smaller traditional probabilistic samples (Meng (2014)). For example, when measuring prices online, one must realise that not all transactions are conducted on the internet. The measurement bias can be problematic if online prices are significantly different from the prices observed in physical stores, or if the products consumers buy online are different to those they buy offline.

These challenges are reinforced by two distinctive features of central banks – the first being their independence and the importance they accord to preserving public trust. Since the **quality** of big data sets may not be at the standard required for official statistics, "misusing" them as the basis of policy actions could raise ethical, reputational as well as efficiency issues. Similarly, if the **confidentiality** of the data analysed is not carefully protected, this could undermine public confidence, in turn calling into question the authorities' competence in collecting, processing and disseminating information derived from big data as well as in taking policy decisions inferred from such data. This implies that central banks would generally seek to provide reassurance that data are used only for appropriate reasons, that only a limited number of staff can access them, and that they are stored securely. The ongoing push to access more detailed data (often down to individual transaction level) reinforces the need for careful consideration of the need to safeguard the privacy of the individuals and firms involved.

A second feature is that central banks are policymaking institutions whose actions are influencing the financial system and thereby the information collected on it. Hence, there is a feedback loop between the financial big data collected, its use for designing policy measures, and the actions taken by market participants in response. As a result, any move to measure a phenomenon can lead to a change in the underlying reality, underscoring the relevance of the famous Lucas critique for policymakers (Lucas (1976)).

[21] See the words of Stephen Jay Gould as quoted in Per Nymand-Andersen's presentation: "*the invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning*".

Looking forward

The workshop and seminar provided a unique opportunity to take stock of the implementation of big data projects in the central bank community. These show that new big data-related sources of information and analytical techniques can provide various benefits for policymakers. Yet big data is still seen as complementing rather than replacing present statistical frameworks. It raises a number of difficult challenges, not least in terms of accuracy, transparency, confidentiality and ethical considerations. These limitations apply to big data sources as well as to the techniques that are being developed for their analysis. In particular, one major drawback of big data analytics is their black-box character, a difficulty reinforced by the frequent use of fancy names even for simple things ("buzzwords"). This can be a challenge for policymakers who need to communicate the rationale behind their analysis and decisions as transparently as possible. Moreover, **important uncertainties** remain on a number of aspects related to information technology and infrastructures, such as the potential use of cloud-based services and the development of new processes (eg cryptography, anonymisation techniques) to facilitate the use of micro-level data without compromising confidentiality.

One important point when discussing these issues is that **central banks do not work in isolation**. They need to explain to the general public how the new data can be used for crafting better policies, say, by providing new insights into the functioning of the financial system, clarifying its changing structure, improving policy design, and evaluating the result of policy actions (Bholat (2015)). But they also need to transparently recognise the associated risks, and to clearly state the safeguards provided in terms of confidentiality protection, access rights and data governance. Ideally, if big data is to be used for policymaking, the same quality of standards and frameworks that relate to traditional official statistics should be applied, such as transparency of sources, methodology, reliability and consistency over time. This will be key to facilitating a greater use of this new information as well as its effective sharing between public bodies.[22]

Looking forward, it is still unclear whether and how far big data developments will trigger a change in the **business models of central banks**, given that they are relatively new to exploiting this type of information and techniques. As noted in the presentation by Bruno Tissot (BIS), central banks have historically focused more on analysing data and less on compiling them. They are now increasingly engaged in statistical activities, reflecting the data collections initiated after the GFC as well as the growing importance of financial channels in economic activities – see, for instance, the substantial involvement of central banks in the compilation of financial accounts, a key element of the SNA framework (van de Ven and Fano (2017)). As both data users and data producers, they are therefore in an ideal position to ensure that big data can be transformed into useful information in support of policy.

[22] On the general data-sharing issues faced by central banks, see IFC (2016b).

References

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

Bean, C (2016): *Independent review of UK economic statistics*, March.

Bender, S, C Hirsch, R Kirchner, O Bover, M Ortega, G D'Alessio, L Teles Dias, P Guimarães, R Lacroix, M Lyon and E Witt (2016): "INEXDA – the Granular Data Network", *IFC Working Papers*, no 18, October.

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, April.

Bholat, D (2015): "Big data and central banks", Bank of England, *Quarterly Bulletin*, March.

Blei, D, A Ng and M Jordan (2003): "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, pp 993–1022.

Cagala, T (2017): "Improving data quality and closing data gaps with machine learning", *IFC Bulletin*, no 46, December.

Carnot, N, V Koen and B Tissot (2011): *Economic Forecasting and Policy*, second edition, Palgrave Macmillan.

Cavallo, A and R Rigobon (2016): "The Billion Prices Project: Using online prices for measurement and research", *Journal of Economic Perspectives*, Spring 2016, vol 30, no 2, pp 151–78.

Chen, T and C Guestrin (2016): *Xgboost: A scalable tree boosting system*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 785–94.

Cielinska, O, A Joseph, U Shreyas, J Tanner and M Vasios (2017): ""Gauging market dynamics using trade repository data: the case of the Swiss franc de-pegging", Bank of England, *Financial Stability Papers*, no 41, January.

Coeuré, B (2017): "Policy analysis with big data", speech at the conference on "Economic and financial regulation in the era of big data", Bank of France, Paris, November.

Committee on Payments and Market Infrastructures (CPMI) and Board of the International Organization of Securities Commissions (IOSCO) (2018): *Framework for supervisory stress testing of central counterparties (CCPs)*, April.

European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development, United Nations and World Bank (2009): *System of National Accounts 2008*.

Financial Stability Board (2011), "Understanding financial linkages: a common data template for global systemically important banks", *FSB Consultation Papers*.

——— (2017): *Artificial intelligence and machine learning in financial services - Market developments and financial stability implications*, November.

Financial Stability Board and International Monetary Fund (2009): *The financial crisis and information gaps*.

——— (2015): *The financial crisis and information gaps – Sixth Implementation Progress Report of the G20 Data Gaps Initiative*.

Groves, R (2011): "Designed data and organic data", in the Director's Blog of the US Census Bureau, www.census.gov/newsroom/blogs/director/2011/05/designed-data-and-organic-data.html.

Haldane, A (2018): "Will big data keep its promise?", speech at the Bank of England Data Analytics for Finance and Macro Research Centre, King's Business School, 19 April.

Hammer, C, D Kostroch, G Quirós and staff of the IMF Statistics Department (STA) Internal Group (2017): "Big data: potential, challenges, and statistical implications", *IMF Staff Discussion Notes*, no 17/06, September.

Hansen, S (2018): "Introduction to text mining", lecture delivered at the workshop on "*Big data for central bank policies*", Bank Indonesia, Bali, 23–25 July.

Hill, S (2018): "The big data revolution in economic statistics: waiting for Godot ... and government funding", *Goldman Sachs US Economics Analyst*, 6 May.

Irving Fisher Committee on Central Bank Statistics (IFC) (2015): *Central banks' use of and interest in 'big data'*, IFC Report, October.

——— (2016a): "Combining micro and macro statistical data for financial stability analysis", *IFC Bulletin*, no 41, May.

——— (2016b): *The sharing of micro data – a central bank perspective*, IFC Report, December.

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

——— (2018): *Central banks and trade repositories derivatives data*, IFC Report, October.

Laney, D (2001): "3D data management: controlling data volume, velocity, and variety", META Group (now Gartner).

Loughran, T and B McDonald (2011): "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks", *Journal of Finance*, vol 66, no 1, pp 35–65.

Lucas, R (1976): "Econometric policy evaluation: A critique", *Carnegie-Rochester Conference Series on Public Policy*, vol 1, no 1, pp 19–46.

Meeting of the Expert Group on International Statistical Classifications (2015): *Classification of Types of Big Data*, United Nations Department of Economic and Social Affairs, ESA/STAT/AC.289/26, May.

Meng, X (2014): "A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it)", in X Lin, C Genest, D Banks, G Molenberghs, D Scott and J-L Wang (eds), *Past, present, and future of statistical science*, Chapman and Hall, pp 537–62.

Nymand-Andersen, P (2016): "Big data – the hunt for timely insights and decision certainty: Central banking reflections on the use of big data for policy purposes", *IFC Working Papers*, no 14.

Petropoulos, A, V Siakoulis, E Stravroulakis and A Klamargias (2018): "A robust machine learning approach for credit risk analysis of large loan level data sets using deep learning and extreme gradient boosting", paper presented at the workshop on "*Big data for central bank policies*", Bank Indonesia, Bali, 23–25 July.

Rigobon, R (2018): "Promise: measuring from inflation to discrimination", presentation given at the workshop on "*Big data for central bank policies*", Bank Indonesia, Bali, 23–25 July.

Schubert, A (2016): "AnaCredit: banking with (pretty) big data", *Central Banking Focus Report*.

Tetlock, P (2007): "Giving content to investor sentiment: the role of media in the stock market", *Journal of Finance*, vol 62, no 3, pp 1139–68.

The Economist (2017): "The world's most valuable resource is no longer oil, but data", 6 May edition.

Tissot, B (2019): "Making the most of big data for financial stability purposes", in S Strydom and M Strydom (eds), *Big Data Governance and Perspectives in Knowledge Management*, IGI Global, pp 1–24.

van de Ven, P and D Fano (2017): *Understanding Financial Accounts*, OECD Publishing, Paris.

Zulen, A and O Wibisono (2018): "Measuring stakeholders' expectations for the central bank's policy rate", paper presented at the workshop on "*Big data for central bank policies*", Bank Indonesia, Bali, 23–25 July.

International Seminar on Big Data "Building Pathways for Policy-Making with Big Data" Bali, 26 July 2018

Opening remarks by Erwin Rijanto, Deputy Governor, Bank Indonesia

Honorables:

- Bapak Fadhil Hasan, Ketua Badan Supervisi Bank Indonesia,
- Mr. Bruno Tissot, Head of IFC Secretariat, Bank for International Settlements,
- Mr. Gabriel Quiros, Statistics Department, International Monetary Fund,
- Our esteemed chairs and speakers,
- Distinguished guests, ladies, and gentlemen,

Good morning to all of you,

It is my pleasure to welcome you to the International Seminar on Big Data, "Building Pathways for Policy-Making with Big Data", co-organized with our wonderful colleagues at the Bank for International Settlements – Irving Fisher Committee for Central Bank Statistics.

I would like to express my sincere gratitude to our prominent speakers who have travelled from around the world to gather here in Bali and share their valuable knowledge and experience to us.

Let me begin by remarking that it would be somewhat understated to say that the Big Data revolution has just begun. It has been a decade since the Google Flu Trend was publicized, from which we learned that Internet search data could predict disease outbreak faster than careful analysis by experts. It has also been a decade since two brilliant fellows at MIT first launched The Billion Prices Project, a robust alternative to the long-established methodology for measuring consumer prices inflation using data scraped from online retailers' website. We are already quite far into the era of Big Data, and it is up to us to make the best use of it.

Although it was first developed and adopted by IT and digital companies, **Big Data tools and methodologies have made their way into public institutions as well.** As of July 2018, there are almost 200 projects listed in the Big Data Project Inventory compiled by the World Bank and the United Nations[1]. Very comprehensive though it is, I am sure the list is not complete and there are many other public sector Big Data projects that are being developed in countries all over the world.

[1] https://unstats.un.org/bigdata/inventory/

Central banks are not missing out from this surge in Big Data adoption as well. I am delighted to realize that the share of central banks who have incorporated Big Data analytics into their policy-making and supervisory processes has gone up significantly, from 30% in 2015[2] to almost 60% in 2017[3]. Even though Big Data implementation presents various challenges, as I will discuss later on, this wider adoption shows that Big Data has proven its merits for us policy-makers.

Distinguished Guests, Ladies, and Gentlemen,

Let us now explore ***why*** **institutions are racing to adopt Big Data** in the first place, mentioning relevant examples along the way.

We observe **that there are at least three key factors that drive the widespread adoption of Big Data. The first and perhaps most significant one is the ubiquitous recording of our activities in digital format.** This is closely related to the increased mobile phone and Internet penetration and the recent trend towards digital commerce and interactions. With the cost of data storage and computing power continuously declining over the past decade, thanks in part to the development of cloud services, it has now become very in-expensive for companies and institutions to log and store the data of all their transactions and activities. These massive and granular sets of records provide a gold-mine equivalent for decision-makers from which they can draw insights and base decisions. E-commerce, fintech, social media, and all kinds of companies and institutions play an active part in producing, distributing, and consuming this explosion of digital data.

The next important factor that drives Big Data adoption is the various shifts in data analytical paradigms. Conventional data and econometric analysis have usually been applied on aggregated datasets and time series. Now, we are seeing some major change **from aggregated analysis towards analytical methods that depend on granular/large datasets,** often individual, transaction-by-transaction, or tick-by-tick data. Examples would be the Billion Prices Project that I mentioned earlier, and the AnaCredit[4] initiative by the European Central Bank, which allows individualized analysis of lending trends and behavior.

The high level of granularity in the data allows us to reveal interesting patterns and behaviors of economic agents. One of the source that readily accessible to central banks is the payment system settlement data. By applying network analysis and entity-matching algorithms, we are able to discover core-periphery structure in interbank payments from our own Real-Time Gross Settlement system. In addition, we can identify flows of funds both within and between groups of corporations. The overall analysis opens up new opportunities for systemic risk assessment.

Besides aggregate-to-granular analysis, we are also seeing **more applications of predictive analytics in addition to descriptive statistics.** Big Data offers added-value that allow us to base our predictions on richer, more granular and more varied types of data, including unstructured data such as texts and images. Big Data is also

2 IFC Report on "Central banks' use of and interest in big data", 2015

3 Central Banking and BearingPoint's Joint Survey on "Big Data in Central Banks", 2017

4 https://www.ecb.europa.eu/stats/money_credit_banking/anacredit/html/index.en.html

well-suited for ***nowcasting***, "predicting" data for the current time period using alternative datasets. Big Data[5] can help alleviate the problem of **data lag** through higher-frequency collection of granular, publicly accessible datasets and subsequent processing on top of streaming and other real-time Big Data technologies. **Bank Indonesia is also developing and has benefited from this approach for measuring the trends in job vacancy, secondary property, and used-car markets,** for which official statistics are published with long time lags or are not available with the desired level of detail.

Finally, it feels incomplete to talk about reasons for Big Data adoption without alluding to **Artificial Intelligence and machine learning**. The recent development and subsequent boom of specialized machine learning algorithms called deep learning[6] have allowed computers to see and discern images and videos, understand and generate human-language texts and speeches, drive cars and control robots, and perform a plethora of other tasks nearly as well as human do. With its potential of automating manual human labors, machine learning will certainly impact employment and the general economy in the years to come.

Ladies and Gentlemen,

How could central banks join the broader industry in this issue? One example would be the application of **text mining**: the automated analysis of text data.

Perhaps the most widely cited and widely replicated application of text mining for economic analysis is the **Economic Policy Uncertainty Index[7]**. It has long been believed that high policy uncertainty undermines macroeconomic performance. The index developed by Baker, Bloom and Davis (2016) provides an important advancement in the area of measuring policy uncertainty. Through machine-reading of newspaper articles, an index that measures policy-related economic uncertainty can be constructed. This new approach has been followed by many central banks, including Bank Indonesia, and several robustness checks has shown that the index well correlated with important events that might affect this uncertainty, despite some shortcomings.

In recent years there were also some number of text mining researches that went into **understanding central banks' statements[8]**. As we know, these statements are an integral part of our policies and being able to extract information in a quantitative way may provide decision-makers and central banks themselves with useful insights on past and future policy decisions and their impacts on the economy.

Distinguished Guests, Ladies, and Gentlemen,

Let us now turn to building pathways for Big Data in our policy-making. There are numerous obstacles that hinder Big Data implementation in central banks and government institutions, as have been brought into attention in previous conferences and surveys. The most common challenges include lack of adequate support of Big

5 https://www.ecb.europa.eu/pub/conferences/html/20140407_workshop_on_using_big_data.en.html

6 https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning

7 https://academic.oup.com/qje/article-abstract/131/4/1593/2468873

8 http://sekhansen.github.io/pdf_files/fomc_transparency.pdf, https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2085.en.pdf

Data technology infrastructures and software, lack of capable human resources (data scientists) and dedicated Big Data unit, lack of procedures in place for ensuring data privacy, not to mention questionable data quality and the complex pre-processing required to clean such data. I will focus on one specific challenge: **how policy-makers should obtain access to Big Datasets.**

It is quite understandable that **institutions and companies are concerned and wish to keep their data private to themselves.** Data has been termed as the new oil in the 21st century, considering its ability to generate and maintain competitive advantage for its owners[9]. Absent overarching regulations that require other institutions and companies to report their proprietary data, government institutions are ill-positioned to obtain most Big Datasets relevant for analysis and policy-making, since they are not themselves producers of such datasets.

As policy-makers, it becomes our task to explain to the greater public that access to Big Datasets are aimed for and only for crafting better policies, and to make good on this promise. A clear regulation in conjunction with clear objectives should help convince the public of the necessity for government access to Big Datasets and elicit more cooperation and compliance.

Central banks, statistics offices, and government institutions should also establish a close coordination for data sharing[10]. It is very burdensome to the public if they are required to report the same datasets multiple times, especially if different government institutions each require different specifications for the reports. This is even more relevant in the case of Big Data, since transferring voluminous amount of data in near real-time to multiple institutions will be impractical, if not impossible. Depending on the cross-cutting of jurisdictions in the country, relevant datasets are also often captured in silos in different government institutions. **Thus, besides reducing reporting load, data sharing allows a more complete picture for analysis and policy-making.**

Of course, all data access and data sharing initiatives should be established with stringent privacy and confidentiality measures that protect both the data producers and the respective individuals. For example, reporting and sharing of identifiable information that are not needed for analysis, such as people's names and exact addresses, should be kept at a minimum and kept private. Sufficient IT standards that cover all relevant aspects of data privacy, including network and database security, should be implemented and duly observed. In addition, internal access to confidential Big Datasets should be assessed and authorized on a need-by-need basis.

Distinguished Guests, Ladies, and Gentlemen,

Allow me to conclude my remarks.

I have discussed what we consider to be the main factors driving Big Data development and its widespread adoption. Big Data analytics is made possible through massive and ubiquitous digital recording of our activities, complemented by the growth of computing power. Big Data can inform better decisions through

9 https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

10 https://www.bis.org/ifc/events/7ifc-tf-report-datasharing.pdf

analysis of richer, more granular data, in more timely manner. Recent progress in machine learning algorithms also allows more accurate analysis of unstructured datasets, such as texts and images.

I have also discussed the need for data access and data sharing for public institutions, along with several considerations that need to be taken into account in the implementation.

I am fully aware that it will not suffice to discuss all aspects of and trends in Big Data in a single speech nor a single day of seminar. I sincerely hope that we all have much to learn from this Seminar, and we can go back to our institutions with concrete strategies and improvements to be put in practice.

Finally, I would also like to mention that this event is held as part of the Voyage to Indonesia (VTI) series of activities that we prepare as the groundwork for the 2018 IMF-WBG Annual Meetings here in Bali. The theme Voyage to Indonesia reflect a journey that will bring the world to the renewed Indonesia; a reformed, resilient, and progressive economy. We hope that you will continue to be engaged in various VTI programs.

With that, allow me to declare the opening of this Seminar.

Thank you very much.

International Seminar on Big Data "Building Pathways for Policy-Making with Big Data" Bali, 26 July 2018

Welcoming remarks by Claudia Buch, IFC Chair and Vice President of the Deutsche Bundesbank

1. Welcoming address

Good morning ladies and gentlemen,

It is my great pleasure to welcome you on behalf of the Irving Fisher Committee to the second seminar on big data.

Big data provides a big opportunity for central banks both in terms of analytical work and statistical work. So I am very pleased that the Central Bank of Indonesia is hosting this event.

2. Reference to IFC work

The BIS All Governors meeting last year discussed big data and implications for Central Banks. Part of the discussion was based on the 2015 IFC survey.[1]

There are a couple of areas where central banks are interested in dealing with big data issues:

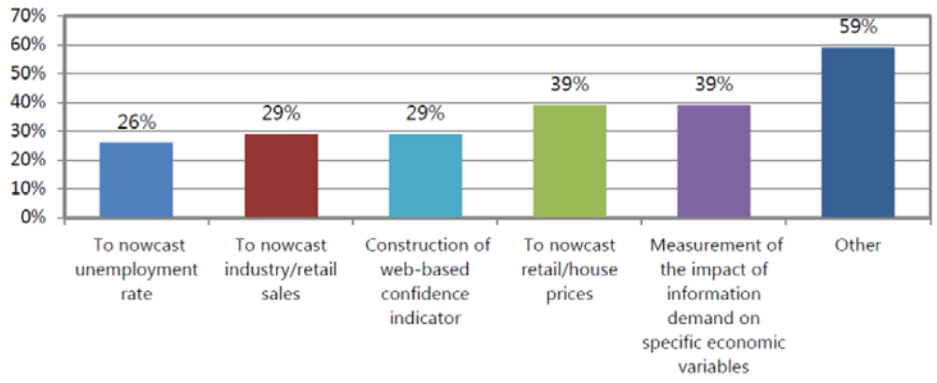
- The most important one for the statistics community is the collection and the provision of statistics.
- In addition, big data techniques are useful for analysis of monetary and financial market developments in line with central banks' mandates for price stability and financial stability.
- Many central banks are also involved in the supervision of the banking system and the financial system. This is another area where big data and new technologies can be useful.
- Central banks provide (financial) infrastructures and are big institutions which provide in-house services such as IT and human resources. All these are areas where we also rely on data, technologies, and where big data can be important.

So more specifically, what did the IFC survey tell us about interest of the central bank community in big data issues?

[1] See Irving Fisher Committee on Central Bank Statistics, "Central Banks use of and interest in big data", October 2015.

What kind of outcomes are you expecting as a result of exploring big data sets?

Note: multiple responses possible.



Roughly 40 % of central banks use big data sets to measure specific economic variables such as retail and house prices. And 60% of respondents use big data for other issues. This includes nowcasting, not only of prices but also economic activity, output, understanding credit risk, and understanding risks to financial stability. Another result of the survey was a high demand for qualified personnel to do the work and carry out the analysis.

3. All Governors' meeting, September 2017

In the BIS All Governors' meeting in September last year, Professor Roberto Rigobon (Massachusetts Institute of Technology, MIT) talked about his experience with big data and new technologies. One of the implications of his discussion was that we should start small. Even though we are talking about big data, we should start with small projects that we can implement and then go step-by-step. In addition, the importance of international cooperation and the exploitation of large administrative data sets have been stressed at the meeting.

4. What needs to be done?

In my opinion, central banks should also make better use of existing data infrastructures.

We should work hard on improving data sharing internationally and nationally. The recommendations on data sharing in the framework of the second phase of the Data Gaps Initiative needs to be implemented in practice. The G20 leaders, Financial

Ministers and Central Bank Governors welcomed these recommendations and are looking forward to receiving progress reports.[2]

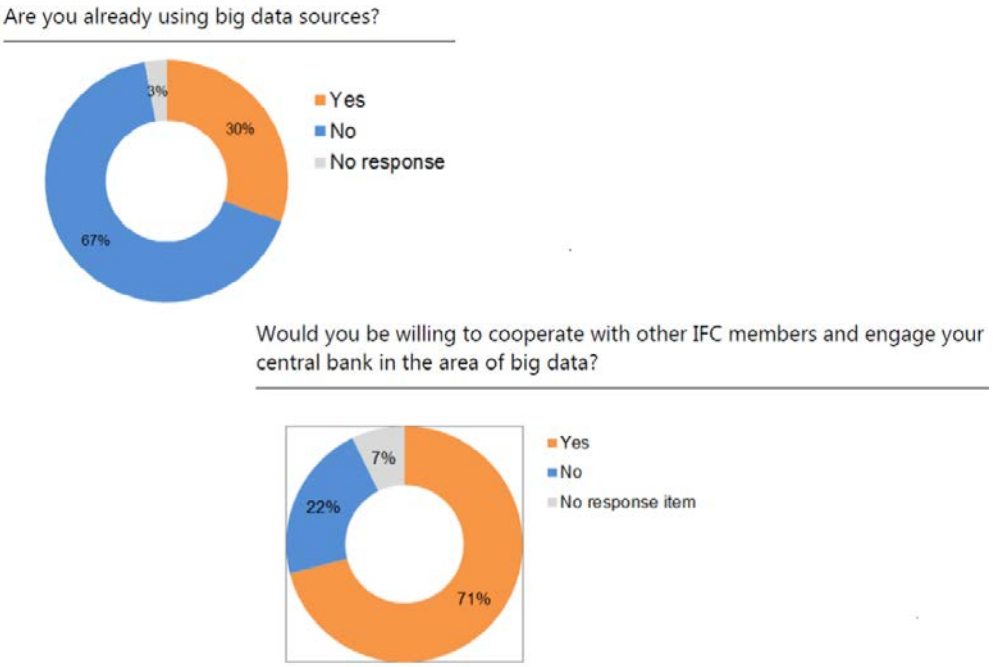
Of course, central banks have to develop and test new analytical tools for the big data sets, and eventually, share the experience with each other.

5. Closing

I would like to close with a last reference to the recent IFC big data survey.

IFC members were asked if they already use big data sources. 30 % of the participants answered "Yes", while 67 % answered "No", the rest gave no response.

The other question asked if participants are willing to cooperate with other IFC members and engage in the area of big data. This question was answered with "Yes" by 71 % and with "No" by 22 %, the rest gave no response.



So in that sense, I think this is exactly what the conference is supposed to do. It is hopefully a fruitful and stimulating exchange and I wish you the best of success in further shaping your ideas. Thank you very much for your attention.

[2] See G20 Hamburg Action Plan (2017). Retrieved 7 May 2019, from https://www.g20germany.de/Content/DE/_Anlagen/G7_G20/2017-g20-hamburg-action-plan-en__blob=publicationFile&v=4.pdf

References

Irving Fisher Committee on Central Bank Statistics, "Central Banks use of and interest in big data", October 2015.

Irving Fisher Committee on Central Bank Statistics, IFC Bulletin No. 44 on Big Data, September 2017.

G20 Hamburg Action Plan (2017). Retrieved 7 May 2019, from https://www.g20germany.de/Content/DE/_Anlagen/G7_G20/2017-g20-hamburg-action-plan-en__blob=publicationFile&v=4.pdf

International Workshop on "Big Data for Central Bank Policies"
Bali, 23 – 25 July 2018

Opening remarks by Yati Kurniati, Executive Director, Head of Statistics Department, Bank Indonesia

Honorables:

- Head of Statistics & Research Support and Head of Irving Fisher Committee Secretariat, Bank for International Settlements, Mr. Bruno Tissot,
- Our Distinguished speakers from the Bank of England, Santa Clara University, the University of Oxford, Massachusetts Institute of Technology, and Financial Network Analytics, and
- Distinguished representatives of all the participating countries from around the world.

Assalaamu'alaikum Wr. Wb.,

Peace be upon us, Om Swastiastu ….

Good Morning and Welcome to the Island of Gods - Bali, one of the most beautiful places on earth.

1. First of all, let us extend our praise to The God Almighty, since only with His permission and blessings we can all get together this morning to attend the *"International Workshop on Big Data for Central Bank Policies"*.
2. We are delighted to have another opportunity to collaborate with BIS, after the Satellite Seminar on Big Data, preceding the ISI Regional Statistics Conference, which was held in March 21st 2017, also in Bali. I believe this event will be as fruitful and productive as the last.
3. We are honored to have the contributions of our prominent instructors of this workshop: Mr. Bruno Tissot from BIS, Mr. Paul Robinson from the Bank of England, Mr. Sanjiv Das from Santa Clara University, Mr. Stephen Hansen from the University of Oxford, Mr. Roberto Rigobon from MIT, and Mr. Kimmo Soramaki from Financial Network Analytics. I hope you would enjoy teaching and speaking at this event, and we will do our best to learn as much as we can from you.
4. This workshop is attended by approximately 70 participants from 18 countries around the world; quite an international audience, I would say. You come from macroeconomic, monetary, supervisory, financial stability, research, and various other departments of your institutions. I am sure your diverse backgrounds will only enrich our discussions and we will have much to gain from each other's experience. I hope we can also establish productive relationships from this

workshop. A knowledge that you think is trivial could be huge and important for another institution.

Distinguished Ladies and Gentlemen,

5. I believe we all know or at least have heard about Big Data. It is a very popular topic in recent years, and it has gained significant traction both in the industry and in academia.

6. But is Big Data Analytics really relevant for our work as central bankers and government officials? Will it bring value to our current practices of policy-making and supervision? We believe so, and that is why Bank Indonesia decided to organize this important workshop on Big Data for policy-making.

7. As you will learn in the coming days, Big Data Analytics is not separate from "traditional" statistics, and indeed it borrows many long-established methodologies from statistics.

8. The one characteristic that certainly sets Big Data Analytics apart is its application on "modern" datasets. Today's digital technologies have resulted in data being produced in massive amounts, in real-time, in a variety of formats, by various institutions and individuals. Extracting relevant information from these sources is not straightforward and will require a distinct set of skills. The workshop curriculum that we have designed aims to introduce some of these concepts.

Distinguished Ladies and Gentlemen,

9. Let me briefly go through the workshop's sessions. The first session by Mr. Bruno Tissot will introduce the definition of Big Data, and provide examples of Big Data Analytics as well as the challenges in its implementation. For the second session we will have Mr. Paul Robinson, who will continue from Bruno's main points and present further examples of Big Data Analytics. Paul will also introduce us to several methodologies for Big Data Analytics.

10. The third session will be delivered by Mr. Sanjiv Das. It will cover some of the most popular Big Data Analytics algorithms. Although quite technical, the session nicely illustrates the analytical tools that we have for Big Data, and we will see how they differ from the usual statistics and econometrics methodologies.

11. Mr. Stephen Hansen's 3-hour session on the second day will focus on text mining: the set of methodologies for understanding texts written in human language. Text data, which include newspaper articles, official reports, and social media posts, contain a lot of information that may not be available elsewhere in conventional, structured data. Mr. Hansen will also teach one application of text mining that is very relevant to central banks, namely how we could measure people's opinion and sentiment that they express in text data.

12. Besides Mr. Stephen, we will also have Mr. Roberto Rigobon in the second day. He is widely known for his Billion Prices Project, and hopefully we can learn from his insightful experience about macroeconomic nowcasting and forecasting with Big Data.

13. On the third day Mr. Kimmo will present another family of Big Data Analytics: network analysis. He will discuss the networked structure of economic and financial activities, and how Big Data can help us identify prominent agents and patterns in such networks.

14. In this workshop we will also have the opportunity to learn specific research and applications of Big Data Analytics that you have implemented in your institutions. We will hear 7 exciting paper presentations on Tuesday and Wednesday, chaired by Paul and Roberto. We hope these will further exemplify Big Data Analytics for policy-making.

15. As you are already aware, this 3-day Workshop will be followed by the International Seminar on *"Building Pathways for Policy Making with Big Data"* on Thursday. This high-level seminar will host prominent chairs and speakers and will be attended by 200 audiences from various backgrounds, from central banks, public institutions, banks and financial institutions, other industries, and academics. It will discuss important insights and topics of Big Data Analytics for policy-making, including implementation challenges as well as the strategies adopted by central banks. We are also hosting a Gala Dinner for the seminar on Wednesday evening.

16. I am honored to invite you to attend and I sincerely hope you could participate in the seminar and the Gala Dinner.

Distinguished Ladies and Gentlemen,

17. We really hope that this program would be able to strategically contribute to our knowledge enhancement on Big Data, as well as to provide a strong basis for better policy-making going forward. I hope this program might reveal the best result for all of us. I am confident that your in-depth discussions and the outcomes of this workshop will further enable us to realize this workshop's objectives.

18. Before I end my remarks, on behalf of Bank Indonesia, I would like to thank BIS again for their strong support and collaboration in delivering the success of this event. Last but not least, I would like to say: "Have a nice workshop and I wish you all get fruitful days during this program. Have a pleasant stay in Bali."

And finally, allow me to declare the official opening of the workshop.

Thank you very much.

Wabillahi taufiq wal hidayah. Wassalamu alaikum Wr. Wb.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Understanding big data: fundamental concepts and framework[1]

Paul Robinson,
Bank of England

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANK OF ENGLAND

Understanding Big Data: Fundamental Concepts and Framework

International Workshop on Big Data for Central Bank Policies

Paul Robinson, Bank of England

23 July 2018

Outline

- What do we mean by ‘Big Data’?
- Several different dimensions that we can classify its use:
  - Different types of data
  - Different uses of the data sets
  - Different analytical techniques
- Are central bank needs’ different from other organisations?
- Lots of opportunities but also challenges

What do we mean by ‘Big Data’?

- First page of a Google search for “V’s of big data” included:
  - Infographic: The Four V's of Big Data | IBM Big Data & Analytics Hub
  - The 10 Vs of Big Data | Transforming Data with Intelligence
  - Understanding the 3 Vs of Big Data - Volume, Velocity and Variety
  - The 42 V's of Big Data and Data Science - Elder Research
  - The five V's of big data | BBVA
  - How many V's are in big data?

What do we mean by ‘Big Data’?

- First page of a Google search for “V’s of big data” included:
  - Infographic: The Four V's of Big Data | IBM Big Data & Analytics Hub
  - The 10 Vs of Big Data | Transforming Data with Intelligence
  - Understanding the 3 Vs of Big Data - Volume, Velocity and Variety
  - The 42 V's of Big Data and Data Science - Elder Research
  - The five V's of big data | BBVA
  - How many V's are in big data?

Different types of data

- Despite the confusion and hype the ‘V’s structure does offer a framework to consider the opportunities and challenges
- In particular, the following 5 ‘V’s set up is useful:
  - Volume
  - Velocity
  - Variety
  - Value
  - Veracity

What central banks do

- Regulate important institutions
  - Banks, insurance companies, FMIs, …
- Set policy
  - Monetary policy, macroprudential policy, microprudential policy
  - Engage in international policy setting
- Implement policy
  - Markets, PRA, …
- Run important functions
  - Payment systems, currency issuance …
  - Manage national reserves
  - Act as a bank to key institutions (eg the government)
- Run a large, (singular) institution

- Most central banks have similar responsibilities

How do central banks go about discharging these responsibilities?

- Understand the current situation
  - Combine information with an understanding of how it fits together
- Forecast what would happen holding policy unchanged
- Consider possible policy changes
- Model how they would affect the economy/financial system, …
- Set policy
- Monitor the effects of policy
  - Update our understanding of the current situation and the structure of the system

Why it's difficult

- Imperfect measurement
  - Noise, biases, blind spots, out of date information, (near) simultaneity of cause and effect
- "Too much" data, too little information
- Imperfect theory
- Complex, adaptive system with lots of feedback
  - Leads to "chaotic" behaviour
- Internal frictions

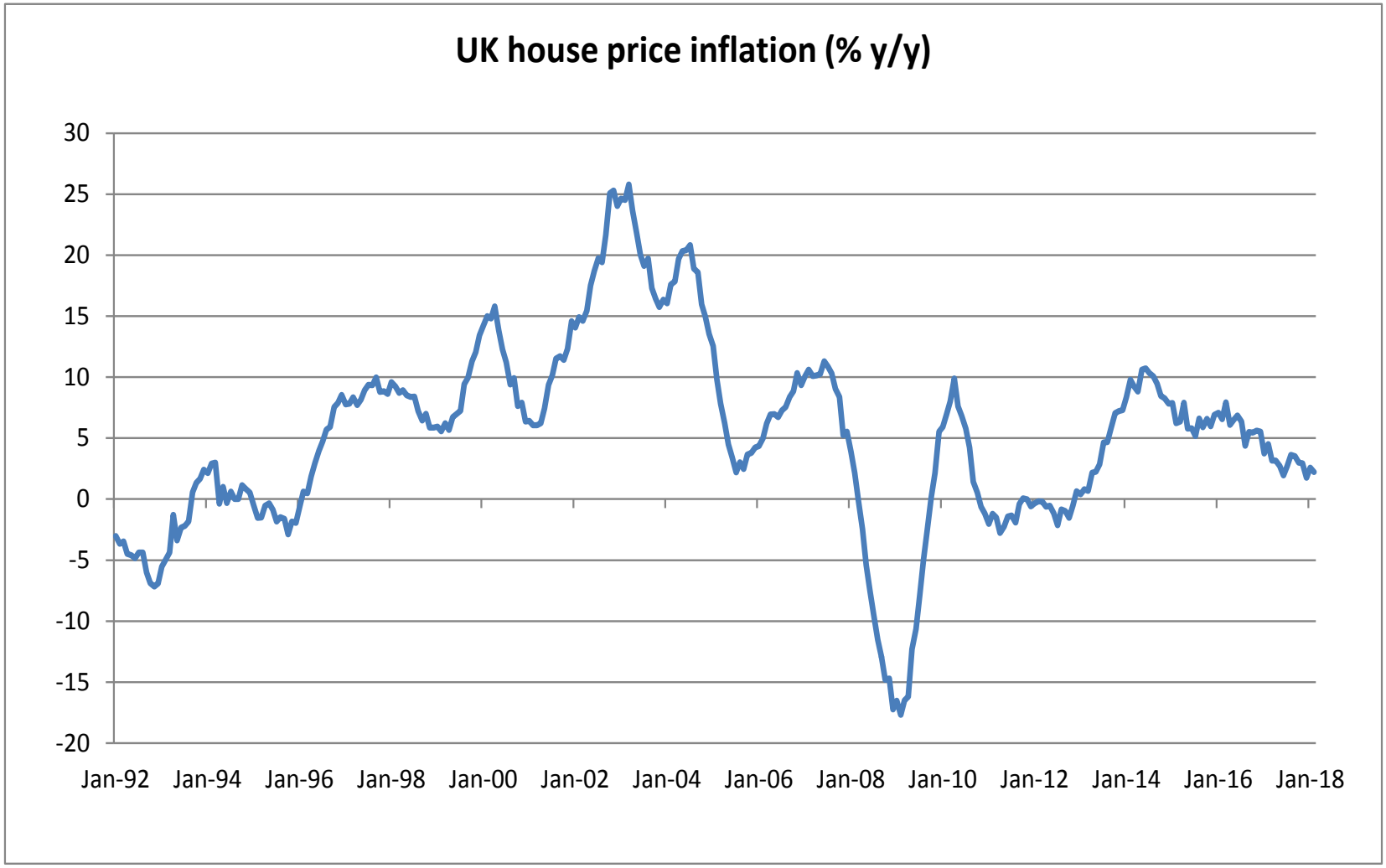
How can Big Data help?

- Imperfect measurement
  - Insight into previously hidden phenomena
  - Combining different types of data
  - Speed and completeness of coverage
- "Too much" data, too little information. Use data science methods to:
  - Improve processing large data sets
  - Help separate the signal from the noise
- Imperfect theory
  - Hypothesis generation
  - Alternative modelling approaches (eg Agent-based models)
- Complex, adaptive system with lots of feedback
  - Difficult to cope with, but more accurate understanding of initial conditions and more frequent updating help a lot
- Internal frictions
  - Improved management information

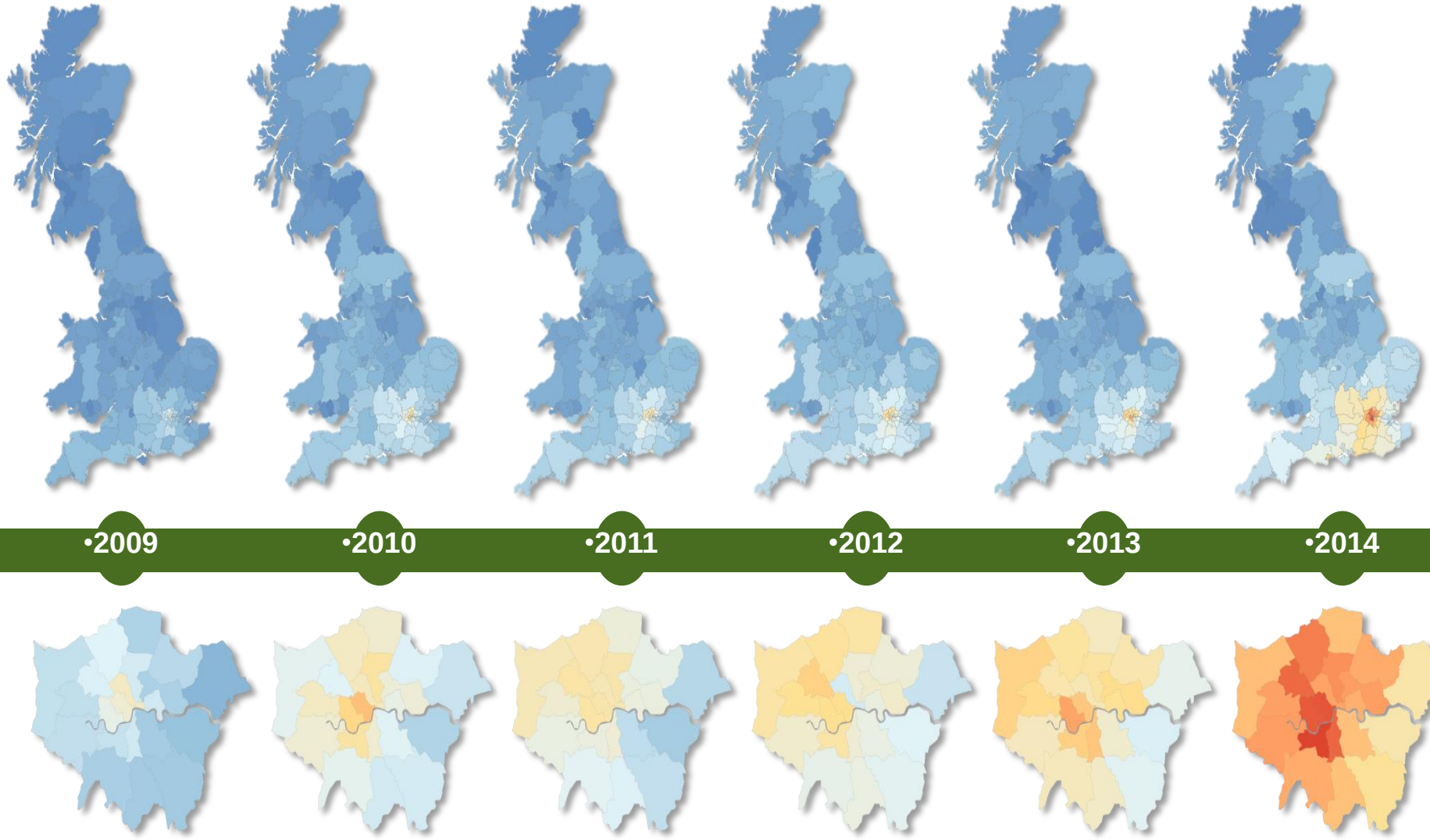
Big data sets offer significant potential advantages

- Greater **detail** (Volume, Velocity, Variety)
- Allow insights that aggregate numbers might obscure
- Examples:
  - UK housing market
  - Market dynamics around the abolition of the EUR/CHF floor
  - Market liquidity around large market moves

UK housing market

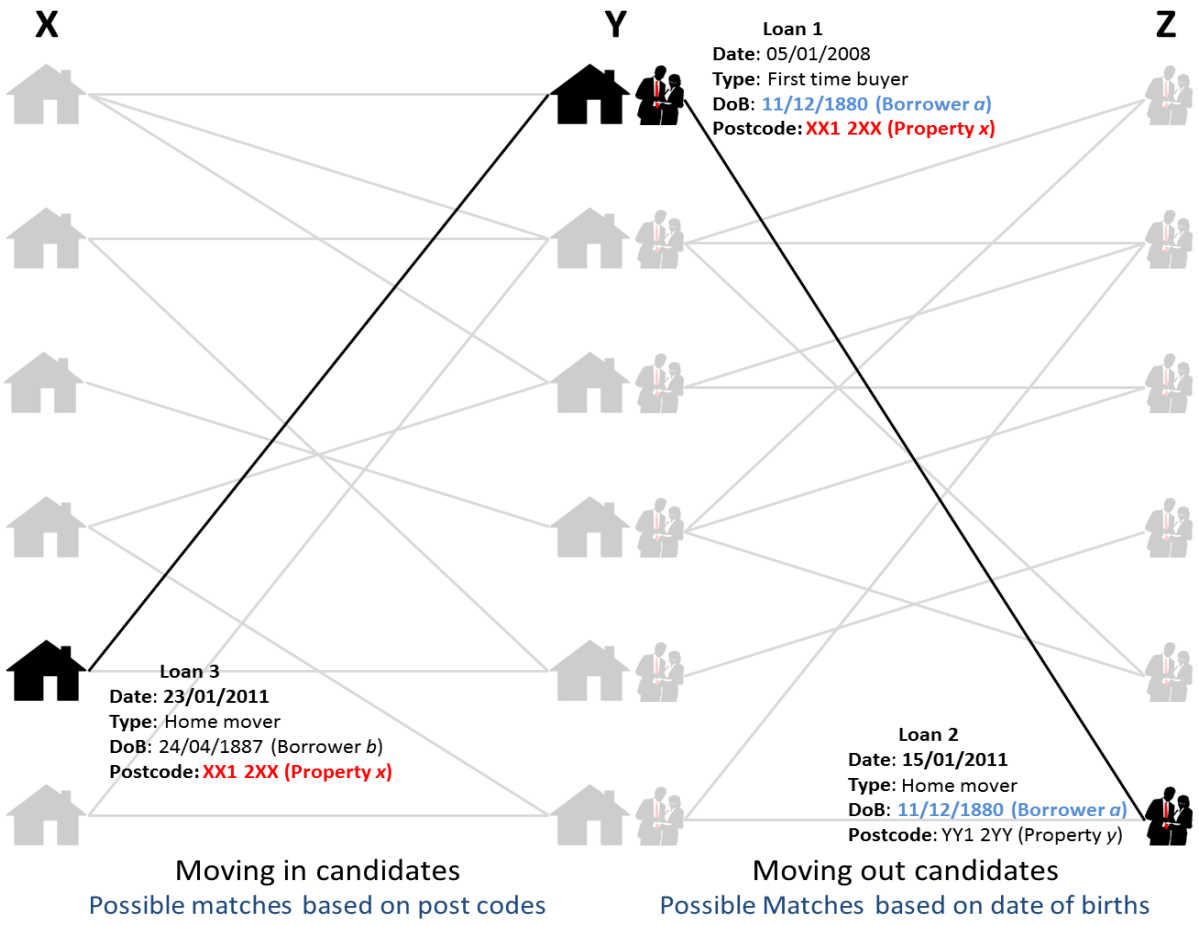


Source: Average of HBOS and Nationwide measures

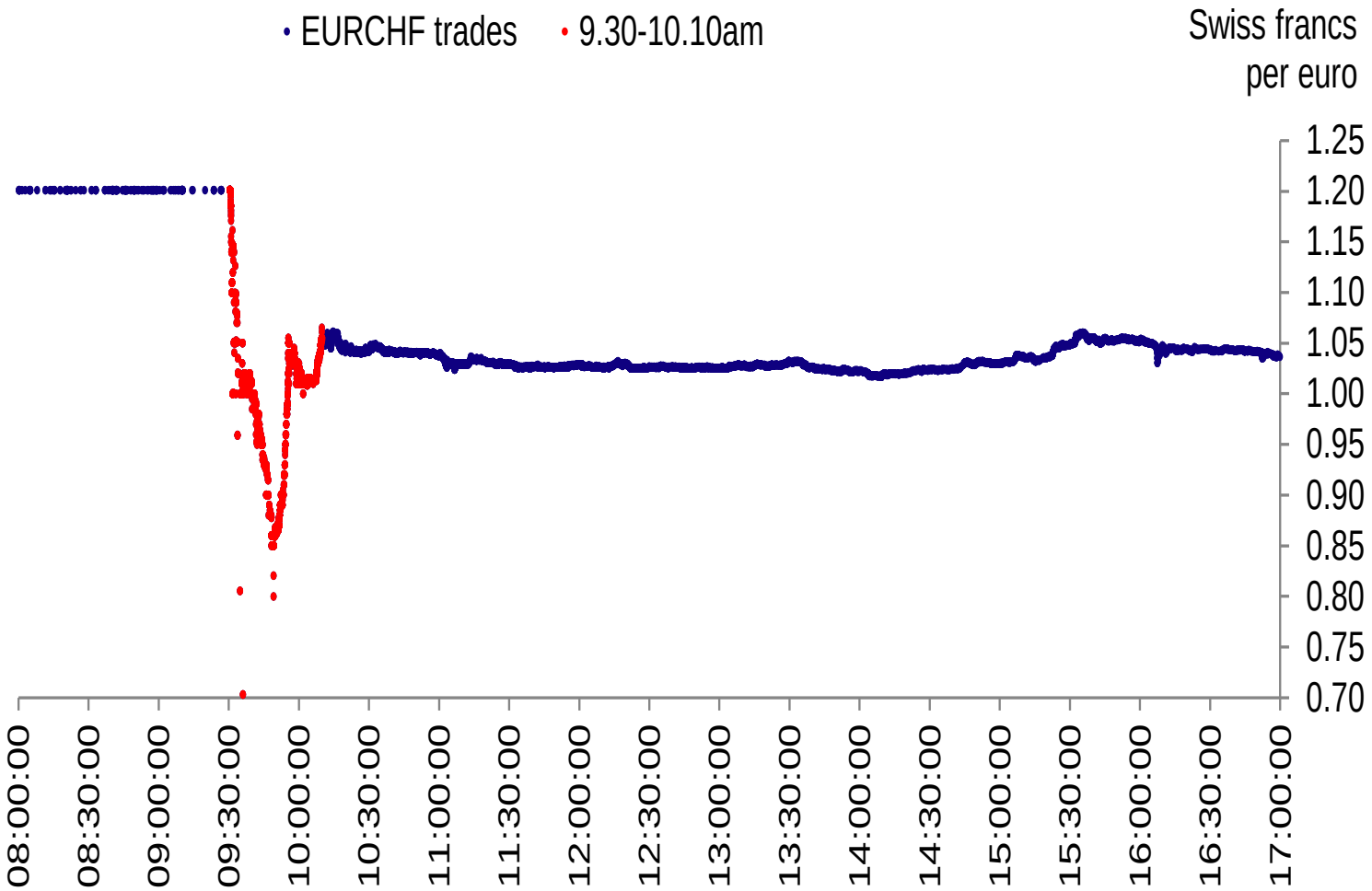


•**Key:** % of mortgages with loan more than 4.5x income

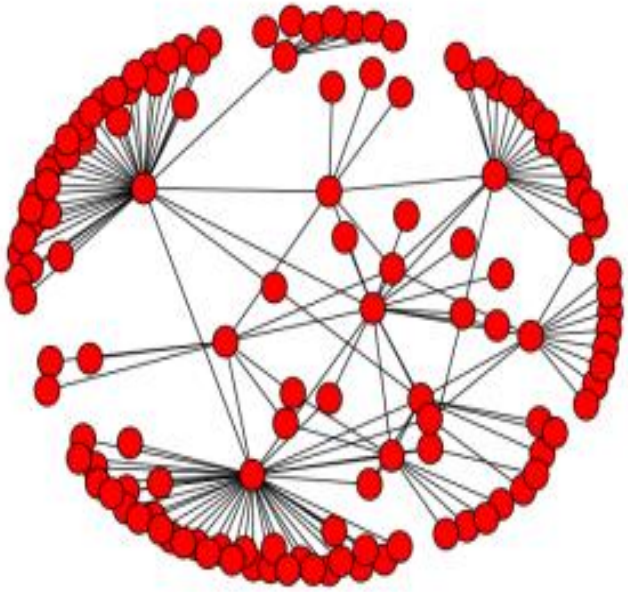
Tracking home movers



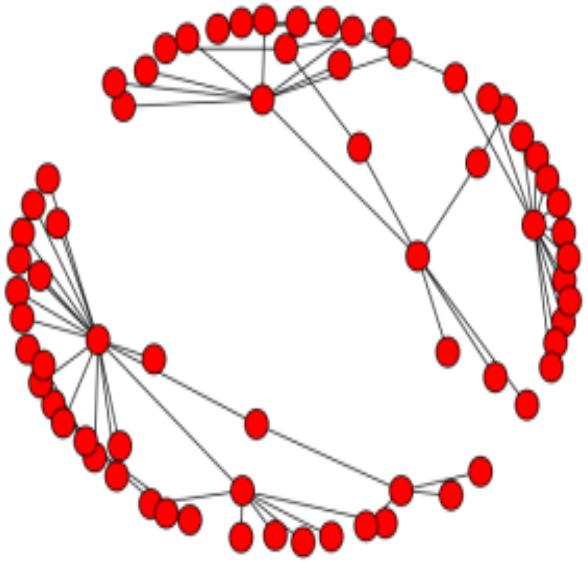
Large-scale data analysis



Network of CHF derivatives contracts

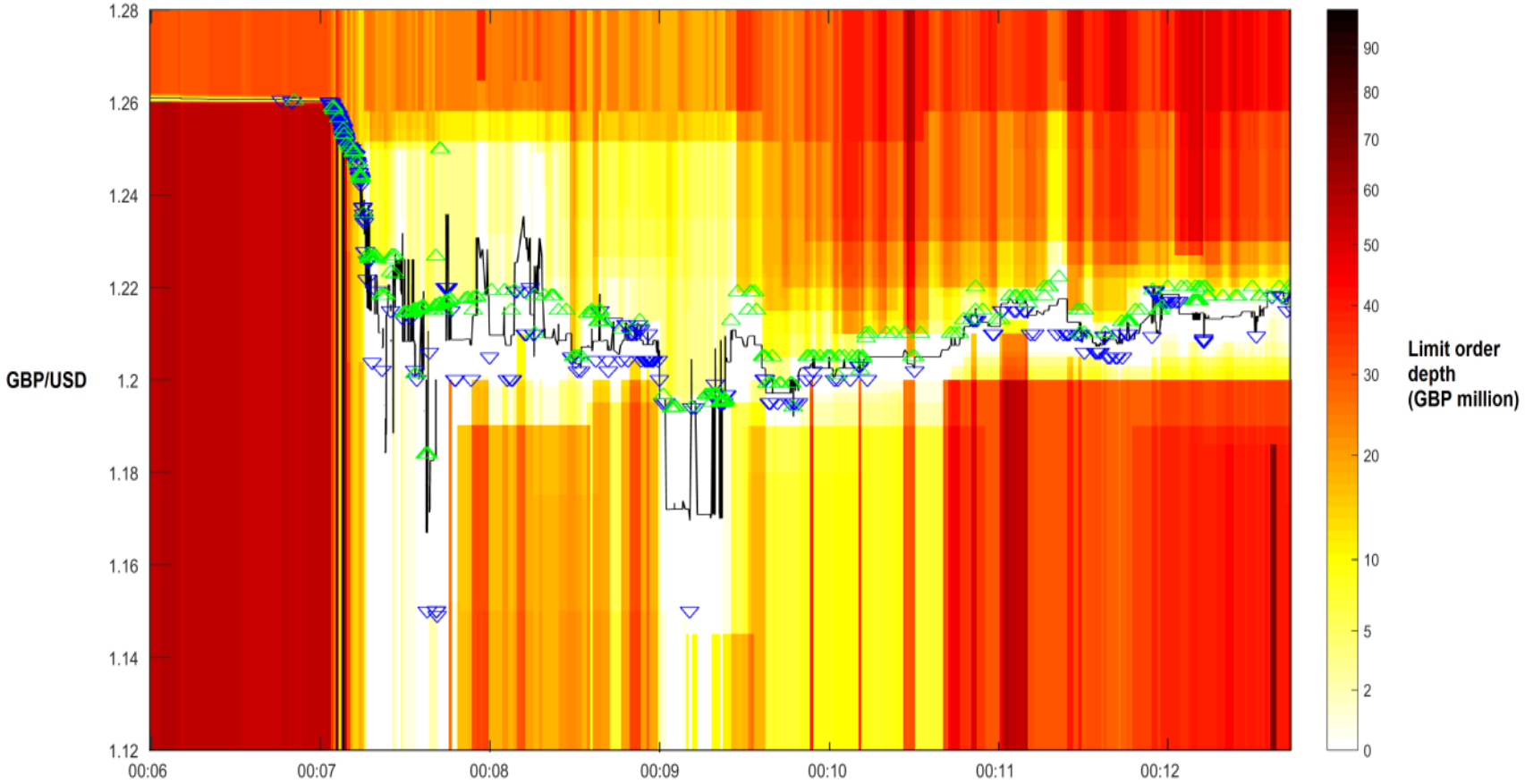


15 January 2015



22 January 2015

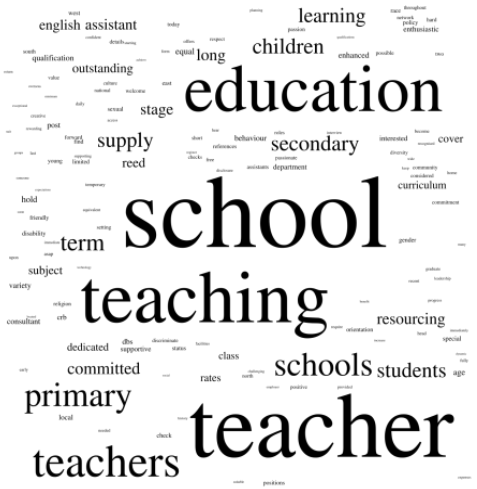
Market depth around sterling "flash crash" episode (7 Oct 2016)

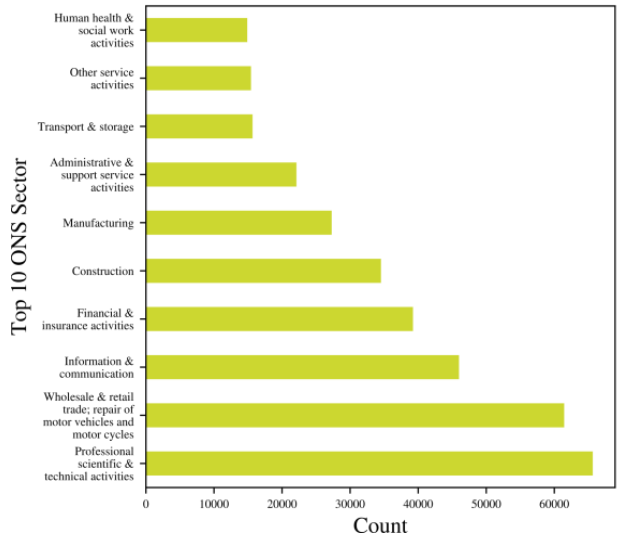


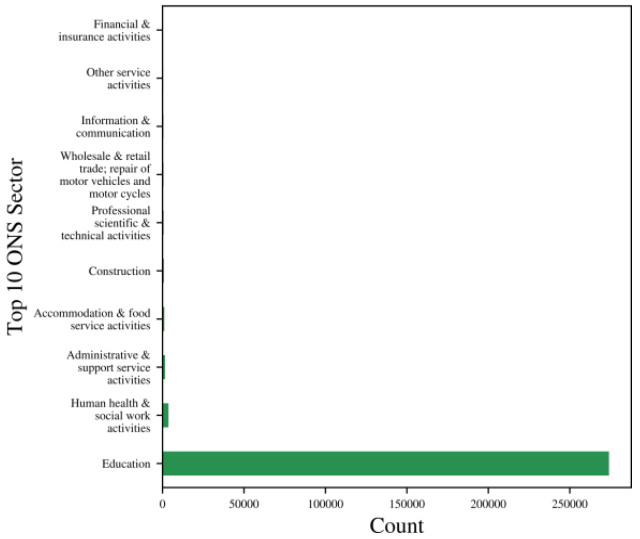
Big data sets offer significant potential advantages

- Greater **flexibility** (Velocity, Variety)
- Gives a window into changing structure of the economy
- Example:
  - Using job adverts to understand changing labour market dynamics

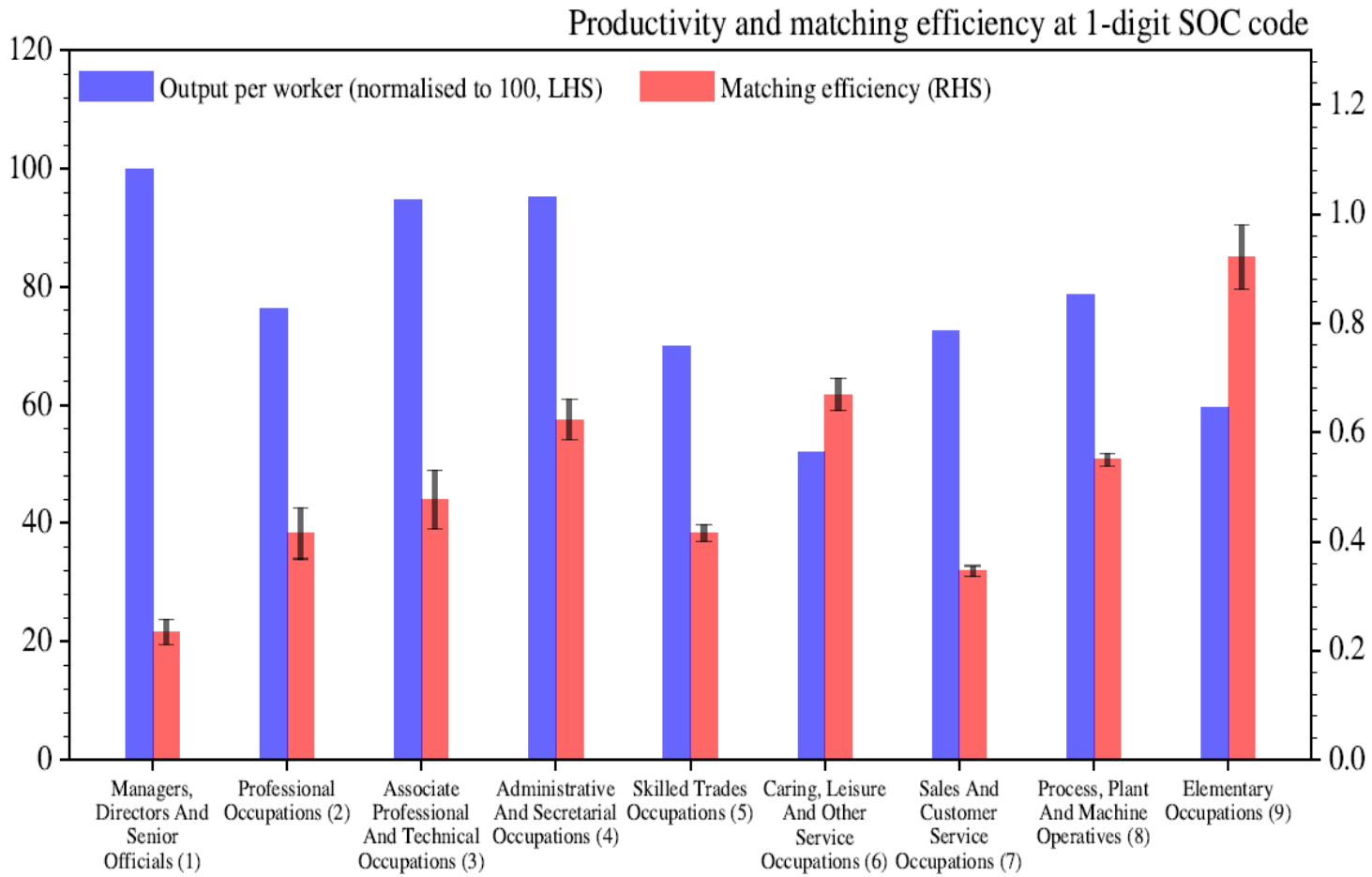
Understanding the labour market using job ads







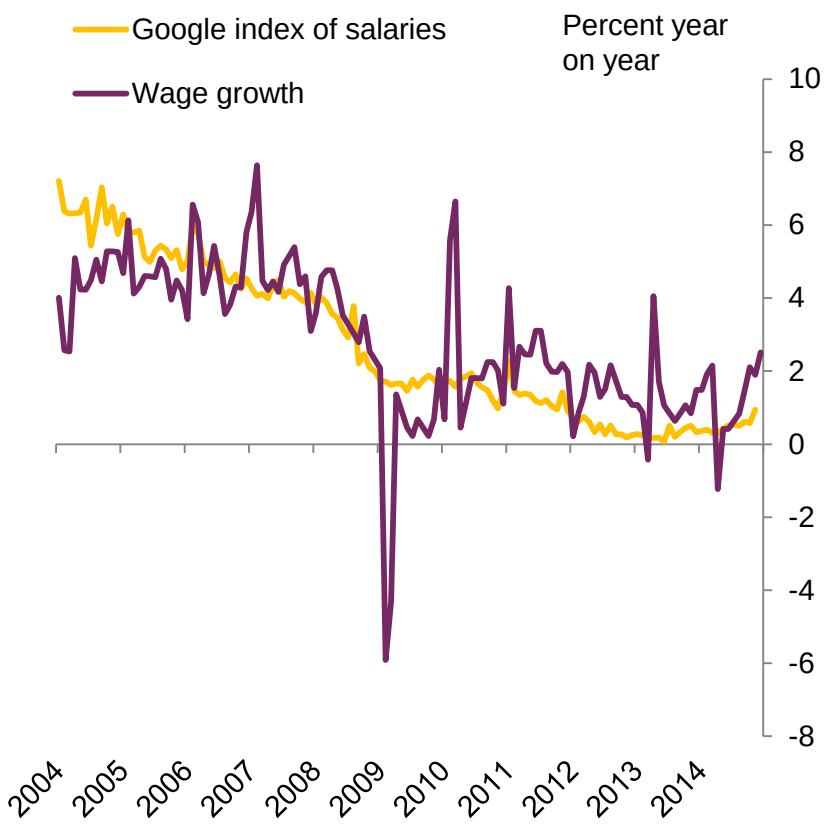
Understanding the labour market using job ads

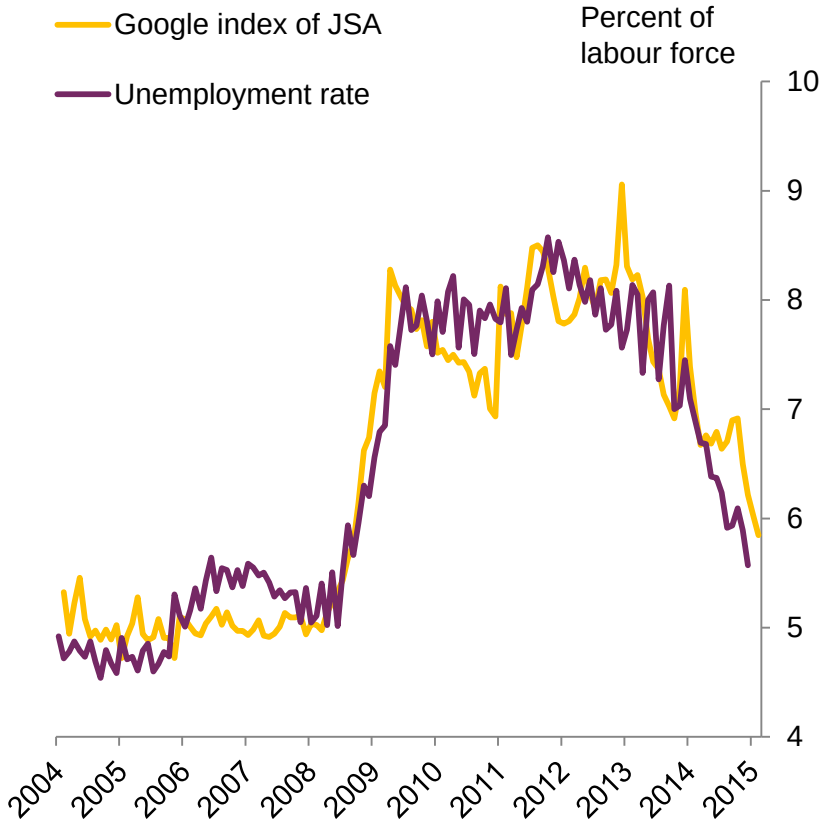


Big data sets offer significant potential advantages

- Greater **timeliness** (Velocity)
  - ‘Nowcasting’ and ‘nearcasting’
  - Always important, especially in times of crisis
- Greater **efficiency** / value for money (Value)
  - Using administrative data
  - ‘Found’ data

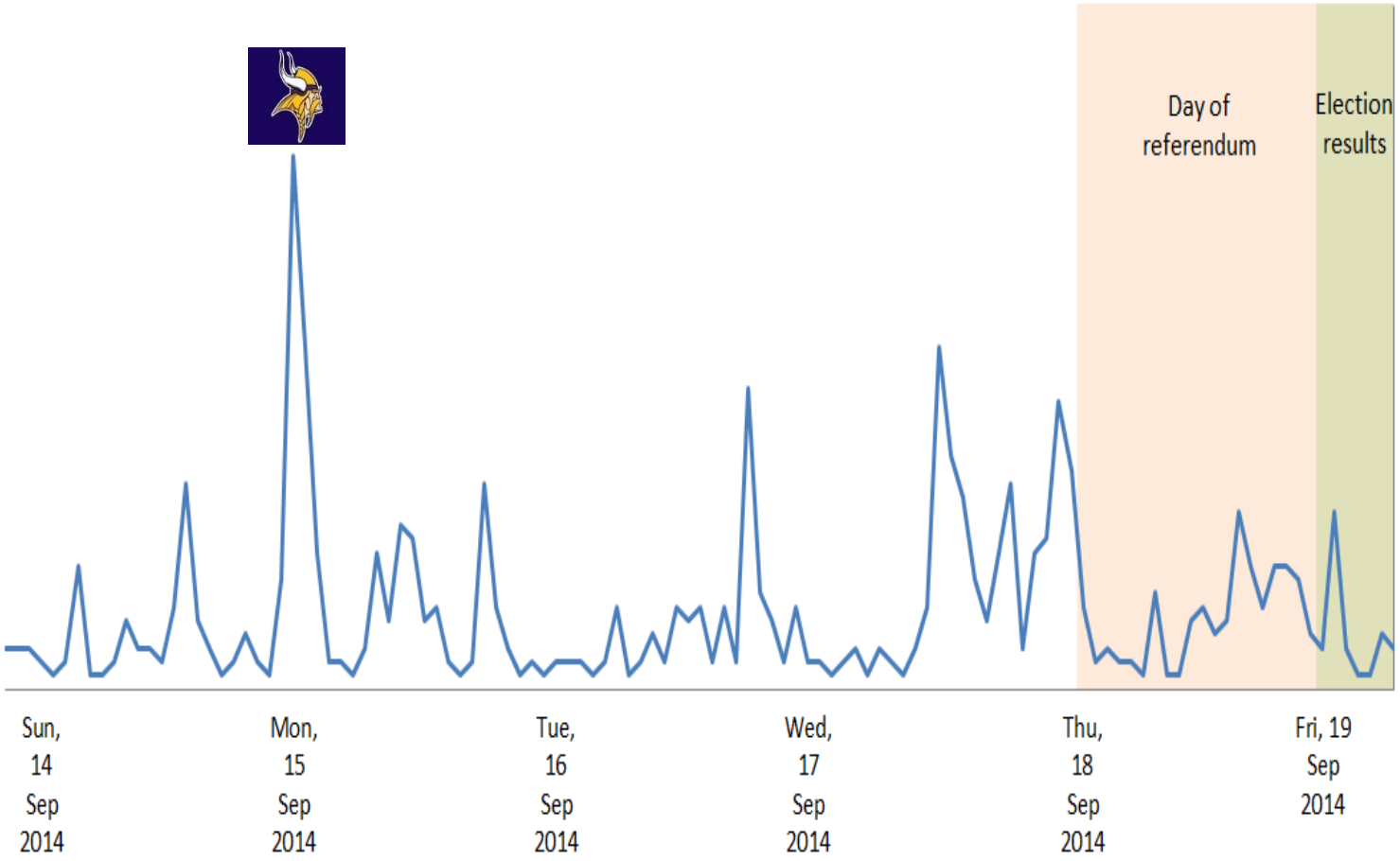
Googling the Labour Market





Source: ONS; Google. Notes: The Google indices are mean and variance adjusted to put on the same scale as the unemployment rate and wage growth. The Google indices are drawn from searches containing the terms “salaries” and “job seekers allowance”. See Mclaren and Shanbhogue (2011) for further details.

Exploiting novel datasets



Big data sets offer significant potential advantages

- New statistical / modelling approaches:
  - Machine learning
  - Network analysis
  - Agent-based modelling

Machine learning

- Different flavours:
  - Supervised
  - Unsupervised
  - Reinforcement learning
- Differences from conventional econometrics:
  - Typically focussed on prediction rather than identifying causal relationships
    - Individual parameter values are generally of limited interest
  - Use the algorithm and data to choose the model rather than theory
  - Use goodness of fit outside the 'training set' to determine the quality of the model rather than the familiar statistical tests
- Some key issues:
  - Feature selection
  - Regularisation
  - Researcher judgement vs 'letting the data speak'
    - 'Pure' objectivity is unusual

Machine learning models: supervised learning

- Typical approach:
  - Partition data into three sets:
    - Training set – used to choose the model
    - Validation – used to calibrate it
    - Testing – used to assess it
  - Often repeat the process many times

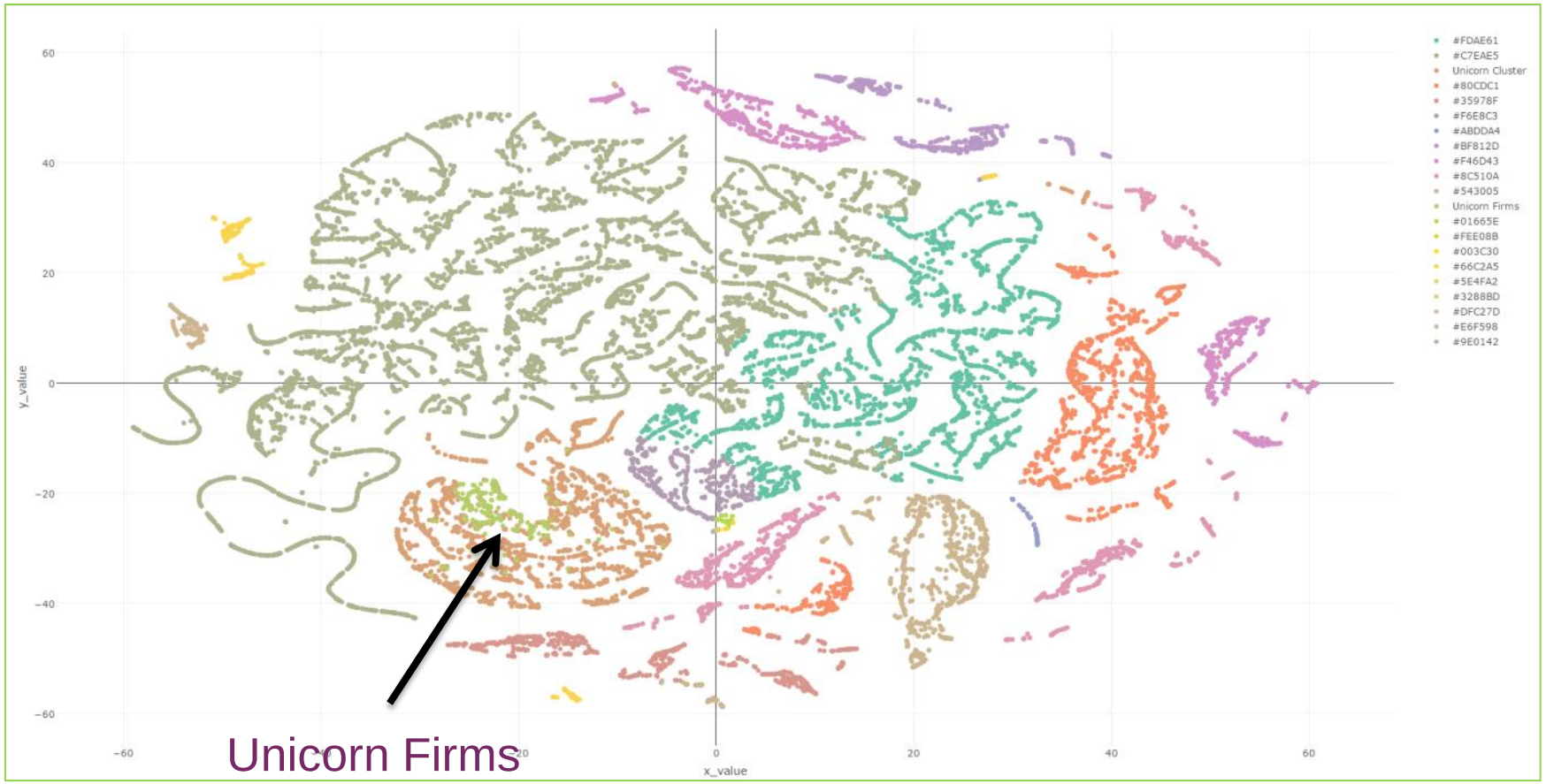
Machine learning models: supervised learning

- Some common models:
  - Linear regression-based:
    - Numerical solution of high dimensional models
    - Penalised regressions where number of explanatory variables is large relative to the number of observations (eg LASSO, Ridge, Elastic Net)
  - Non-linear regression:
    - Support vector machines
    - *K*-nearest neighbours
  - Tree-based:
    - Decision trees
    - Random forests
  - Neural networks

Machine learning models: unsupervised learning

- Classification and pattern identification
- Examples:
  - K-means
  - Hierarchical clustering
  - Neural networks (again)
  - Topic modelling

Cluster Analysis – Identifying potential financial disruptors



Identifying occupations

Three steps for grouping jobs based on the demand expressed in individual vacancies:

1. The text associated with each job vacancy is **cleaned** and the title and job description are combined into a single ‘document’ per vacancy

2. A **topic model** creates N topics to help determine **type** of segmentation

3. Group vacancies into K clusters (final sub-market **types**) using the K-means algorithm

Topic models and the LDA

- We model sectors using a topic model based on the Latent Dirichlet Allocation (LDA)
- Topics are identified by the use of common words and phrases
- Sectors are identified by being made up of common topics

$$\boldsymbol{\theta} = \underbrace{\begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \cdots & \theta_{1,N} \\ \theta_{2,1} & \theta_{2,2} & \cdots & \theta_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{D,1} & \theta_{D,2} & \cdots & \theta_{D,N} \end{bmatrix}}_{\text{Topics (columns)}} \Bigg\} \text{Documents (rows)}$$

Document-topic matrix

$$\boldsymbol{\beta} = \underbrace{\begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,N} \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{L,1} & \beta_{L,2} & \cdots & \beta_{L,N} \end{bmatrix}}_{\text{Topics (columns)}} \Bigg\} \text{Words (rows)}$$

Term-topic matrix

The topics

Word clouds of topics found using Latent Dirichlet Allocation.

Topic 0

Topic 1

Topic 2

Topic 3

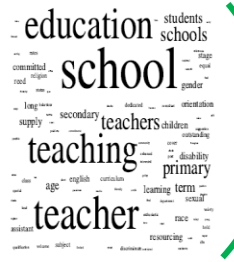
Topic 8

Topic 9

Topic 10

Topic 11

Topic 4



Topic 5

Topic 6

Topic 7

Topic 12

Topic 13

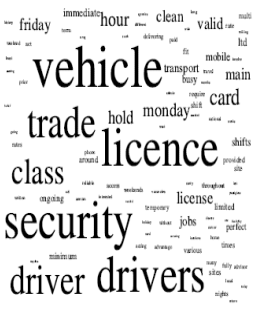
Topic 14

Topic 15

Topic 16

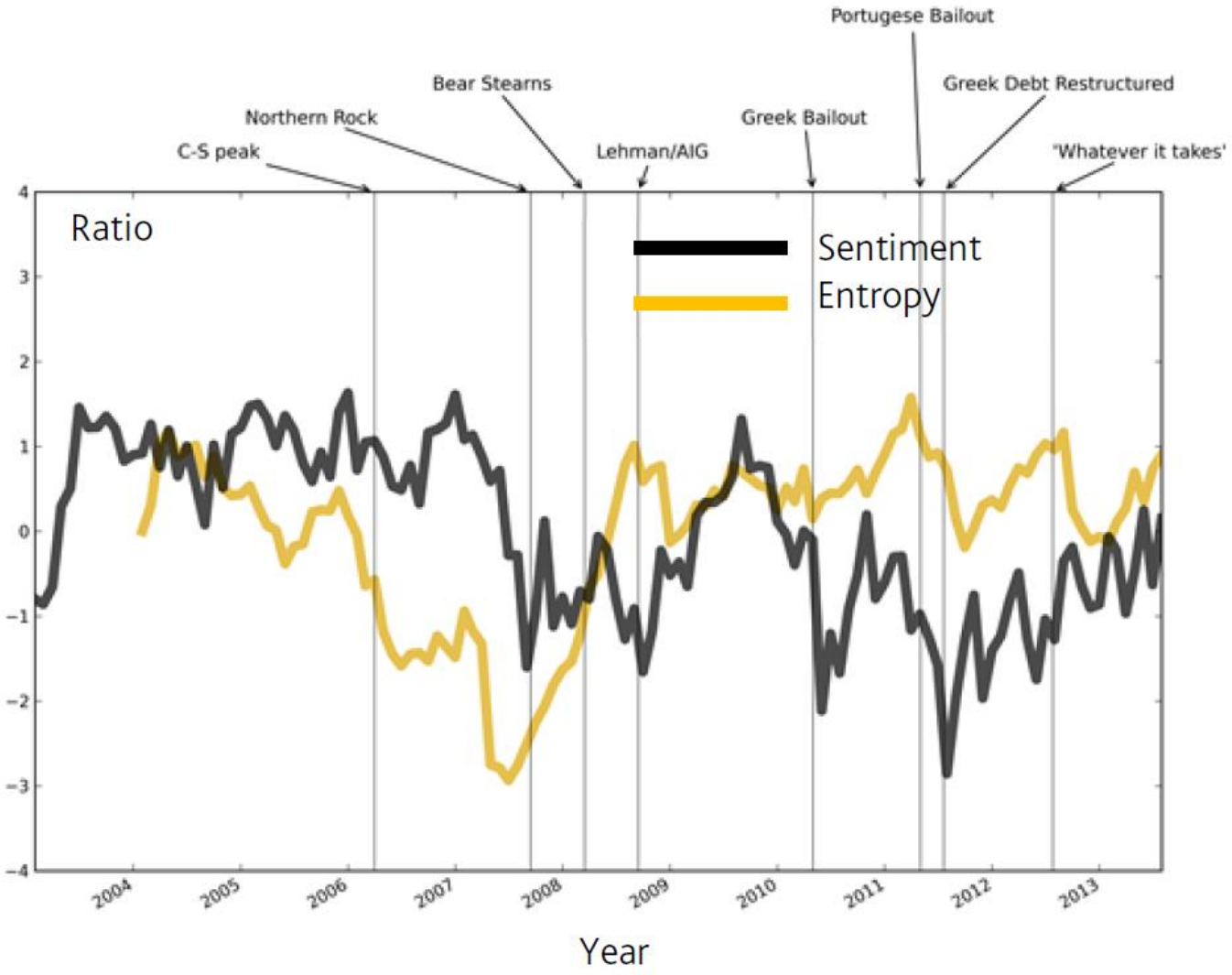
Topic 17

Topic 18



Topic 19

Sentiment or agreement?

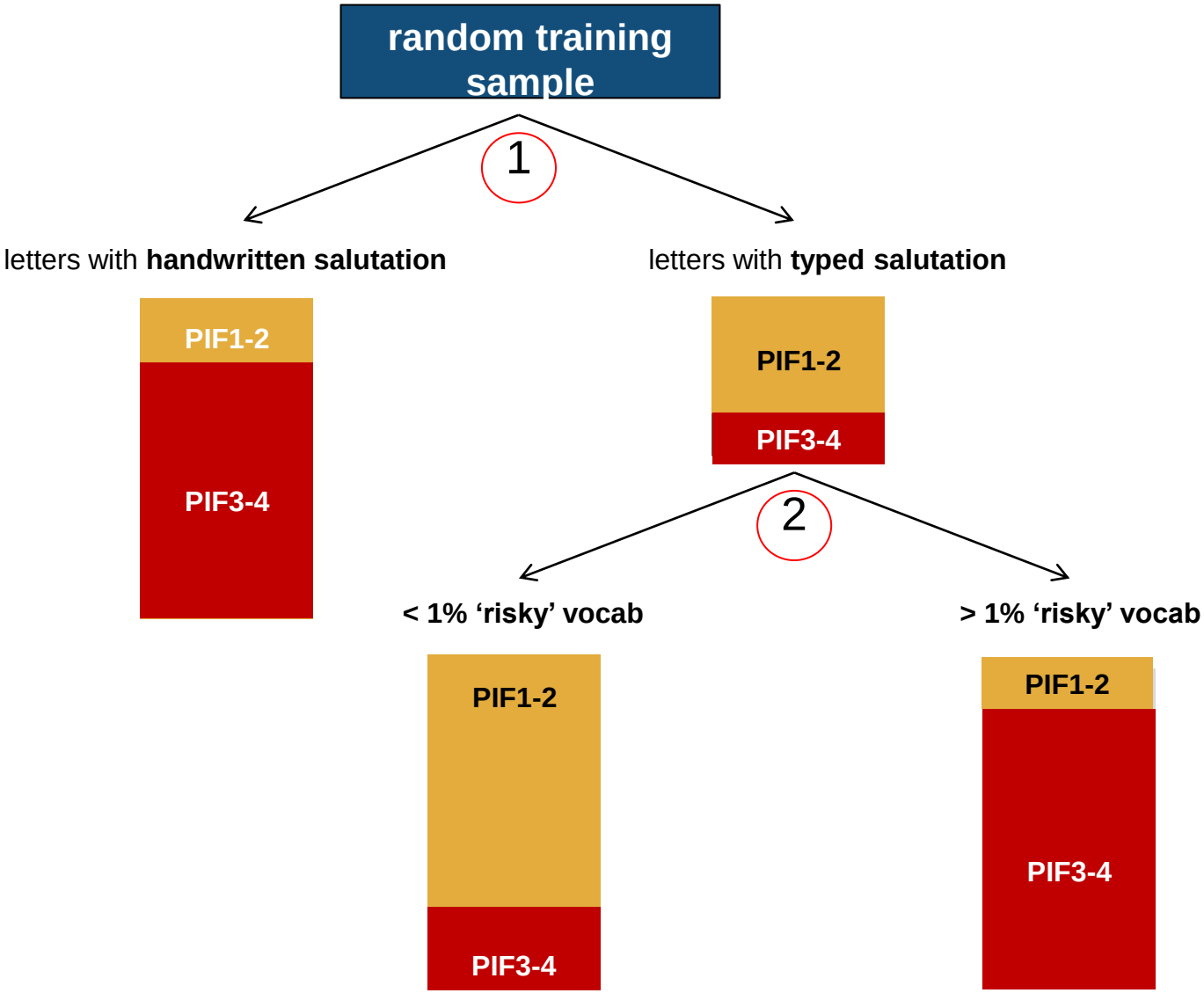


Using text and random forests to understand our own communications

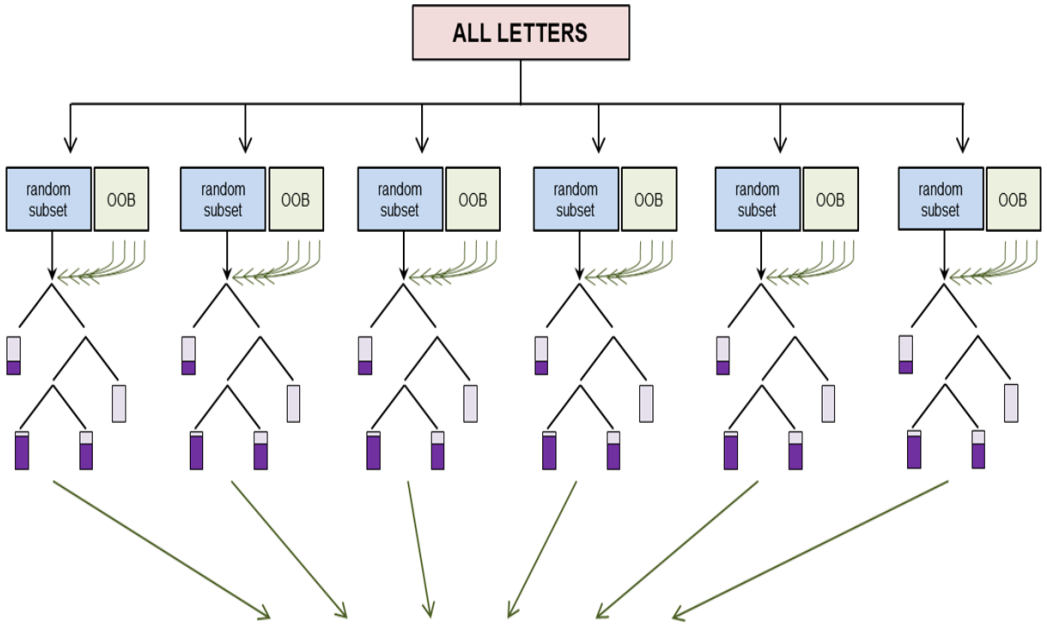
- Analysed periodic summary meeting (PSM) letters from the PRA to the supervised firms
- Are they written differently to firms with different risk profiles?
  - If so, what linguistic features distinguish sub-genres of PSM letters?
- We expected PSM letters to vary depending on firm riskiness
  - consistent with the PRA's principle of proportionality
- We expected higher risk firms to receive letters that were:
  - more complex
  - more negative in sentiment
  - more directive

Linguistic features considered

- Sentiment
  - Positive vs negative words
- Complexity
  - Length of sentences, number of subordinated clauses
- Directiveness
  - Instructions vs suggestions
- Formality
  - Eg “To Whom it may concern” [typed] vs “Dear Jane” [hand-written]
- Forward-lookingness
  - Future focus vs discussion of past developments


random training sample
1
letters with handwritten salutation
letters with typed salutation
PIF1-2
PIF3-4
PIF1-2
PIF3-4
2
< 1% 'risky' vocab
> 1% 'risky' vocab
PIF1-2
PIF3-4
PIF1-2
PIF3-4

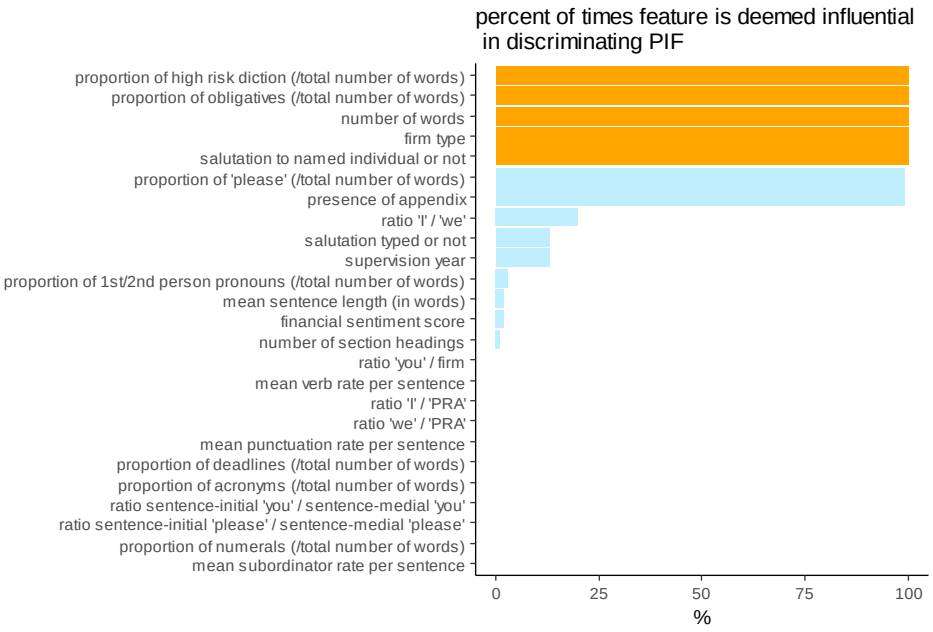
Random forests and text analytics in a regulatory context



Obs.	Obs. Class	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5	Tree 6	Combined Prediction	Prediction == Observed
1	PIF 1-2		PIF 1-2		PIF1-2		PIF 3-4	PIF 1-2	TRUE
2	PIF 1-2	PIF 1-2				PIF 3-4	PIF 3-4	PIF 3-4	FALSE
3	PIF 1-2			PIF 1-2		PIF 1-2		PIF 1-2	TRUE
4	PIF 1-2		PIF 1-2		PIF 3-4		PIF 1-2	PIF 1-2	TRUE
5	PIF 1-2			PIF 1-2	PIF 1-2			PIF 1-2	TRUE
6	PIF 3-4	PIF 1-2	PIF 3-4				PIF 1-2	PIF 1-2	FALSE
7	PIF 3-4			PIF 1-2	PIF 1-2			PIF 1-2	FALSE
8	PIF 3-4		PIF 3-4			PIF 3-4		PIF 3-4	TRUE
9	PIF 3-4	PIF 3-4		PIF 1-2		PIF 3-4		PIF 3-4	TRUE
10	PIF 3-4	PIF 1-2						PIF 1-2	FALSE

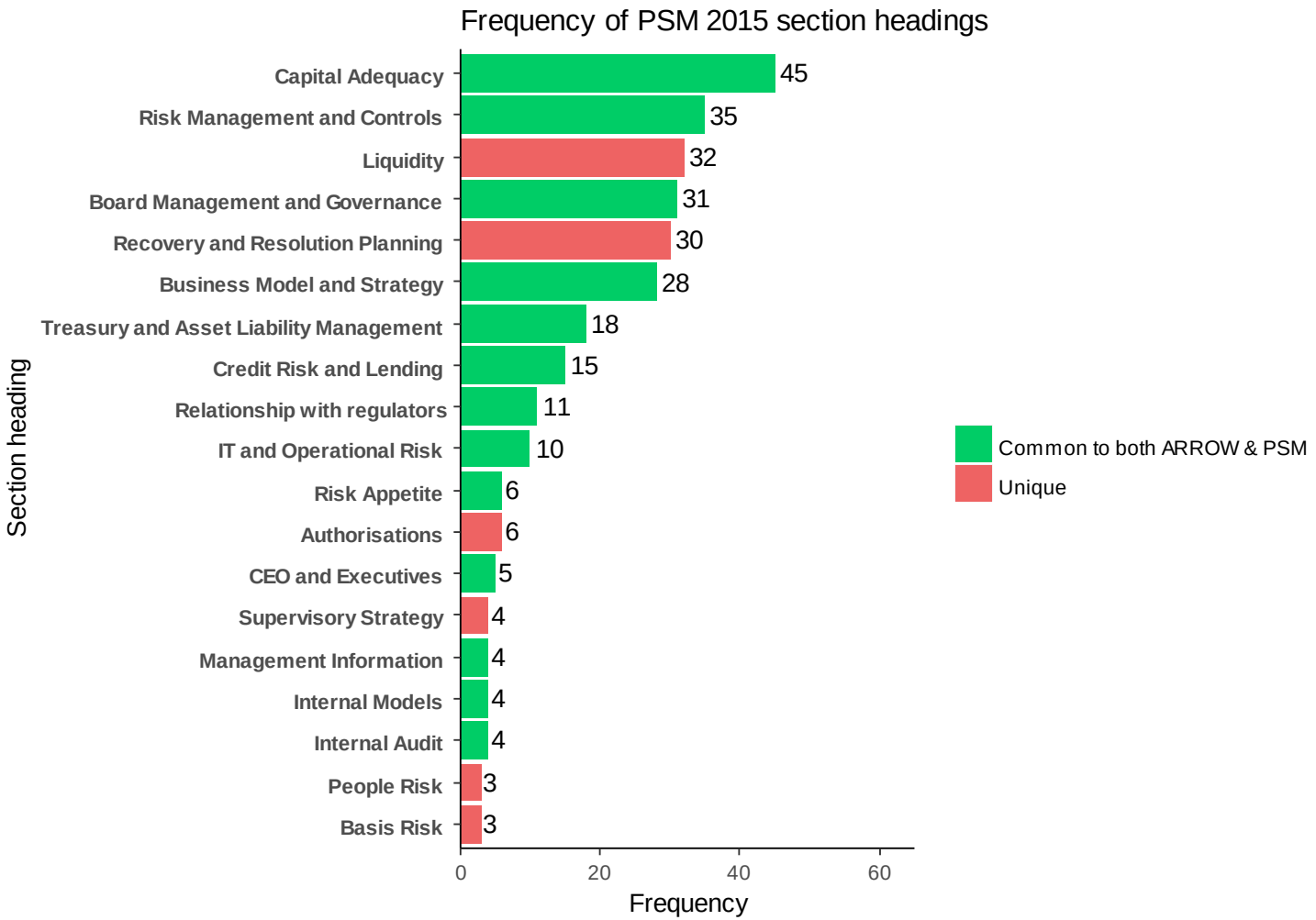
OOB Predictive Accuracy = 60%

PIF 3-4 PSM letters different from PIF 1-2 letters



- More complex
- More ‘high-risk’ vocabulary
- Less directive
- Less formal

PSM letters different from ARROW letters in content

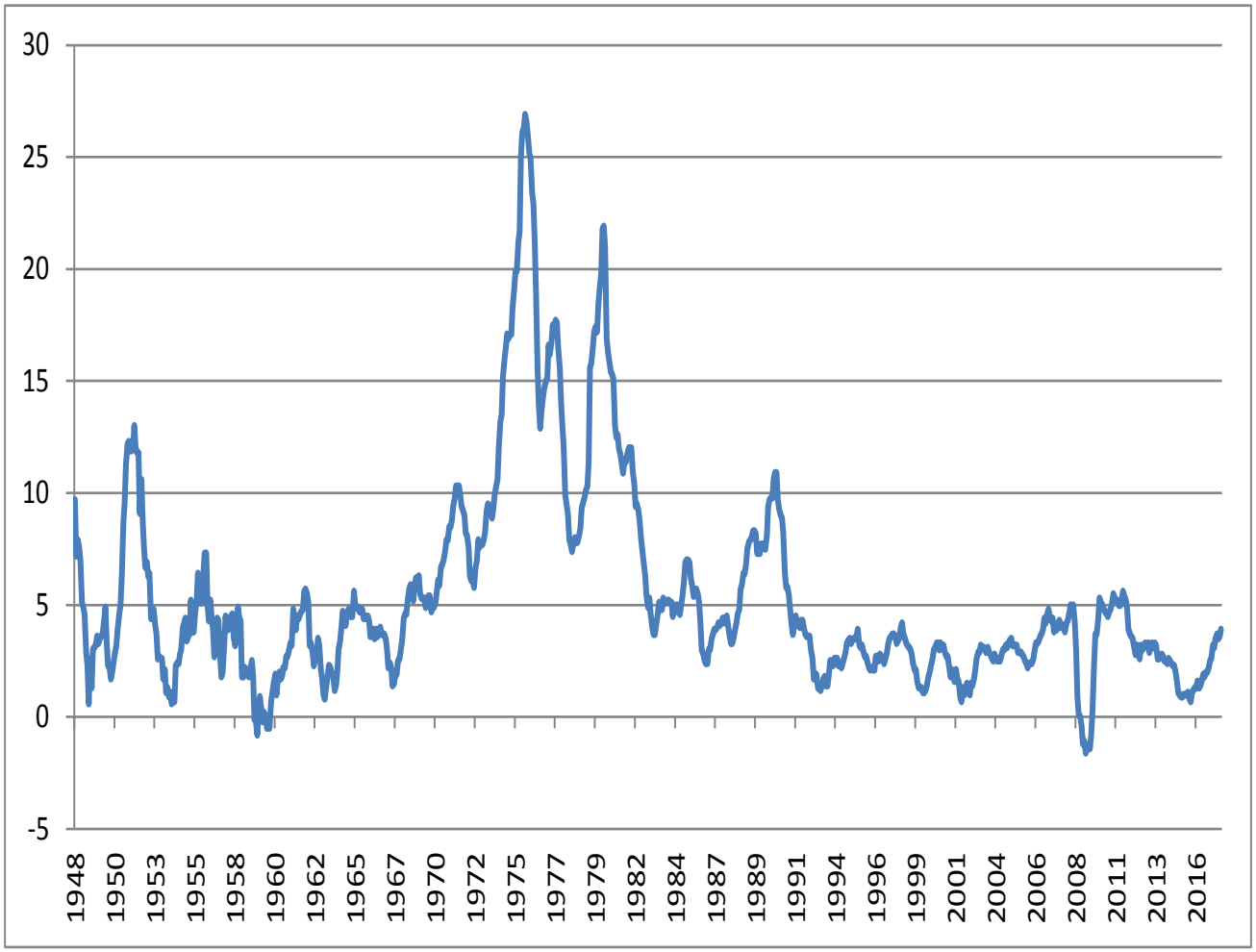


But there is no such thing as a free lunch …

Lots of data == lots of information?

- Example: CPI micro-data
- The ONS has produced a data set comprising:
  - 215 months (Feb 1996-Dec 2013)
  - ~110,000 prices collected per month (not the same number each month)
  - 1,113 items (not the same items each year)
  - 71 COICOP classes
  - various other meta-data (eg type of shop, region etc)
  - in total: 24,442,988 records with 25 fields
  - 611,074,700 pieces of data

Lots of data == lots of information?

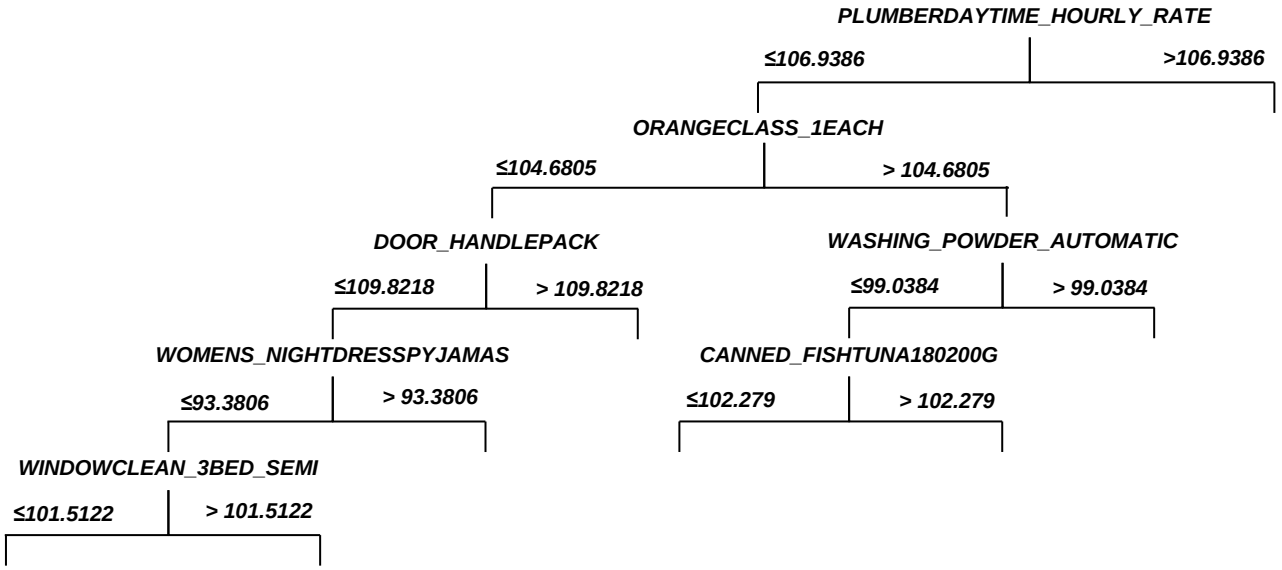


Correlation versus Causality

- ML focuses on prediction
  - Not on structural models
  - But central banks set policy and a policy intervention may change the structure of the economy
  - Beware the ‘Lucas critique’ (and structural breaks)
- This does not mean that ML is not a good fit for central banks
  - Forecasts often matter
  - Intermediate targets can be useful

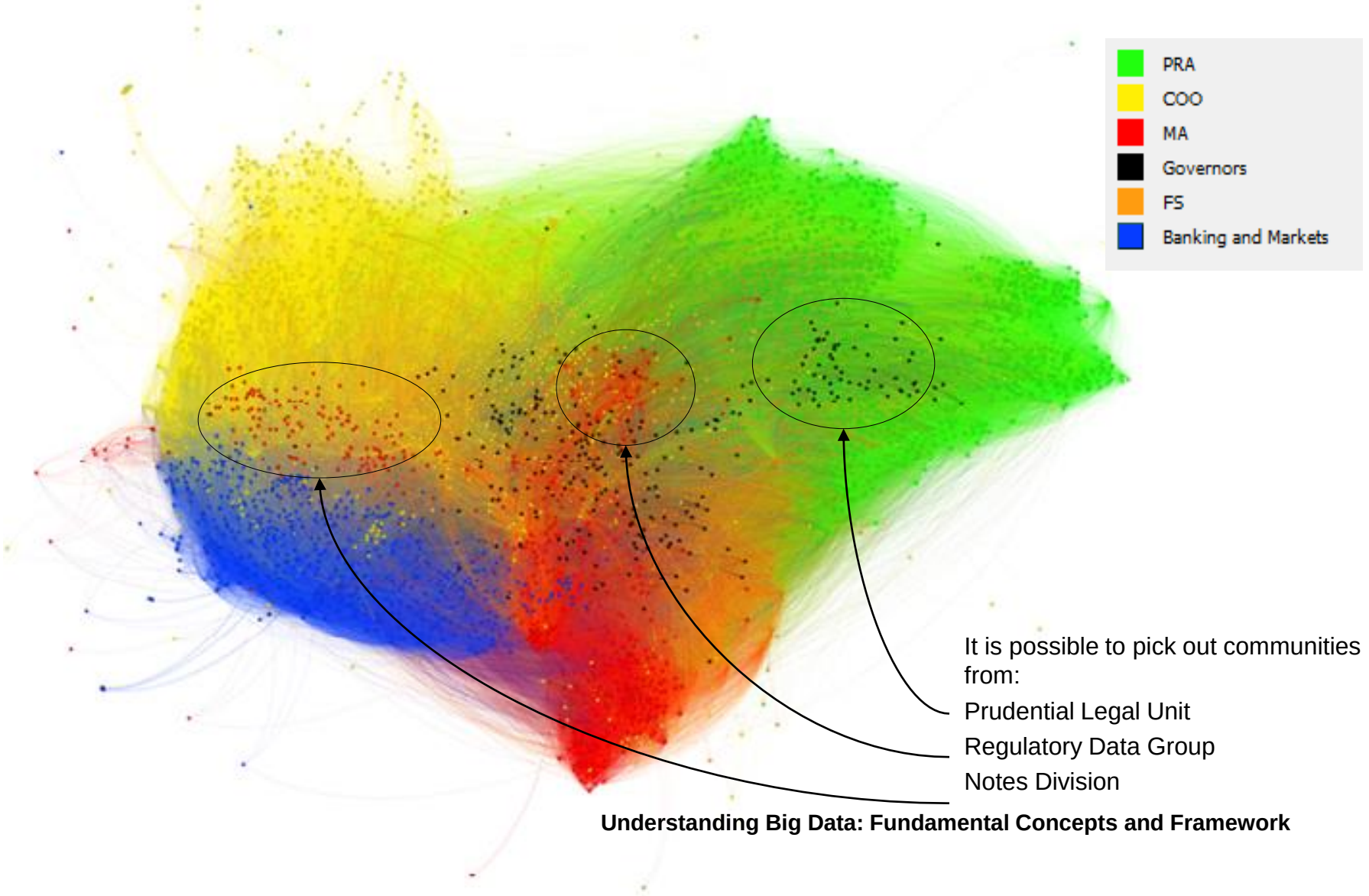
Overfitting

- Large data sets contain huge numbers of correlations …



Interpreting complicated, often highly non-linear relationships

Email connections: January 2015



It is possible to pick out communities from:

Prudential Legal Unit

Regulatory Data Group

Notes Division

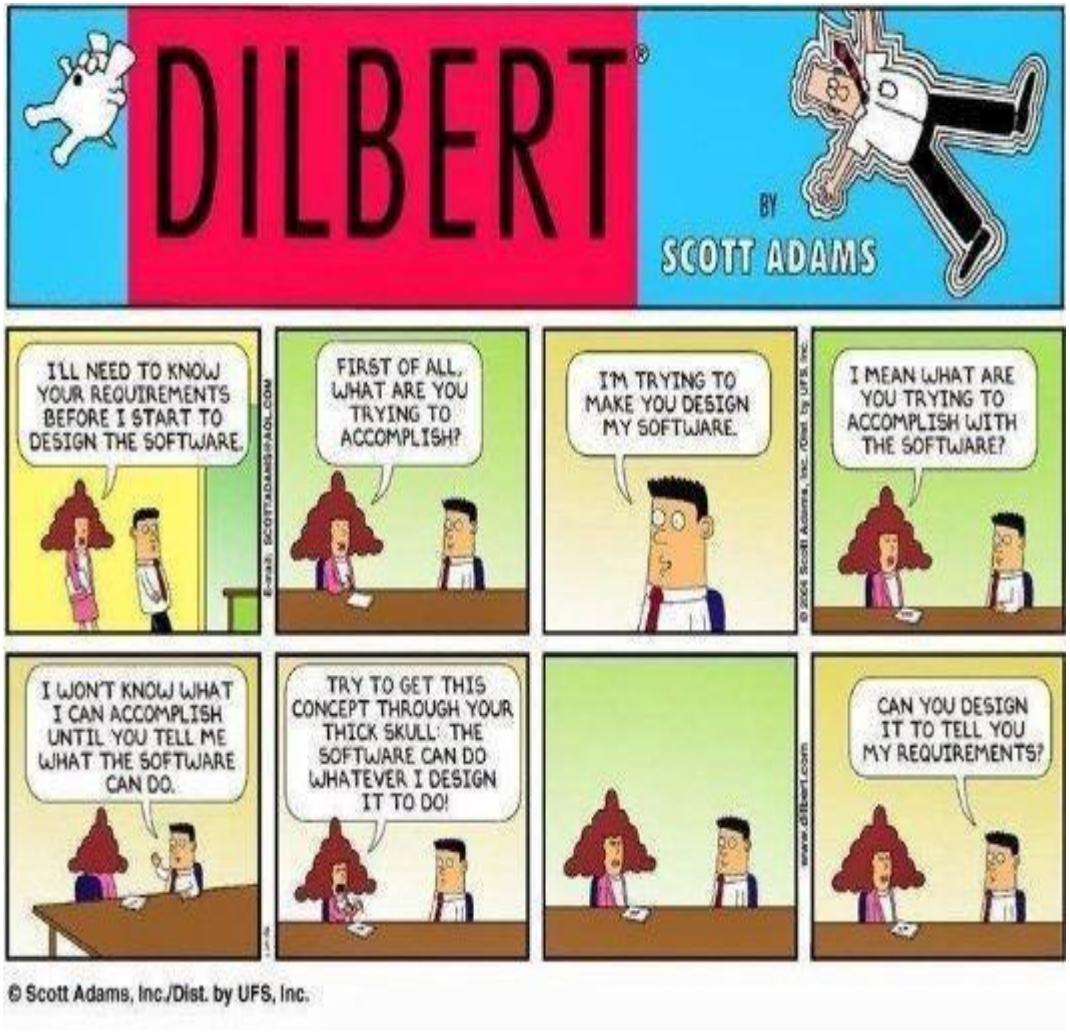
“Veracity”

- Big data sets are often populations, not samples
  - Therefore no sampling error
- But the observed population characteristics may not be typical of the underlying data generating process
- Or it may be biased relative to the true population of interest

Confidentiality / ‘Big Brother’ state

- This was not relevant to the CPI work
- In general, the more detailed and granular the data set is, the more likely it is to contain confidential information
- We must ensure that:
  - we only use data for appropriate reasons
  - the minimum number of people are able to see any confidential data given the needs of the situation
  - data are stored securely and professionally

Engage with AA, but there are no free lunches …



Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Big data for central banks[1]

Bruno Tissot,

Bank for International Settlements

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

Big Data for Central Banks

Bruno TISSOT
Head of Statistics and Research Support, BIS
Head of Secretariat, Irving Fisher Committee on Central Bank Statistics (IFC)

International Workshop on Big Data for Central Bank Policies – Bali, 23-25 July 2018
Session 1

The views expressed are those of the author and do not necessarily reflect those of the BIS or the IFC.

Overview

- Introduction
- Financial Big Data
- Three key developments
- Challenges in handling and using big data
- Analysing CBs' experiences
- Annexes: Selected references/ BD projects by CBs

Introduction – Big Data…

- General & increasing **policy interest** for "Big Data" (BD)

→ *the world's most valuable resource is no longer oil, but data* (The Economist)

- Term **usually describes**
  - Extremely large data-sets
  - Often a by-product of commercial or social activities
  - Huge amount of granular information, typically transaction-level
  - Data available in, or close to, real time
  - Used to identify behavioural patterns / economic trends

→ ***Growing impact*** *on information creation, storage, retrieval, methodology, analysis*

Introduction – … for Central Banks…

- **Private sector** use big data to produce new & timely indicators

- New opportunities **also for Central Banks (CBs)** – as well as macro-prudential authorities and financial supervisors?
  - Broader and timelier range of indicators
  - New statistical methodologies

- Extraction of **new type of information** supporting
  - Economic forecasts & analysis
  - Financial stability work
  - Policy impact evaluation

Introduction – … with significant opportunities…

- **Focus on sources** that can effectively support micro- and macro-economic as well as monetary and financial stability analyses
  - Other big data – eg geospatial information – of lower interest
- **Feedback loop** inherent to policy-making authorities
  - Big data sources can affect policy-making
  - In turn policies implemented can generate new data-sets
- Big data provide **new "business opportunities"** for CBs, such as:
  - Qualitative statements to decipher central banks' communication
  - Large number of big data pools generated by financial regulations
  - In turn, big data can strengthen supervisors' capacity

Introduction – … but also challenges…

- Specific **challenges faced in handling and using** big data
  - Public nature of financial authorities and public trust
  - Central banks concerned about ethical & reputational consequences
  - Risk of misusing big data for policy actions?

- Different **data quality concerns** compared to private sector
  - Ex: online retailers targeting potential customers based on past web searches might find it acceptable to be "right" 20% of the time
  - Such a low accuracy level looks inadequate for official statisticians

Introduction – … not least due to security concerns…

- Increasing **security concerns linked** to internet / big data, such as:
  - Risk that large private records of individual information could be accessed and potentially misused by unauthorized third-parties
  - Resilience of financial market infrastructures

- **Influence on central banks'** actions
  - Preserve public trust, especially when collecting data
  - Supervise firms' capability to gather and interpret security-related information
  - Set standards and best practices
  - Promote cyber threat intelligence and modelling techniques

Introduction – … with the risk of being behind…

- **CBs' constraints** compared to private firms
  - Basic resources needs (IT budget, staff)
  - Concerns about the lack of transparency in methodologies
  - Poor quality of some data sources hampering public use
- **IFC survey** of central banks
  - Big data work still on an exploratory mode
  - Regular production of big data-based information likely to take time
  - Yet increased interest esp. at senior policy level

Introduction – … and the need to be proactive

- Key objective for central banks is to **better understand**
  - The new data-sets and related methodologies for their analysis
  - The value added in comparison with "traditional" statistics
- Focus on **pilot projects** to assess how big data can help to
  - Better monitor the economic and financial situation
  - Enhance the effectiveness of policy
  - Assess the impact of policy actions
- Possible tasks may well **further expand**
  - Constant creation of new information/research needs
  - Cf Haldane (2018): exploring behaviours in a "virtual economy"

I – What is Financial Big Data?

- **Broad approach for BD**: by-product of commercial or social activities, providing a huge amount of very granular information

- **Yet**:
  - Not sufficient to be large to qualify as "big data" – cf census
  - Unstructured data require new tools to be processed
  - Structured data-sets handled with "traditional" techniques?

- Choice of the **relevant metric**
  - Volume of data?
  - Specific characteristics of big data-sets?
  - Timing issue? "Big data" 10 years ago versus today

I –Financial Big Data: 3 main BD groups…

- Definition by United Nations Department of Economic and Social Affairs

- Big data type of information **classified in three groups,** as a product of:
  1. **Social networks** (human-sourced information, eg blogs, videos, searches)
  2. **Traditional business systems** (process-mediated data, such as data produced by commercial transactions, e-commerce, credit cards)
  3. **The internet of things** (machine-generated data, such as data produced by pollution/traffic sensors, mobile phone information, computer logs)

I – Financial Big Data: … with a key distinction…

- 1. **Unstructured data-set** (often quite large):
  - By-product of a non-statistical activity – "produced **organically**"
  - Different from the datasets produced for traditional statistics, which are structured by **design**

- 2. **Data-set with large records, relatively** well-structured
  - **Difficult** to handle because of size, granularity or complexity
  - Even "simple" structured datasets can **benefit from big data techniques**

I – Financial Big Data: … some judgment…

- Room for **judgment**, depends on **features such as the "Vs"**
  - **Volume** (number of records and attributes)
  - **Velocity** (speed of data production, eg tick data)
  - **Variety** (for instance structure and format)
  - **Veracity** (accuracy / uncertainty of large individual records)
  - **Valence** (interconnectedness of the data)
  - **Value** (often a by-product of an activity, can trigger a monetary reward)
- Features characterising big data can be very **diverse**
- **Information content** also quite heterogeneous

I – Financial Big Data: … 2 main sources for CBs…

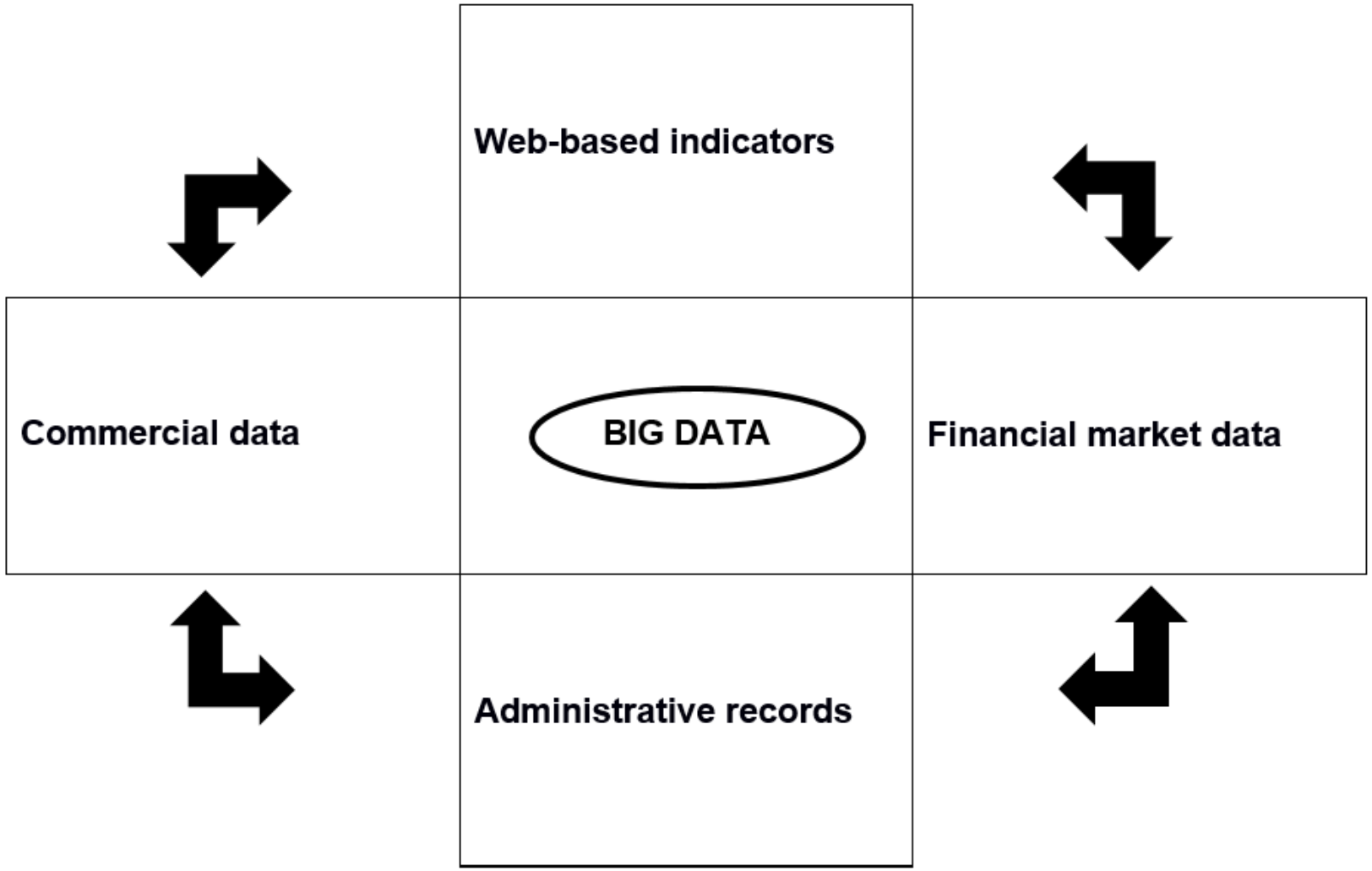
- CBs see Big Data as comprising the variety of large-scale information requiring/benefiting from **"non-traditional" tools** to be processed & analysed
- **Two data sources** relevant for central banks:
  - Restricted view: **the "internet of things"-type** of unstructured data, heavily used by the private sector

  → *Public interest eg Google Trends, but not really the core*

  - **Large registers, by-products** of 3 types of activities: financial, commercial & administrative

  → *Key issues include confidentiality and quality*

I – Financial Big Data: … and 4 main types of BD-sets…



I – Financial Big Data: …with overlaps…

- Increasing part of the **information collected on the web** can be the result of financial, commercial or administrative activities
- Cf recent expansion of "**Fintech**"
  - "*Technology-enabled innovation in financial services that could result in new business models, applications, processes or products with an associated material effect on the provision of financial services*" (FSB, 2017)
  - Parallel innovations: big data, mobile phone, internet, artificial intelligence
- Multiple **applications** that blur traditional boundaries
  - Digital currencies (Bitcoin)
  - Various applications in payments, crowdfunding, smart contracts, robot advice, credit risk assessments & contract pricing

I – Financial Big Data: … practical issues…

- In practice **CBs deal with various & heterogeneous "big data"**
  - Usually **not directly** produced for a specific statistical purpose, as in the cases of traditional census or survey exercises
  - **Indirectly**, data sources can be exploited for addressing statistical information needs that may independently exist
  - "**Smart data**": treatment of the raw, "organic" data is key
- Public authorities just be at the beginning of making sense of all the **increasing volume and variety of data**
  - Use of specific data sources depend on policy questions
  - Eg payment systems data: of interest to supervisors and tourism analysis

I – Financial Big Data: … complexity…

- **Micro-level BD universe** is complex and evolves over time
  - Interaction between data available and specific policy needs

- **Transforming data into information** requires
  - **Merging** different sources, with common identifiers
  - Dealing with **inconsistent** observations
  - Choosing a particular **source**
  - **Aggregating**, by using parent relationship and rules
  - Choices may depend on **circumstances** (“time dependency”)

I – Financial Big Data: … example

- Example: BIS **International Debt Security** issuance statistics
  - Micro aggregation derived from large security-by-security data-sets
  - Data collection based on a "traditional" residency concept…
  - … and a "nationality basis" (include debt issued by foreign affiliates)
- **Constructing nationality-based statistics** requires to
  - Identify the perimeter of global firms
  - Reclassify individual units
  - Consolidate granular information at the group level
  - Tasks both time-dependent and source-dependent

→ Handling large & complex data can benefit from **BD techniques**

II – Three key developments

- Big Data as a result of the **combination of three key developments in the financial area**
  - The **internet** of things
  - **Digitalisation**
  - Expansion of **micro financial data-sets** in the aftermath of the Great Financial Crisis (GFC) of 2007-09

II – 3 Developments: Internet of things (1: new data)

- **Significant experience in recent years** in collecting information generated by the wide range of web and electronic devices
  - Search queries, clicks on specific pages, display of information and text online, social media messages…

- Can be **used to complement "standard" statistical processes**
  - In general, the (near) real time availability of web data can allow for getting rapid information and improving timeliness
  - Approach to estimate current patterns and forecast them in advance of actual publication dates (nowcasting exercises)

II – 3 Developments: Internet of things (2: inflation)

- Example: **"scraping" prices** posted online by retailers
  - Exercises typically limited to specific inflation components (eg volatile fresh vegetables' prices)
  - Process appears robust, scalable and can be automatised
  - Important challenges: capturing unit-level prices, product characteristics, quantities, adequate weights

- **Billion Prices Project** (MIT):
  - Enhanced international comparisons of price indexes
  - Dealing with measurement biases
  - Addressing distortions in international relative prices

II – 3 Developments: Internet of things (3: house prices)

- Example: collection of **housing prices** on the web
  - Scraping prices displayed by real estate agencies
  - Capturing the various housing characteristics posted in advertisements facilitates the calculation of quality effects (hedonic prices)
- **Challenges**
  - Collecting the information in a comprehensive & structured way
  - Weighting schemes

→ Particularly relevant for economies with *less developed statistics*

→ In more *advanced countries* also: property prices often derived from low frequency surveys / for a limited number of cities

II – 3 Developments: Internet of things (4: real activity)

- **Real-side economic indicators**
  - Job web announcements: indicators of business activity & unemployment
  - Monitor consumption of durable goods (eg cars)
  - Overall level of the economy / specific sectors (eg tourism) / areas

- But this use has been relatively **incremental and limited,** even for national statistical agencies in advanced economies, and often targeted at:
  - **Methodological** improvements (eg quality adjustment)
  - Reducing reporting **lags** and data revisions
  - **Alternative** to the organisation of large surveys (eg India)

II – 3 Developments: Internet of things (5: new insights)

- Possibility of capturing **unsuspected data patterns**
  - "Traditional" statistical modelling to infer economic relationships
  - BD algorithms to incorporate various effects without ex ante assumptions
  - Techniques can be implemented easily and in an automated way
- Opportunity to **incorporate qualitative information**
  - Clicks from web searches, twitter messages or posts on social medias
  - Incorporating sentiment and agents' expectations for measuring risks, changing preferences, causality patterns…
  - Factors that play an important role during crises and are quite difficult to model (non-linearities, network effects)

II – 3 Developments: Internet of things (6: drawbacks)

- Data **quality issues**
  - Errors, typos and self-fulfilling expectations
  - Need to collect consistent information but goods are not kept identical
  - Announcement prices can differ from actual transaction prices
  - Advertisements remain posted after economic transactions are settled
  - Accuracy of the information that individuals (or robots!) input to the web
  -
- Key limitation is that the data are **not well structured**
  - Details on the location of a transaction / job offer difficult to get
  - Underlying information can be collected several times

II – 3 Developments: Internet of things (7: challenges)

- **Technical** challenges
  - Use of new techniques (eg web-scraping) and methodologies
- **More fundamental** challenges?
  - Limited interpretability of "black box" calculations
  - Mining data and the need to derive meaningful conclusions from an economic perspective
  - CBs need to present a consistent "story" when communicating policy

II – 3 Developments: Digitalisation (1: new information)

- **Expanded access** to digitalised information
  - Rise in textual information moving to the web (while not produced by internet activities strictly speaking)
  - Reference documents can be digitalised, accessed and analysed like "web-based" indicators
- Can be more easily and automatically **exploited through ad hoc BD techniques:** eg text semantic analysis
  - Extraction of textual information of interest
  - Characterising text attributes and similarities
  - Classifying information content (eg tone of central banks' messages)
  - Assessing the impact of external factors (eg circumstances, policy actions)

II – 3 Developments: Digitalisation (2: new opportunities)

- Techniques can also be used to **measure impact** on economic agents' expectations
- Structured way to assess **policy communication**
  - Perceived stance of public authorities' communication
  - Impact of this communication / action in view of the messages expressed in reaction by stakeholders
  - Formation of public expectations

→ **Complement** traditional "event studies" (eg central bank actions)

→ Provides **opportunities** for fine-tuning policy communication

II – 3 Developments: New financial statistics (1: a revolution?)

- **Revolution in financial statistics** observed since the GFC
  - Limitations of aggregated data: consider those institutions that are systemic on an individual basis
  - Need to measure the distribution of macro indicators, look at "fat tails" and go "beyond the aggregates"
  - Revolution comparable to the 1930s for the real accounts?

- Unprecedented **efforts to collect more** information on the financial sector – the Data Gaps Initiative (DGI) endorsed by G20
  - High demand for large, granular and complex data-sets
  - Collected at the level of institutions, transactions & instruments

II – 3 Developments: New financial statistics (2: CBs' interest)

- **Fundamental factors explaining why CBs' have been leading the way** for collecting such financial big datasets
  - Go beyond aggregated indicators
  - Make a better use for policy of available/expanding micro-level datasets
  - Realisation that a huge amount of information is already available and could be better exploited (eg administrative data)

- Focus on **very granular information**, derived from various sources, and more complex compared to "typical" web-based data

II – 3 Developments: New financial statistics (3: CCRs)

- Example: rising **demand for detailed loan-by-loan / security-by-security** information
  - Central credit registries (CCRS) have become the largest data-sets maintained by some central banks
  - Europe's AnaCredit: "analytical credit dataset"
  - US FRBNY Consumer Credit Panel: detailed information on consumer debt and credit derived from individuals' reports
- Data are well structured, but reporting is **highly granular**
  - Multiple attributes: 200 attributes per data point on a monthly basis (and on a daily basis for a subset) for AnaCredit
  - Often complex to aggregate / analyse

II – 3 Developments: New financial statistics (4: specificities)

- Information often derived from **confidential operations** (tax registers, banks' books)
  - Richness across the population of interest (eg capturing very small enterprises)
  - Usually collected regularly over a long period of time
  - But need for anonymization / confidentiality protection
- **CBs learning from private sector**
  - Increased experience in dealing with large data-sets (eg production of "stress tests")
  - Supervisors of financial firms to develop their expertise in these areas too

III – Challenges

- **Handling big datasets** requires significant resources and proper arrangements for managing the information
- **Using big data in policy-making** creates opportunities but is not without risks
- Key **implications**
  - Explains why public authorities' actual use of big data is still limited, at least in comparison to the private industry
  - Significant time and effort needed before any regular production of big data-based information for supporting CBs' statistical and analytical work on a large scale

III – Challenges in handling big data (1)

- **Resources** and proper arrangements for managing BD
  - Sheer size of the data-sets
  - Lack of structure
  - Often limited quality of raw data
- The statistical **production process** itself has to be adapted
  - Work to appropriately collect, clean, reconcile and store BD
  - Usually, BD produced without standard quality controls of "traditional" statistics (while public authorities put a lot of attention on those issues)
  - Significant number of false/inconsistent/missing records

III – Challenges in handling big data (2)

- Need to **set up a clear and comprehensive information management process**
  - Data acquisition
  - Data preparation
  - Data processing
  - Data validation

- A **major area is IT**
  - Large processing costs, difficult & expensive technology choices
  - Sophisticated statistical techniques: “BD algorithms”, “ML techniques”, “AI”
  - Public authorities with less budget compare to private sector

III – Challenges in handling big data (3)

- New issues in terms of **confidentiality protection and security**
  - Large amount of data provided by users through their web-based activities
  - Large financial datasets require the handling of transaction-level, potentially highly confidential, information
  - Data privacy issues may increase with the development of big data and Fintech firms

- Potentially **wider implications**
  - Operational incidents can lead to significant privacy and legal issues, with financial consequences
  - Cf European General Data Protection Regulation (GDPR, 2018)

III – Challenges in handling big data (4)

- A key risk: **reputation risk**
  - Peculiar position of central banks if private information is reported to them but not protected adequately
  - Especially for regulatory-type data collections
  - But internet-based information, often a by-product of commercial activities, can also pose significant legal, financial, reputational & ethical issues
- **Operational implications**
  - Public statisticians tend to be "cloud computing-adverse"
  - Preference to operate in a "secluded" data environment
  - But could reduce opportunities to use BD techniques in the marketplace

III – Challenges in handling big data (5)

- Ongoing substantial **internal organisational changes** to deal with big data
  - Cf creation of internal centres for big data statistics, "data lakes", "internal clouds"

- Another key area is **staff**
  - Various skills needed: IT, data science and methodology, legal expertise…
  - A "war for talent"? A competition with the private sector that may be difficult for CBs…
  - Additional issues: compensation, career path, management

III – Challenges in handling big data (6)

- How to **enhance existing information management processes**?
  - Goal: flexible production of relevant information out of data points
  - "Traditional", template-driven data collections to be replaced by accessing granular data from various sources

- **Requirements**
  - Greater harmonisation of data-sets, statistical standards, identifiers and dictionaries
  - International efforts eg to develop global Legal Entity Identifiers & automated data exchanges standards (XBRL, SMDX, ISO 20022)

III – Challenges in handling big data (7)

- Better **integration of various IT systems** among both authorities and reporting entities

- Recent "**Fintech innovations**" to facilitate secure data transfer mechanisms:
  - Distributed ledger technology (DLT) to enable network participants to securely propose, validate and record information to a synchronised ledger distributed across the network
  - Each transaction can be recorded in a batch (a "block") and added to the full transactions' history (the "blockchain")

- Involvement of **private service providers ("regtech"** industry)

III – Challenges in using big data (1)

- Big data **opportunities** for policy use but is not without risks
  - Immediate benefits: lower production costs, new insights, production speed
  - To be balanced against potential large economic and social costs of misguided policy decisions

- **Key question**: does "big data" provide a more accurate picture of economic reality?

III – Challenges in using big data (2)

- Risk of **conveying a false sense of accuracy** and precision
  - Problem exacerbated by the organic nature of BD:
  - Data often self-reported or by-product of social activities
  - Coverage bias unknown, can be significant (eg social media users)

→ *Extremely **large big data samples may thus compare unfavourably** with (smaller) traditional probabilistic samples – precisely designed to be representative of the population of interest*

→ *Key **misperception** of the intrinsic value of big data*

III – Challenges in using big data (3)

- Risk of **undermining public policy**?
  - Effectiveness (if data are providing wrong signals)
  - Reputation/legitimacy

- Might systematically **alter decision-making**?
  - Greater ability to monitor the economy in real time: bias towards responding quickly and more frequently to news, encouraging shorter horizons?
  - Greater reliance on "big data"-based analyses of sentiment: risk of excessively fine-tuning policy communication based on perceived expectations rather than actual economic developments?

IV – Analysing CBs' experiences (1)

- Proper **information management frameworks** needed to make the most of big data so as to:
  - Address challenges faced when handling and using big data
  - Avoid the risk of focussing on cumbersome management tasks – cleaning, documenting, organising data – instead of *using* the information

- Key is to **make sense of the data** collected, with a coherent information management framework

- CBs are following **step-by-step** approaches, with specific use cases instead of "big bang" solutions

IV – Analysing CBs' experiences (2)

- **Pressing challenges**
  - Combination of internet / digitalisation / new post-crisis initiatives
  - Authorities just at the beginning of making sense of the increasing volume, granularity and variety of information
  - "*Connecting the dots is as important as collecting the dots, meaning the right data*" (Caruana, 2017)
- Fundamental **distinction between "data" and "information"**
  - "Traditional" official statistics were "designed data" collected for a specific purpose
  - By definition were organised in order to extract meaningful information
  - Key difference with "organic"-type big datasets

IV – Analysing CBs' experiences (3)

- Avoiding the **risk of confusing "data" and "information"**
  - Need to complement, not replace, designed data with (organic) BD sets
  - Calls for ensuring a continuum: from the collection of BD to statistical processing and the extraction of valuable information for policy use

- Various **ingredients**:
  - Proper IT infrastructure
  - Adequate statistical applications (including big data analytics)
  - Legal and HR support in terms of skill-sets
  - Good co-ordination to have a consistent and holistic information production chain

IV – Analysing CBs' experiences (4)

- Central banks have **already started to rethink** their information management processes to:
  - Be able to access internet data-sets and big data techniques
  - Handle the new data collections initiated after the GFC

- **No one-size-fits-all approach**, as it depends on
  - Characteristics of each data collection
  - Country circumstances
  - Actual policy needs

Annex (1): Selected references

Bank for International Settlements (BIS) (2018). *Cryptocurrencies: looking beyond the hype*, BIS Annual Economic Report, chapter V.

Bean, C. (2016). *Independent review of UK economic statistics*, March.

Bholat, D. (2015). *Big data and central banks*, Bank of England, Quarterly Bulletin, March 2015.

Borio, C. (2013). *The Great Financial Crisis: setting priorities for new statistics*, Journal of Banking Regulation.

Caruana, J. (2017). *International financial crises: new understandings, new data*, Speech at the National Bank of Belgium, Brussels, February.

Cavallo, A., & Rigobon, R. (2016). *The Billion Prices Project: Using Online Prices for Measurement and Research*, Journal of Economic Perspectives, Spring 2016, Vol 30(2): 151-78.

Cœuré, B. (2017). *Policy analysis with big data*, speech at the conference on "Economic and Financial Regulation in the Era of Big Data", Banque de France, Paris, November.

Financial Stability Board (FSB) (2017). *Financial Stability Implications from FinTech*.

Glass, E. (2016): *Survey analysis – Big data in central banks*, Central Banking Focus Report, 2016.

Haldane, A. G. (2018). *Will Big Data Keep Its Promise?* Speech at the Bank of England Data Analytics for Finance and Macro Research Centre, King's Business School, 19 April.

Hammer, C., Kostroch, D., Quiros, G., & Staff of the IMF Statistics Department (STA) Internal Group (2017). *Big data: potential, challenges, and statistical implications*, *IMF Staff Discussion Note*, Staff Discussion Notes (SDN)/17/06, September.

Hill, S. (2018). *The Big Data Revolution in Economic Statistics: Waiting for Godot... and Government Funding*, Goldman Sachs US Economics Analyst, 6 May.

Irving Fisher Committee on Central Bank Statistics (IFC) (2015). *Central banks' use of and interest in 'big data'*, October.

Irving Fisher Committee on Central Bank Statistics (IFC) (2017). *Proceedings of the IFC Satellite Seminar on "Big Data" at the ISI Regional Statistics Conference 2017*, IFC Bulletin, no 44, September.

Meng, X. (2014). *A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it)*, in Lin, X., Genest, C., Banks, D., Molenberghs, G., Scott D., & Wang, J.-L. (eds), *Past, present, and future of statistical science*, Chapman and Hall, 2014, pp 537–62.

Nymand-Andersen, P. (2015). *Big data – the hunt for timely insights and decision certainty: Central banking reflections on the use of big data for policy purposes*, IFC Working Paper, no 14.

The Economist (2017). *The world's most valuable resource is no longer oil, but data*, 6th May edition.

Annex (2): Selected BD projects by central banks

Big data areas	Types of data-sets	Examples of projects
Administrative records	Foreign trade operations / investment transactions	Balance of payments statistics eg tourism, exports
	Taxation / payroll / unemployment insurance	Employment, wages, business formation (SMEs)
	Central balance sheet offices	Performance vulnerabilities assessment
	Loans registers	Measurement of credit risk, FX exposures
	Financial market supervisors	Network analysis, exposures
	Public financial statements	Corporate balance sheet, group-level supervision
	Financial market activity indicators	Payments systems, Trade repositories
Web-based indicators	Internet clicks	Google searches
	social networks	confidence & economic sentiment
	Digitalised content / text	policy communication , analysis of expectations
	Websites' scraping	Various uses
	Job portals	Employment / activity
	Prices posted directly on websites	Measure specific components of the CPI, PPIs, Inflation nowcasting / forecasting, Pricing strategy analysis
	Real estate agencies	House price indices
Commercial data-sets	Credit card operations	Payments patterns, Tourism
	Mobile operators	Mobile positioning data (eg travelers'), Financial inclusion
	Geo spatial information	National statistical system Tasks
Financial data-sets	Credit institutions	Balance sheet exposures, Investor behaviour/expectations
	Settlement operations	Operational risks, Market functioning
	Securities issuance	Security-by-security databases
	Market liquidity	Bid/ask spreads
	Custodians records	Securities holding statistics
	Tick-by-tick data	Real-time analysis of financial patterns

Thank you!!

Questions?

bruno.tissot@bis.org

IFC.secretariat@bis.org

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Machine learning: classification and clustering[1]

Sanjiv R. Das,
Santa Clara University

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Machine Learning: Classification and Clustering

Sanjiv R. Das
Santa Clara University
http://srdas.github.io

Bank of Indonesia
IFC Workshop on "Big Data for Central Bank Policies"
July 2018

Outline, Slides, Code

Machine learning ⊆ artificial intelligence

ARTIFICIAL INTELLIGENCE

Design an intelligent agent that perceives its environment and makes decisions to maximize chances of achieving its goal.
Subfields: vision, robotics, machine learning, natural language processing, planning, …

MACHINE LEARNING

Gives "computers the ability to learn without being explicitly programmed" (Arthur Samuel, 1959)

SUPERVISED LEARNING

Classification, regression

UNSUPERVISED LEARNING

Clustering, dimensionality reduction, recommendation

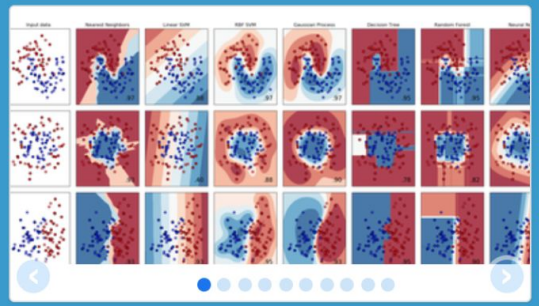
REINFORCEMENT LEARNING

Reward maximization

scikit learn

Home | Installation | Documentation | Examples

Google Custom Search



scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

http://scikit-learn.org/stable/

Classification

Identifying to which category an object belongs to.

Applications: Spam detection, Image recognition.
Algorithms: SVM, nearest neighbors, random forest, … — Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.
Algorithms: SVR, ridge regression, Lasso, … — Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, … — Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency
Algorithms: PCA, feature selection, non-negative matrix factorization. — Examples

Model selection

Comparing, validating and choosing parameters and models.

Goal: Improved accuracy via parameter tuning
Modules: grid search, cross validation, metrics. — Examples

Preprocessing

Feature extraction and normalization.

Application: Transforming input data such as text for use with machine learning algorithms.
Modules: preprocessing, feature extraction. — Examples

Supervised and Unsupervised Learning

1. Machine Learning
2. Linear Models
3. Logistic Regression
4. Discriminant Analysis
5. Bayes Classifier
6. Support Vector Machines
7. Nearest Neighbors (kNN)
8. Decision Trees
9. Random Forest

1. Clustering
   a. K-means
   b. Hierarchical
2. Dimension Reduction
   a. PCA & Factor Analysis
3. Neural Networks and Deep Learning

Ensemble Methods

1. Bagging
2. Stacking
3. Boosting

Small Business Association Loans Dataset

```
#Import the SBA Loans dataset

sba = pd.read_csv("data/SBA.csv")
print(sba.columns)
print(sba.shape)
sba.head()
```

```
Index(['LoanID', 'GrossApproval', 'SBAGuaranteedApproval', 'subpgmdesc',
       'ApprovalFiscalYear', 'InitialInterestRate', 'TermInMonths',
       'ProjectState', 'BusinessType', 'LoanStatus', 'RevolverStatus',
       'JobsSupported'],
      dtype='object')
(527700, 12)
```

The dependent variable is the guaranteed amount as a percentage of the gross loan approved.

```
#Create the dependent variable
Y = sba.SBAGuaranteedApproval.astype("float")/sba.GrossApproval.astype("float")
X = sba[['GrossApproval', 'ApprovalFiscalYear', 'InitialInterestRate', 'TermInMonths',
         'RevolverStatus','JobsSupported']]
```

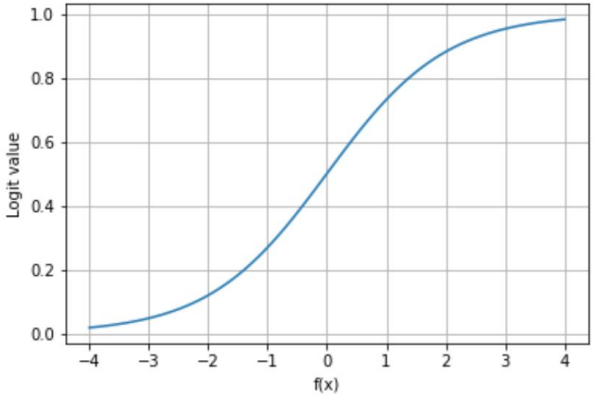
Program code:
http://srdas.github.io/Presentations/ClassClust/Linear_Regression.slides.html#/

Logistic Regression

Limited Dependent Variables

- The dependent variable may be discrete, and could be binomial or multinomial. That is, the dependent variable is limited. In such cases, we need a different approach.
- Discrete dependent variables are a special case of limited dependent variables. The Logit model we look at here is a discrete dependent variable model. Such models are also often called qualitative response (QR) models.

```
#Sigmoid Function
def logit(fx):
    return exp(fx)/(1+exp(fx))

fx = linspace(-4,4,100)
y = logit(fx)
plot(fx,y)
xlabel('f(x)')
ylabel('Logit value')
grid()
```



The Logistic Function

$$y = \frac{1}{1 + e^{-f(x_1, x_2, \ldots, x_n)}} \in (0, 1)$$

where

$$f(x_1, x_2, \ldots, x_n) = a_0 + a_1 x_1 + \ldots + a_n x_n \in (-\infty, +\infty)$$

Program code: http://srdas.github.io/Presentations/ClassClust/LogisticRegression.slides.html#/

Odds Ratio

What are odds ratios? An odds ratio (OR) is the ratio of probability of success to the probability of failure. If the probability of success is p, then

$$OR = \frac{p}{1-p}; \qquad p = \frac{OR}{1+OR}$$

Odds Ratio Coefficients

- In a linear regression, it is easy to see how the dependent variable changes when any right hand side variable changes. Not so with nonlinear models. A little bit of pencil pushing is required (add some calculus too).
- The coefficient of an independent variable in a logit regression tell us by how much the log odds of the dependent variable change with a one unit change in the independent variable. If you want the odds ratio, then simply take the exponentiation of the log odds.

http://srdas.github.io/Presentations/ClassClust/LogisticRegression.slides.html#/12

```
#Example

p = 0.3
OR = p/(1-p)
print('OR old =', OR)

beta = 2
OR_new = OR * exp(beta)
print('OR new =', OR_new)

p_new = OR_new/(1+OR_new)
print('p new =', p_new)
```

```
OR old = 0.4285714285714286
OR new = 3.1667383281131363
p new = 0.7600041276283266
```

Classification Metrics

1. Accuracy: the number of correctly predicted class values.
2. ROC and AUC: The Receiver-Operating Characteristic (ROC) curve is a plot of the True Positive Rate (TPR) against the False Positive Rate (FPR) for different levels of the cut-off posterior probability. This is an essential trade-off in all classification systems.
3. TPR = sensitivity or recall = TP/(TP+FN)
4. FPR = (1 − specificity) = FP/(FP+TN)

1. Precision = $\frac{TP}{TP+FP}$
2. Recall = $\frac{TP}{TP+FN}$
3. F1 score = $\frac{2}{\frac{1}{Precision}+\frac{1}{Recall}}$

(F1 is the harmonic mean of precision and recall.)

Confusion matrix

```
print(metrics.confusion_matrix(y_test, predicted))
print(metrics.classification_report(y_test, predicted))
```

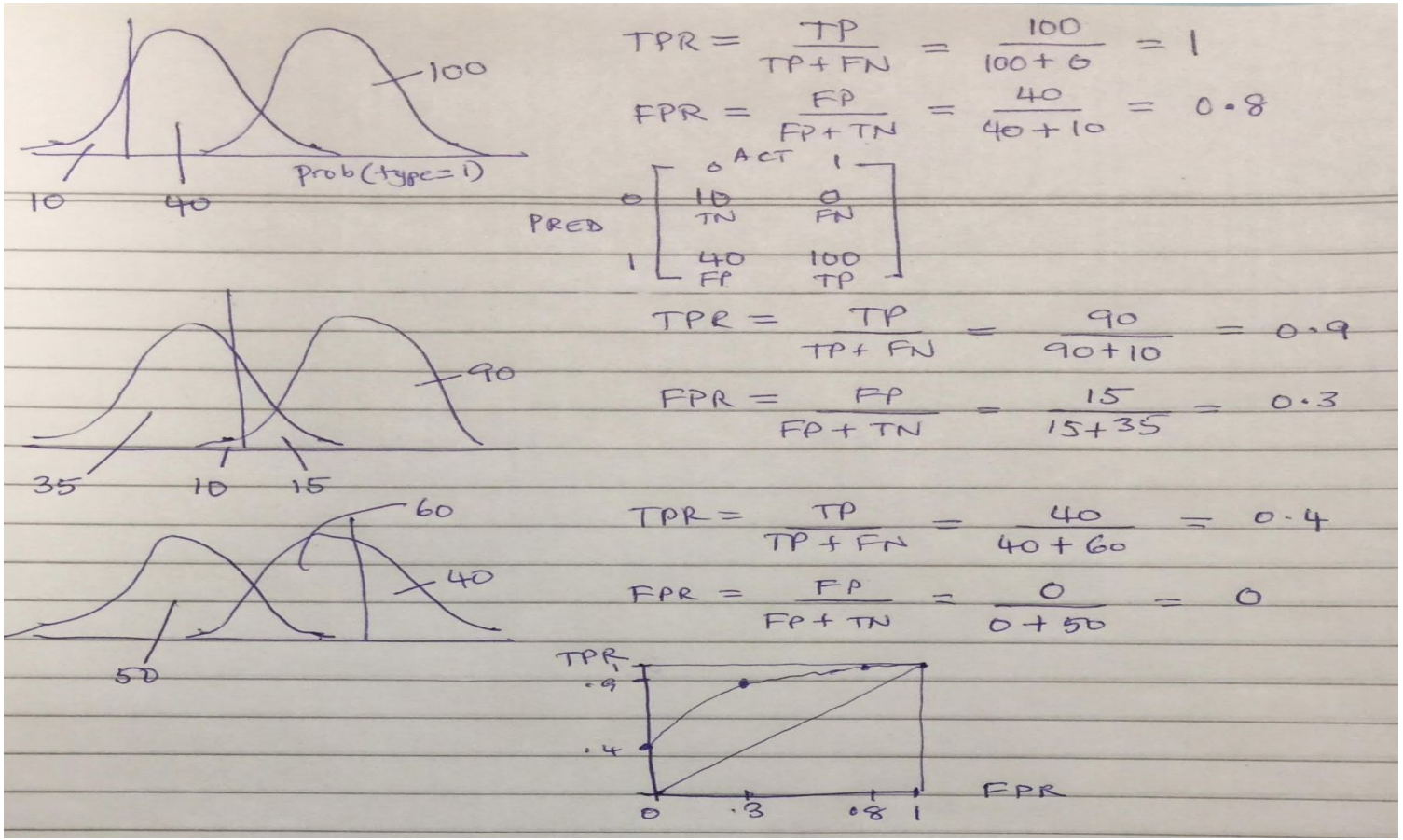
```
[[43753  3435]
 [ 8168 11739]]
             precision    recall  f1-score   support

          0       0.84      0.93      0.88     47188
          1       0.77      0.59      0.67     19907

avg / total       0.82      0.83      0.82     67095
```

http://srdas.github.io/Presentations/ClassClust/LogisticRegression.slides.html#/17

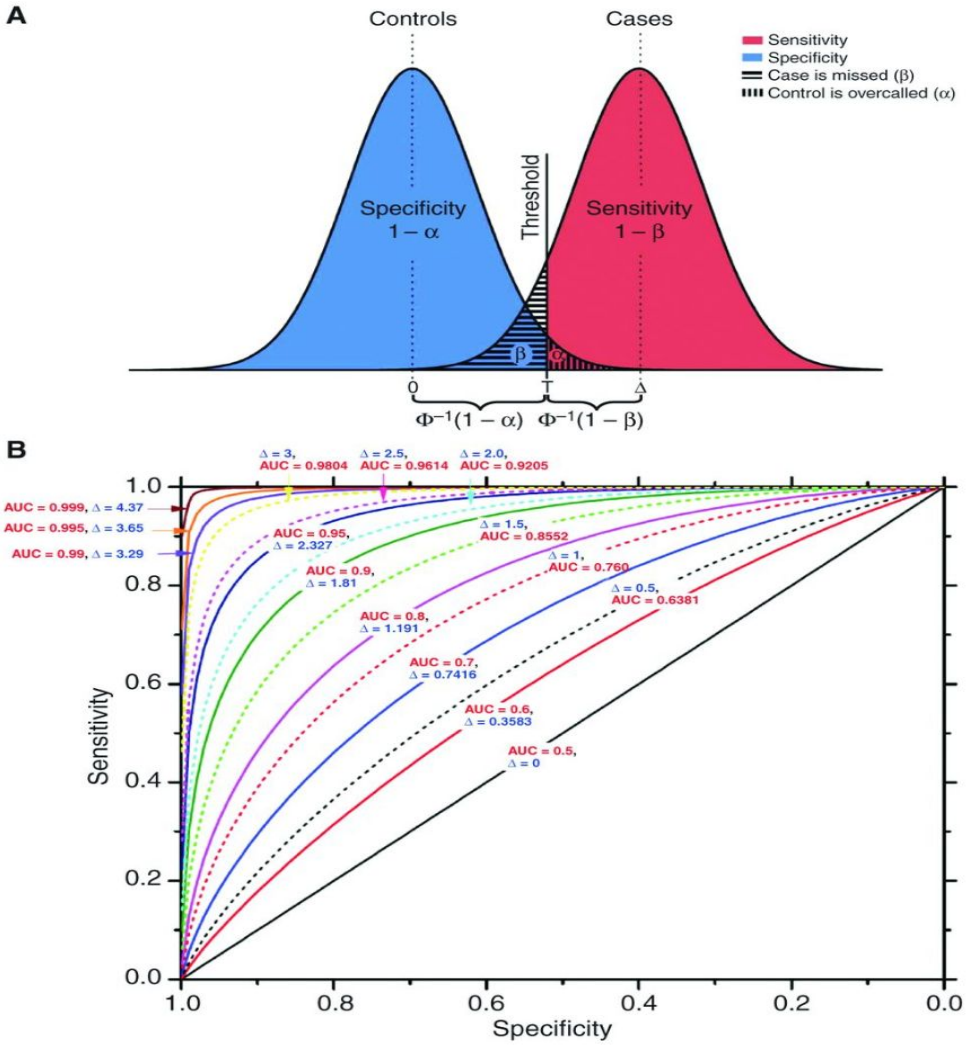
		True condition			
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$
Predicted condition	Predicted condition positive	**True positive**, Power	**False positive**, Type I error	Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	True negative rate (TNR), Specificity (SPC) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$	F_1 score $= \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$

https://en.wikipedia.org/wiki/Receiver_operating_characteristic

ROC Curve


100
Prob(type=1)
10
40
90
35
10
15
60
40
50
TPR = TP / (TP + FN) = 100 / (100 + 0) = 1
FPR = FP / (FP + TN) = 40 / (40 + 10) = 0.8
ACT
PRED
0
1
10 TN
0 FN
40 FP
100 TP
TPR = TP / (TP + FN) = 90 / (90 + 10) = 0.9
FPR = FP / (FP + TN) = 15 / (15 + 35) = 0.3
TPR = TP / (TP + FN) = 40 / (40 + 60) = 0.4
FPR = FP / (FP + TN) = 0 / (0 + 50) = 0
TPR
.9
.4
0
.3
.8
1
FPR

ROC and AUC


A
Controls
Cases
Sensitivity
Specificity
Case is missed (β)
Control is overcalled (α)
Threshold
Specificity
1 − α
Sensitivity
1 − β
β
α
0
T
Δ
Φ⁻¹(1 − α)
Φ⁻¹(1 − β)
B
Δ = 3,
AUC = 0.9804
Δ = 2.5,
AUC = 0.9614
Δ = 2.0,
AUC = 0.9205
AUC = 0.999, Δ = 4.37
AUC = 0.995, Δ = 3.65
AUC = 0.99, Δ = 3.29
AUC = 0.95,
Δ = 2.327
Δ = 1.5,
AUC = 0.8552
AUC = 0.9,
Δ = 1.81
Δ = 1,
AUC = 0.760
Δ = 0.5,
AUC = 0.6381
AUC = 0.8,
Δ = 1.191
AUC = 0.7,
Δ = 0.7416
AUC = 0.6,
Δ = 0.3583
AUC = 0.5,
Δ = 0
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

Multinomial Logit

The probability of each class $(0, 1, \ldots, k)$ for $(k + 1)$ classes is as follows:

$$Pr[y = j] = \frac{e^{a_j^\top x}}{\sum_{i=1}^{k} e^{a_i^\top x}}$$

and

$$Pr[y = 0] = \frac{1}{\sum_{i=1}^{k} e^{a_i^\top x}}$$

Note that $\sum_{i=1}^{k} Pr[y = i] = 1$.

http://srdas.github.io/Presentations/ClassClust/LogisticRegression.slides.html#/28

Discriminant Analysis

Kaggle credit card fraud dataset

Quick Class counts

```
data[["Class","V1"]].groupby(["Class"]).count()
```

	V1
Class	
0	284315
1	492

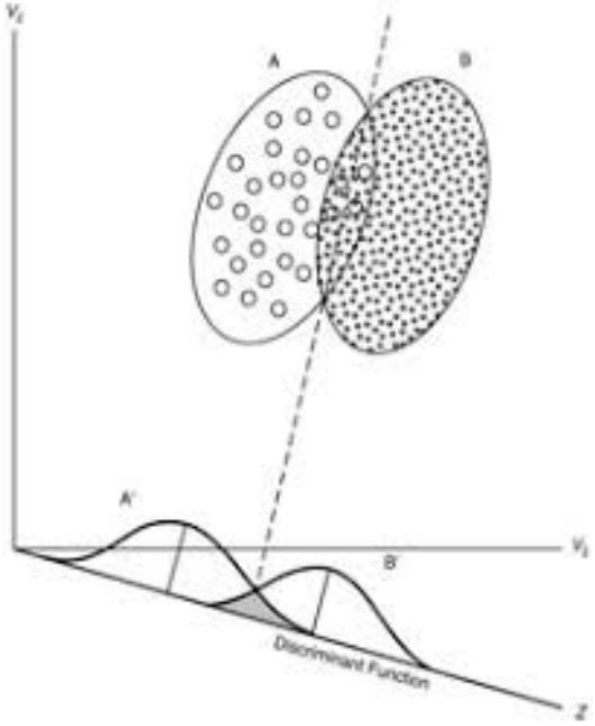
Mean Amount in Each class

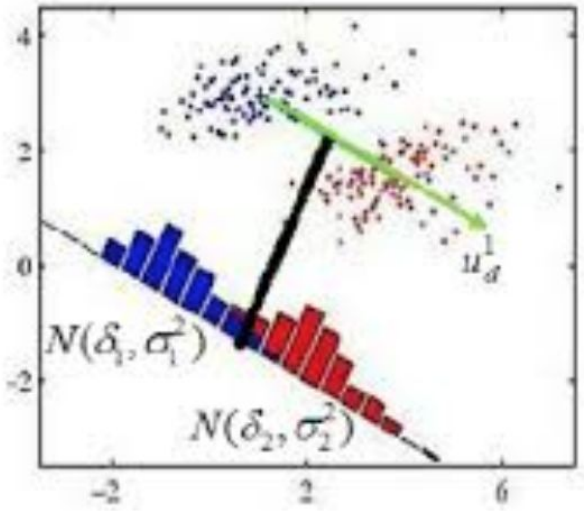
```
data[["Class","Amount"]].groupby(["Class"]).mean()
```

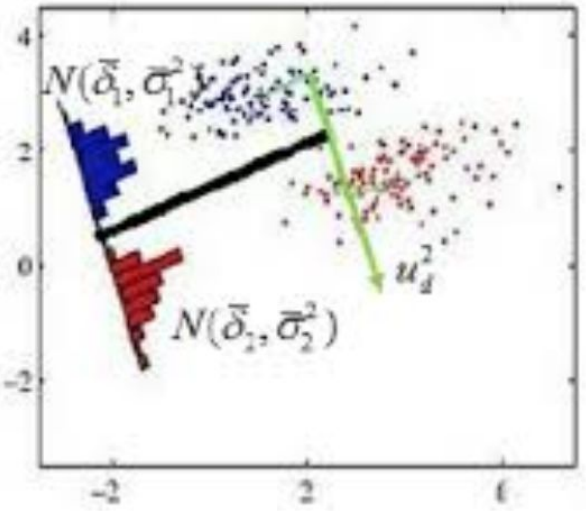
	Amount
Class	
0	88.291022
1	122.211321

http://srdas.github.io/Presentations/ClassClust/Discriminant_Analysis.slides.html#/

Linear Discriminant Analysis







http://srdas.github.io/MLBook/DiscriminantFactorAnalysis.html#discriminant-analysis

NCAA Dataset

http://srdas.github.io/Presentations/ClassClust/Discriminant_Analysis.slides.html#/13

```
ncaa = pd.read_table("data/ncaa.txt")
yy = append(list(ones(32)), list(zeros(32)))
ncaa["y"] = yy
ncaa.head()
```

	No NAME	GMS	PTS	REB	AST	TO	A/T	STL	BLK	PF	
0	1. NorthCarolina	6	84.2	41.5	17.8	12.8	1.39	6.7	3.8	16.7	0.5
1	2. Illinois	6	74.5	34.0	19.0	10.2	1.87	8.0	1.7	16.5	0.4
2	3. Louisville	5	77.4	35.4	13.6	11.0	1.24	5.4	4.2	16.6	0.4
3	4. MichiganState	5	80.8	37.8	13.0	12.6	1.03	8.4	2.4	19.8	0.4
4	5. Arizona	4	79.8	35.0	15.8	14.5	1.09	6.0	6.5	13.3	0.5

Feature Set

```
#CREATE FEATURES
y = ncaa['y']
X = ncaa.iloc[:,2:13]
X.head()
```

	PTS	REB	AST	TO	A/T	STL	BLK	PF	FG	FT	3P
0	84.2	41.5	17.8	12.8	1.39	6.7	3.8	16.7	0.514	0.664	0.417
1	74.5	34.0	19.0	10.2	1.87	8.0	1.7	16.5	0.457	0.753	0.361
2	77.4	35.4	13.6	11.0	1.24	5.4	4.2	16.6	0.479	0.702	0.376
3	80.8	37.8	13.0	12.6	1.03	8.4	2.4	19.8	0.445	0.783	0.329
4	79.8	35.0	15.8	14.5	1.09	6.0	6.5	13.3	0.542	0.759	0.397

Naive Bayes Classifier

Classification based on the class with the highest posterior probability:

$$Pr[C_j|x_1, \ldots, x_n] = \frac{Pr[x_1, \ldots, x_n|C_j] \cdot Pr[C_j]}{\sum_i Pr[x_1, \ldots, x_n|C_i] \cdot Pr[C_i]}$$

and

$$Pr[x_1, \ldots, x_n|C_j] = f[x_1|C_j] \cdot f[x_2|C_j] \cdots f[x_n|C_j]$$

where the last equation encapsulates "naivety", i.e., $x_1, \ldots, x_n$ are independent and Gaussian with density function $f(x) \sim N(\mu_x, \sigma_x^2)$, computed from the raw data.

http://srdas.github.io/Presentations/ClassClust/Naive_Bayes.slides.html#/

Support Vector Machines

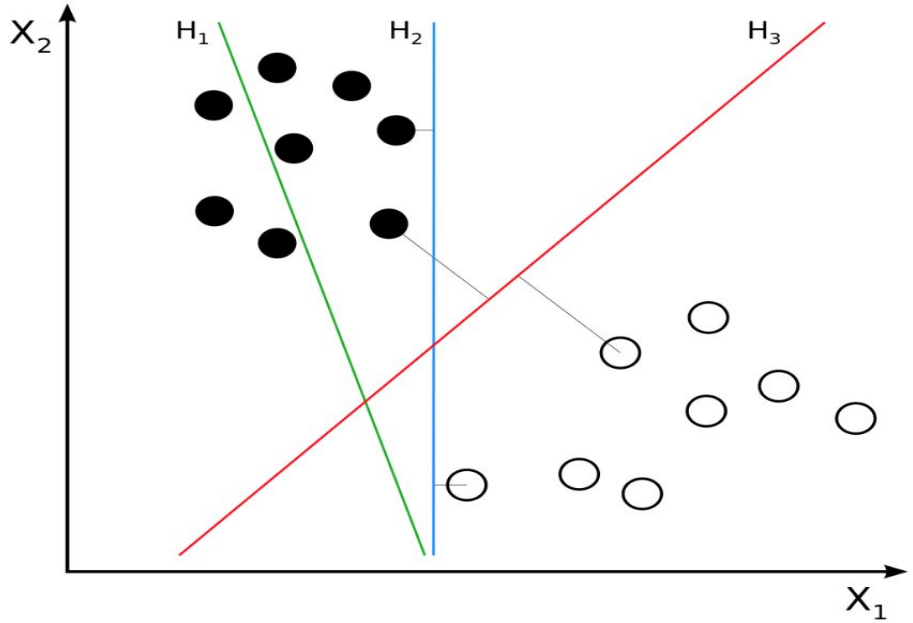
What is a SVM?

The goal of the SVM is to map a set of entities with inputs $X = \{x_1, x_2, \ldots, x_n\}$ of dimension n, i.e., $X \in R^n$, into a set of categories $Y = \{y_1, y_2, \ldots, y_m\}$ of dimension m, such that the n-dimensional X-space is divided using hyperplanes, which result in the maximal separation between classes Y. A hyperplane is the set of points $\mathbf{x}$ satisfying the equation

$$\mathbf{w} \cdot \mathbf{x} = b$$

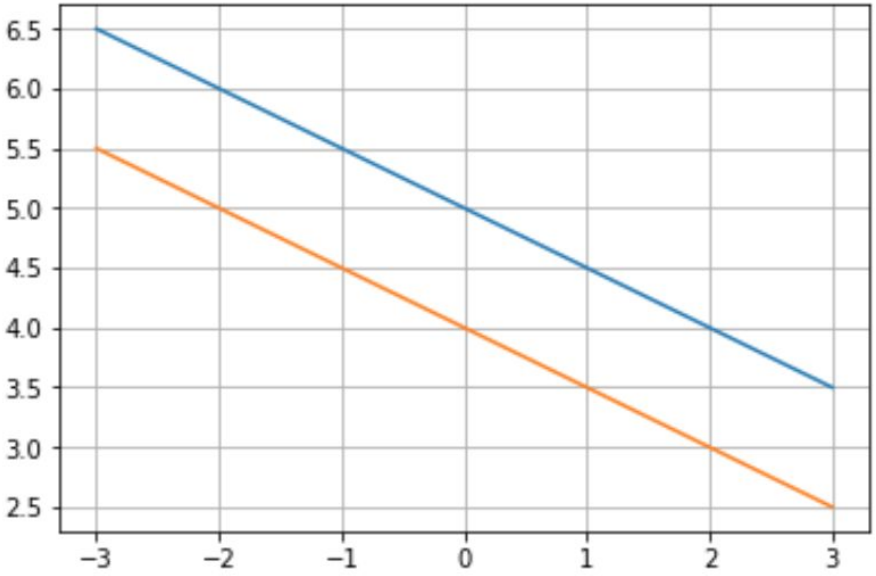
where b is a scalar constant, and $\mathbf{w} \in R^n$ is the normal vector to the hyperplane, i.e., the vector at right angles to the plane. The distance between this hyperplane and $\mathbf{w} \cdot \mathbf{x} = 0$ is given by $b/||\mathbf{w}||$, where $||\mathbf{w}||$ is the norm of vector $\mathbf{w}$.

http://srdas.github.io/Presentations/ClassClust/SVM.slides.html#/



H_3 is the best separating hyperplane.

- Suppose we have two categories of data, i.e., $y = \{y_1, y_2\}$.
- Assume that all points in category y_1 lie above a hyperplane $\mathbf{w} \cdot \mathbf{x} = b_1$, and all points in category y_2 lie below a hyperplane $\mathbf{w} \cdot \mathbf{x} = b_2$.
- Then the distance between the two hyperplanes is $\frac{|b_1 - b_2|}{||\mathbf{w}||}$.




```
#Example of hyperplane geometry
w1 = 1; w2 = 2
b1 = 10
#Plot hyperplane in x1, x2 space
x1 = linspace(-3,3,100)
x2 = (b1-w1*x1)/w2
plot(x1,x2)
#Create hyperplane 2
b2 = 8
x2 = (b2-w1*x1)/w2
plot(x1,x2)
grid()
#Compute distance to hyperplane 2
print('Distance between two hyperplanes =',abs(b1-b2)/sqrt(w1**2+w2**2))
```

```
Distance between two hyperplanes = 0.89442719099999159
```

Regularization

- Of course, there may be no linear hyperplane that perfectly separates the two groups. Hence, L2 regularization.

$$\min_{b_1, b_2, \mathbf{w}, \{\eta_i\}} \frac{1}{2}||\mathbf{w}||^2 + C_1 \sum_{i=1}^{n} \eta_i + C_2 \sum_{i=1}^{n} \eta_i$$

- where C_1, C_2 are the costs for slippage in groups 1 and 2, respectively.
- Often implementations assume $C_1 = C_2$.
- The values η_i are positive for observations that are not perfectly separated, i.e., lead to slippage.

K Nearest Neighbors

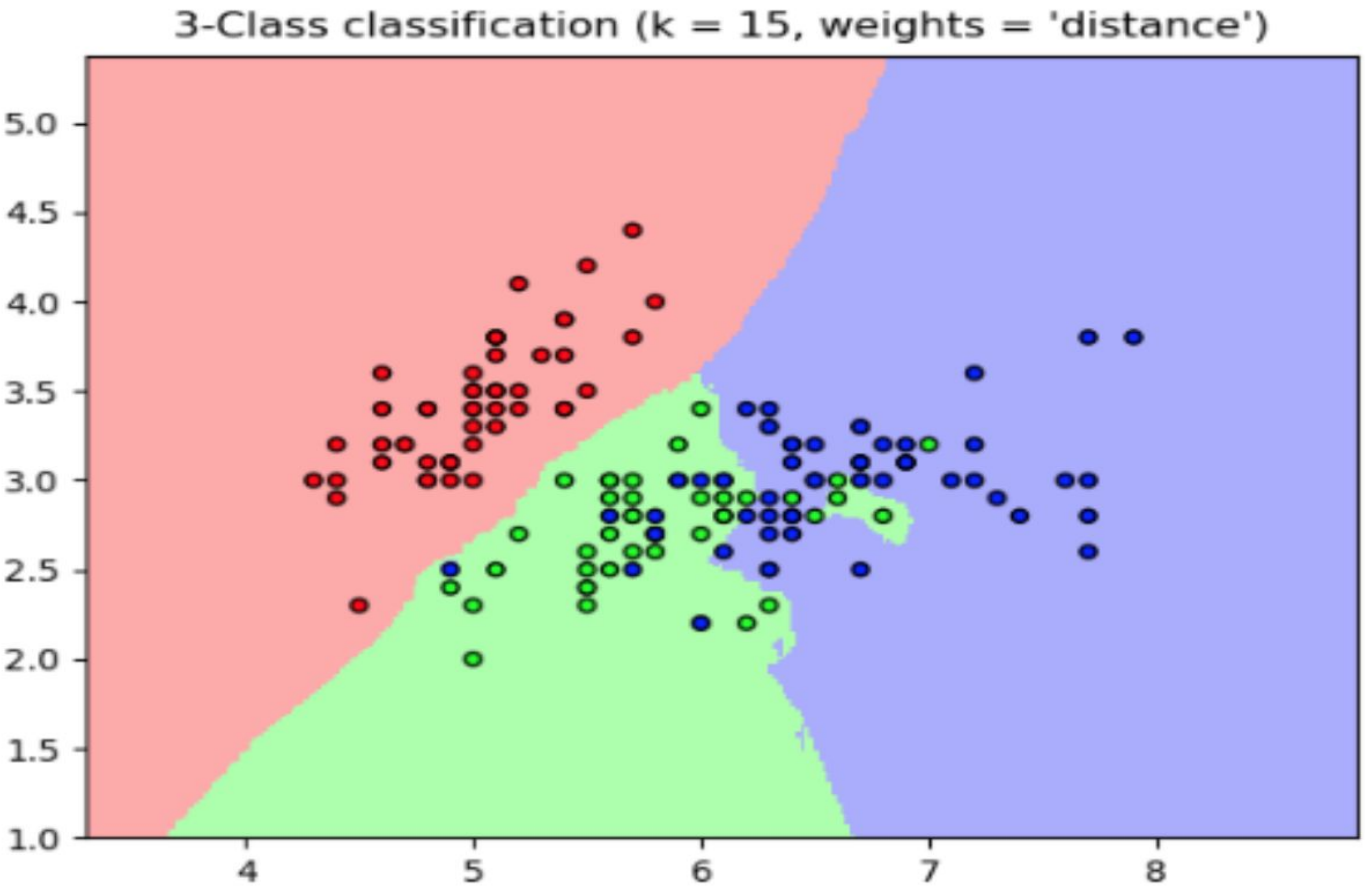
- This is one of the simplest algorithms for classification and grouping.
- Simply define a distance metric over a set of observations, each with M characteristics, i.e., $x_1, x_2, \ldots, x_M$.
- Compute the pairwise distance between each pair of observations, using any of the standard metrics. For example, Euclidian distance between data x and y:

$$d = \sqrt{\sum_{i=1}^{M} (x_i - y_i)^2}$$

- Next, fix k, the number of nearest neighbors in the population to be considered.
- Finally, assign the category based on which one has the majority of nearest neighbors to the case we are trying to classify.

http://srdas.github.io/Presentations/ClassClust/kNN.slides.html#/

Classification



Decision Trees

- A natural outcome of recursive partitioning of the data.
- CART, which stands for classification analysis and regression trees.
- Prediction trees are of two types: (a) Classification trees, where the leaves of the trees are different categories of discrete outcomes. and (b) Regression trees, where the leaves are continuous outcomes.
- We may think of the former as a generalized form of limited dependent variables, and the latter as a generalized form of regression analysis.

http://srdas.github.io/Presentations/ClassClust/Decision_Trees.slides.html#/

Recursive Partitioning

- Bifurcate the data into two categories such that the additional information from categorization results in better **information** than before the binary split.
- Raw frequency p of how many people made donations, i.e., and number between 0 and 1. The **information** of the predicted likelihood p is inversely related to the sum of squared errors (SSE) between this value p and the $x_i = 0, 1$ values of the observations.

$$SSE_1 = \sum_{i=1}^{n} (x_i - p)^2$$

- Second bifurcation:

$$SSE_2 = \sum_{i,Income<K} (x_i - p_L)^2 + \sum_{i,Income \geq K} (x_i - p_R)^2$$

- By choosing K correctly, our recursive partitioning algorithm will maximize the gain, i.e., $\delta = (SSE_1 - SSE_2)$. We stop branching further when at a given tree level δ is less than a pre-specified threshold.

C4.5 Classifier

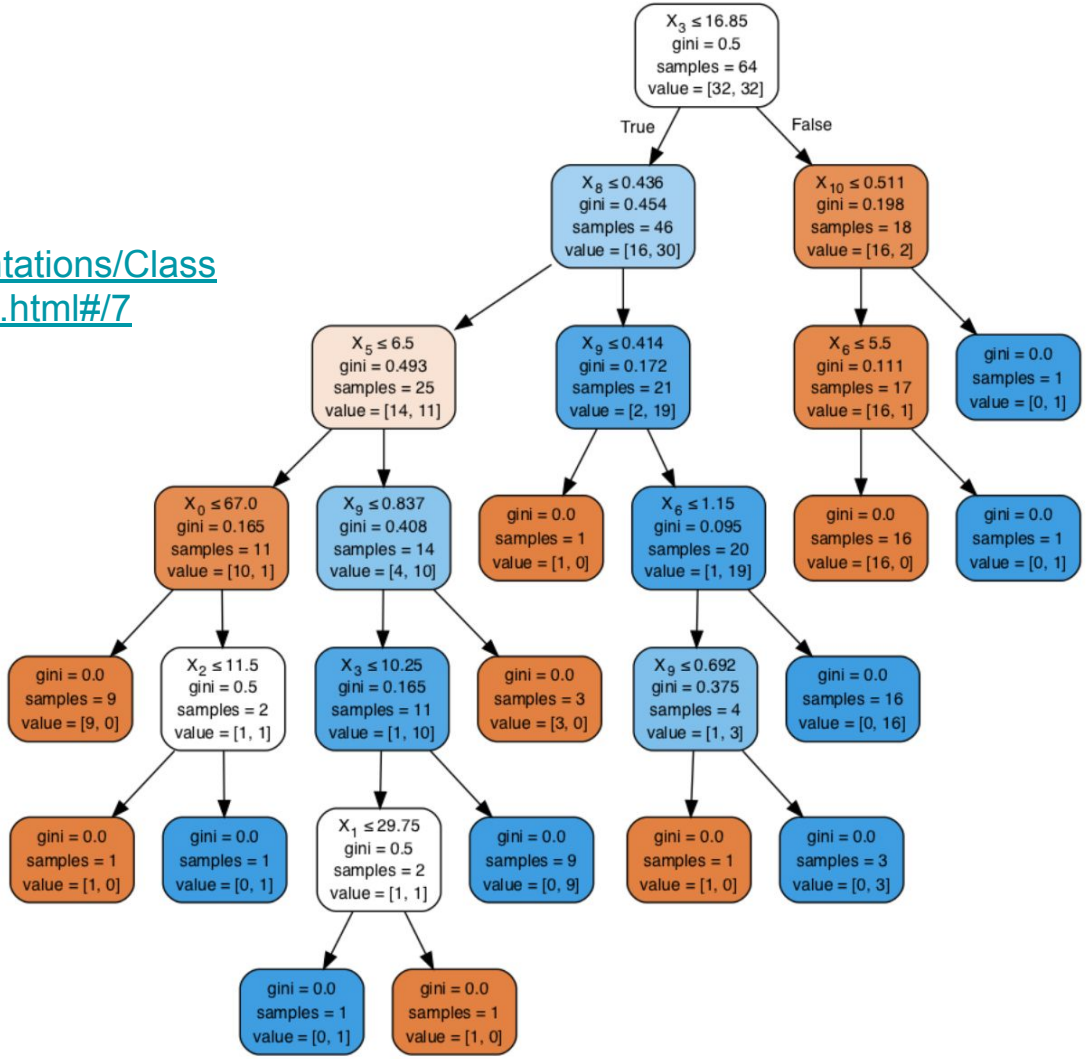
Recursive partitioning as in the previous case, but instead of minimizing the sum of squared errors between the sample data x and the true value p at each level, here the goal is to minimize entropy. This improves the information gain. Natural entropy (H) of the data x is defined as

$$H = -\sum_{x} f(x) \cdot ln\, f(x)$$

where $f(x)$ is the probability density of x. This is intuitive because after the optimal split in recursing down the tree, the distribution of x becomes narrower, lowering entropy. This measure is also often known as "differential entropy."

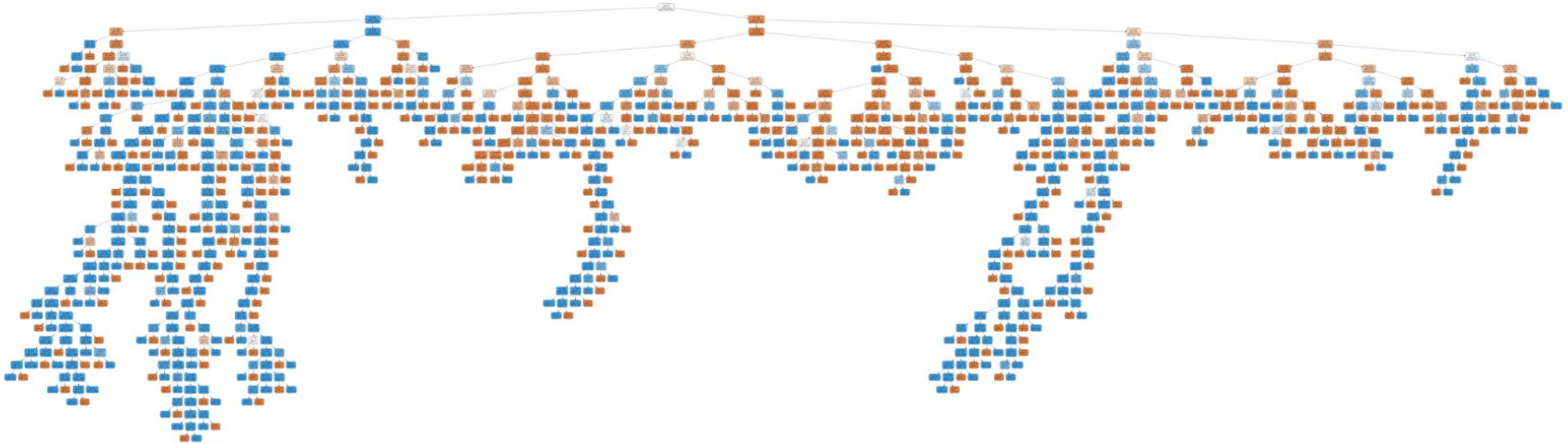
NCAA Tree

http://srdas.github.io/Presentations/ClassClust/Decision_Trees.slides.html#/7



Credit Card Dataset

```
Image(graph.create_png())
```



http://srdas.github.io/Presentations/ClassClust/Decision_Trees.slides.html#/15

Random Forest Classifier

```
%pylab inline
import pandas as pd
from sklearn.model_selection import train_test_split
from imblearn.combine import SMOTEENN
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
from sklearn.metrics import roc_curve,auc
from sklearn.metrics import confusion_matrix
```

```
Populating the interactive namespace from numpy and matplotlib
```

http://srdas.github.io/Presentations/ClassClust/RandomForest_CC_Fraud.slides.html#/

Rebalance sample with under- or over-sampling

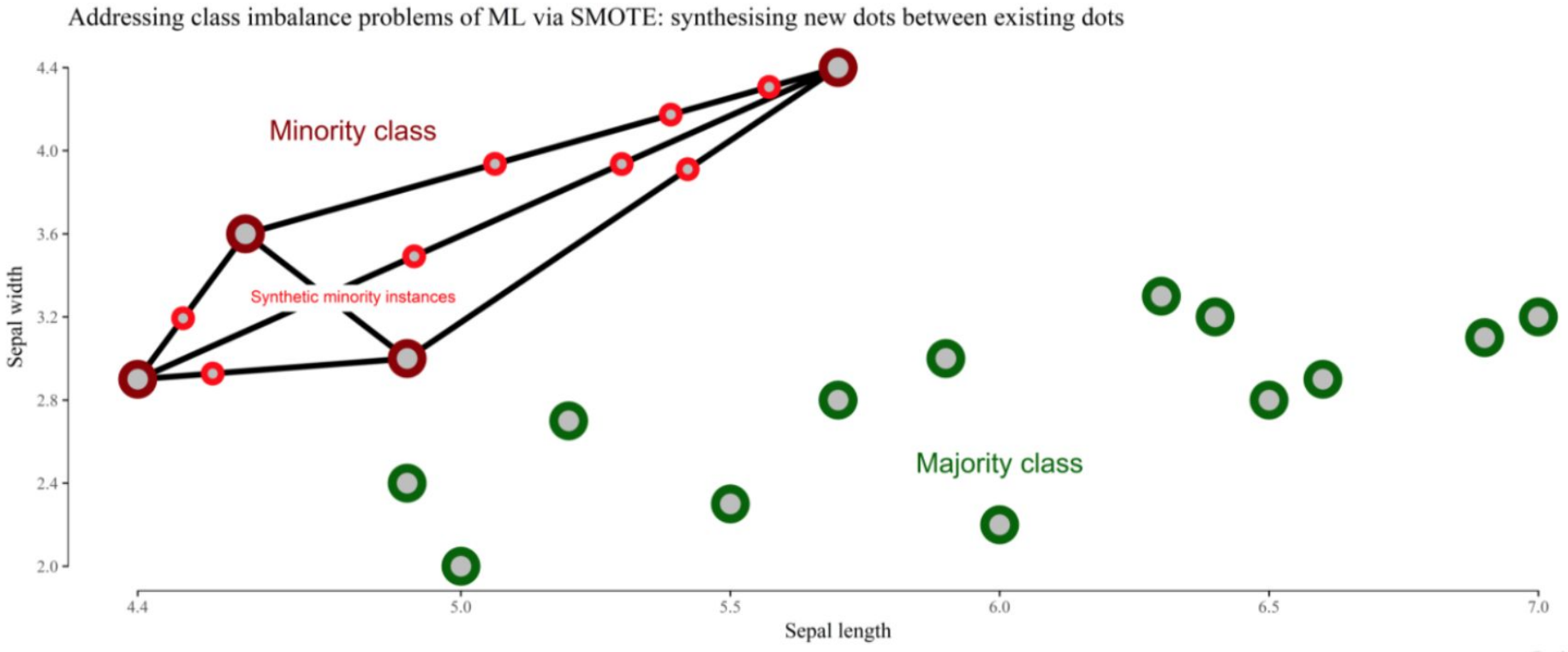
While a Random Forest classifier is generally considered imbalance-agnostic, in this case the severity of the imbalance resuts in overfitting to the majority class.

The Synthetic Minority Over-sampling Technique (SMOTE) is one of the most well-known methods to cope with it and to balance the different number of examples of each class.

The basic idea is to oversample the minority class, while trying to get the most variegated samples from the majority class.

http://srdas.github.io/Presentations/ClassClust/RandomForest_CC_Fraud.slides.html#/7

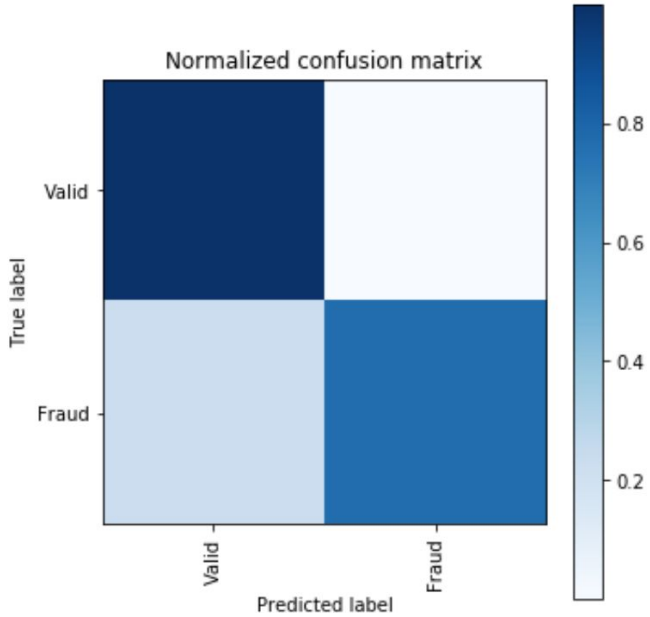
SMOTE



http://rikunert.com/SMOTE_explained

Confusion Matrix

```
cm = confusion_matrix(y_test, y_test_hat)
cm_normalized = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
plt.figure(figsize=(5,5))
plot_confusion_matrix(cm_normalized, title='Normalized confusion matrix')
```



http://srdas.github.io/Presentations/ClassClust/RandomForest_CC_Fraud.slides.html#/21

Dimension Reduction

A Matrix Reduction

Suppose we reduce the $k = 11$ dimensional feature space X to reduced factor space R with $k = 3$. We translate with a matrix L.

$$R = X \cdot L$$

where F is (64×3), X is (64×11), and L is (11×3).

Where does matrix L come from?

- From Principal Components Analysis.
- Based on an Eigenvalue Decomposition of the covariance matrix of the features, i.e., $C = Cov(X)$, which is size (11×11).
- Decomposition is based on solving the following equation:

$$\lambda l = C \cdot l$$

- There are 11 solutions l and the first 3 will form the matrix L.

http://srdas.github.io/Presentations/ClassClust/Dimension_Reduction.slides.html#/

Principal Components Analysis (PCA)

```
#REDUCED DATA
from sklearn import decomposition
pca = decomposition.PCA(n_components=3)
pca.fit(Xs)
R = pca.transform(Xs)
print(R.shape)
```

```
(64, 3)
```

http://srdas.github.io/Presentations/ClassClust/Dimension_Reduction.slides.html#/7

```
#LOADINGS MATRIX L
L = pca.components_.T
print(L.shape)
print(X.columns)
L
```

```
(11, 3)
Index(['PTS  ', 'REB  ', 'AST  ', 'TO  ', 'A/T  ', 'STL  ', 'BLK  ', 'PF  ',
       'FG  ', 'FT  ', '3P'],
      dtype='object')

array([[-0.43884425,  0.02285078, -0.18631231],
       [ 0.1867903 , -0.43294301, -0.21835274],
       [-0.47238137,  0.04962787, -0.18252437],
       [ 0.17651088, -0.02325077, -0.68627945],
       [-0.45266018,  0.08602947,  0.37681287],
       [ 0.03888779,  0.57289362, -0.29086594],
       [ 0.02703794,  0.08087251, -0.16285993],
       [ 0.05607815,  0.51652087, -0.12761847],
       [-0.44993945, -0.18657986, -0.21900567],
       [ 0.03599791,  0.40620447,  0.17479897],
       [-0.32279124, -0.01667957, -0.25572689]])
```

```
#CHECK THAT DECOMPOSITION IS CORRECT
sum(R - Xs.dot(L))
```

```
0   -1.893624e-14
1    3.152340e-14
2    1.458902e-15
dtype: float64
```

Treasury Rates Dataset

```
rates = pd.read_table("data/tryrates.txt")
print(rates.shape)
rates.head()
```

```
(367, 9)
```

	DATE	FYGM3	FYGM6	FYGT1	FYGT2	FYGT3	FYGT5	FYGT7	FYGT10
0	Jun-76	5.41	5.77	6.52	7.06	7.31	7.61	7.75	7.86
1	Jul-76	5.23	5.53	6.20	6.85	7.12	7.49	7.70	7.83
2	Aug-76	5.14	5.40	6.00	6.63	6.86	7.31	7.58	7.77
3	Sep-76	5.08	5.30	5.84	6.42	6.66	7.13	7.41	7.59
4	Oct-76	4.92	5.06	5.50	5.98	6.24	6.75	7.16	7.41

http://srdas.github.io/Presentations/ClassClust/Dimension_Reduction.slides.html#/11

```
X.corr()
```

	FYGM3	FYGM6	FYGT1	FYGT2	FYGT3	FYGT5	FYGT7	FYGT10
FYGM3	1.000000	0.997537	0.991125	0.975089	0.961225	0.938329	0.922041	0.906564
FYGM6	0.997537	1.000000	0.997350	0.985125	0.972844	0.951266	0.935603	0.920542
FYGT1	0.991125	0.997350	1.000000	0.993696	0.984692	0.966859	0.953130	0.939686
FYGT2	0.975089	0.985125	0.993696	1.000000	0.997767	0.987892	0.978651	0.968093
FYGT3	0.961225	0.972844	0.984692	0.997767	1.000000	0.995622	0.989403	0.981307
FYGT5	0.938329	0.951266	0.966859	0.987892	0.995622	1.000000	0.998435	0.994569
FYGT7	0.922041	0.935603	0.953130	0.978651	0.989403	0.998435	1.000000	0.998493
FYGT10	0.906564	0.920542	0.939686	0.968093	0.981307	0.994569	0.998493	1.000000

```
#PCA
X = rates.drop("DATE",axis=1)
pca = decomposition.PCA(n_components=2)
pca.fit(X)
Y = pca.transform(X)
Y.shape
```

```
(367, 2)
```

Clustering

Overview

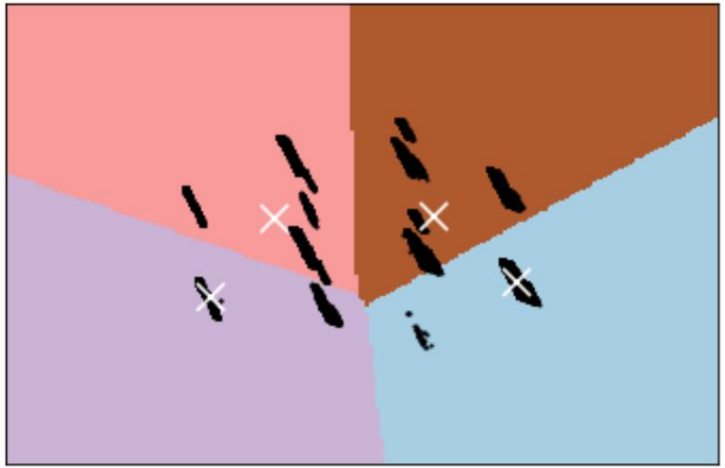
- Grouping individuals, firms, projects, etc.
- Cluster analysis comprises a group of techniques that uses distance metrics to bunch data into categories.
- Two approaches.

1. **Partitioning or Top-down**: In this approach, the entire set of n entities is assumed to be divided into k clusters. Then entities are assigned clusters.
2. **Agglomerative or Hierarchical or Bottom-up**: In this case we begin with all entities in the analysis being given their own cluster, so that we start with n clusters. Then, entities are grouped into clusters based on a given distance metric between each pair of entities. In this way a hierarchy of clusters is built up and the researcher can choose which grouping is preferred.

http://srdas.github.io/Presentations/ClassClust/Clustering.slides.html#/

K-means

1. Form a distance matrix.
2. Initialize cluster centroids with evenly spaced items.
3. Assign each observation to closest cluster (to centroid, closest or farthest mamber).
4. Repeat a few times to re-assign until the scheme stabilizes.



http://srdas.github.io/Presentations/ClassClust/Clustering.slides.html#/5

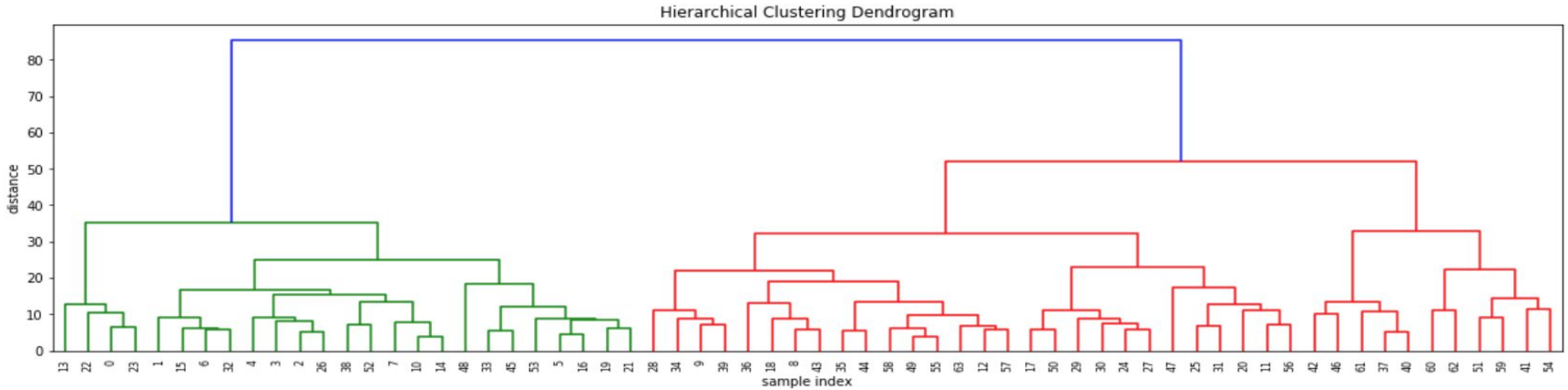
Hierarchical Clustering

1. Get distance matrix for n observations. Each in its own cluster.
2. Club the two closest observations into a cluster. Now we have $(n-1)$ clusters.
3. Recalculate centroids.
4. Repeat to get hierarchical structure.

http://srdas.github.io/Presentations/ClassClust/Clustering.slides.html#/13

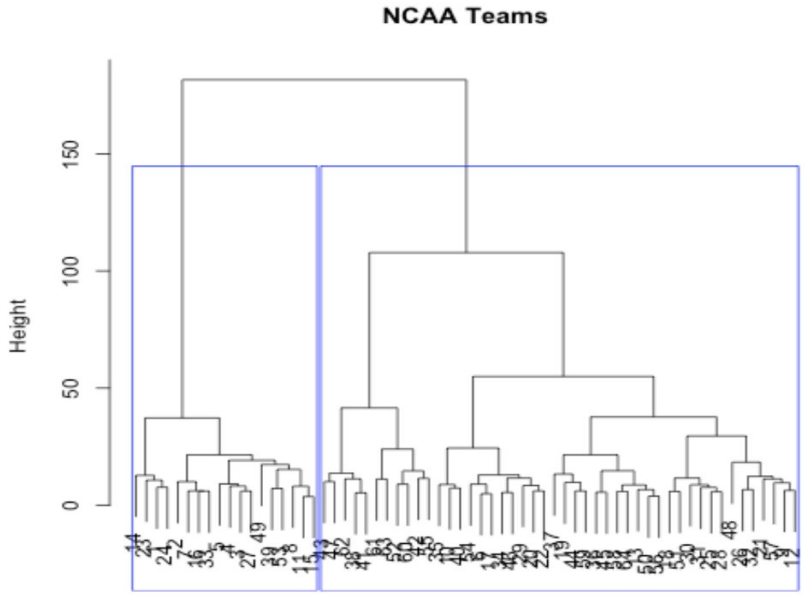
Dendrogram

```
#DENDROGRAM
figure(figsize=(20, 5))
title('Hierarchical Clustering Dendrogram')
xlabel('sample index')
ylabel('distance')
dendrogram(Z,
    leaf_rotation=90.,  # rotates the x axis labels
    leaf_font_size=8.,  # font size for the x axis labels
)
show()
```

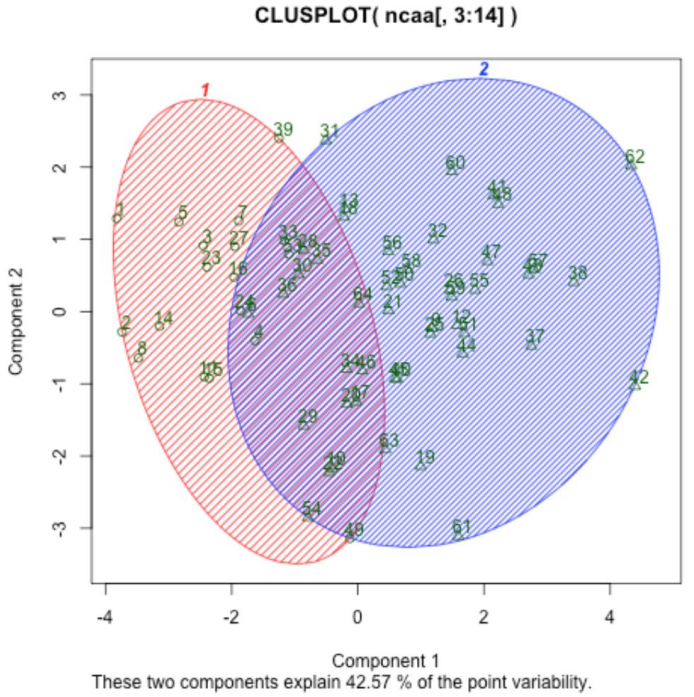


Hierarchical Clustering in R

```
%%R
#CLUSTER
ncaa = read.table("data/ncaa.txt",header=TRUE)
d = dist(ncaa[,3:14], method="euclidian")
fit = hclust(d, method="ward.D")
plot(fit,main="NCAA Teams")
groups = cutree(fit, k=2)
rect.hclust(fit, k=2, border="blue")
```




```
%%R
#CLUSTER PLOT
library(cluster)
clusplot(ncaa[,3:14],groups,color=TRUE,shade=TRUE,
```



http://srdas.github.io/Presentations/ClassClust/Clustering.slides.html#/19

Deep Learning with Neural Networks

Deep Learning is Pattern Recognition

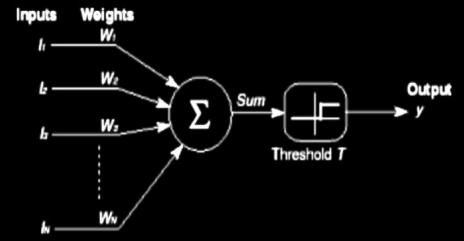


Figure 1.1: McCullough Pitts neuron

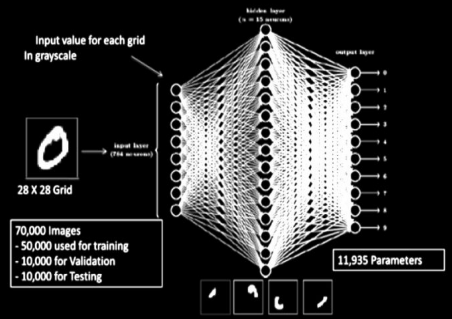


Figure 2.2: The NN used for Solving the MNIST Digit Recognition Problem

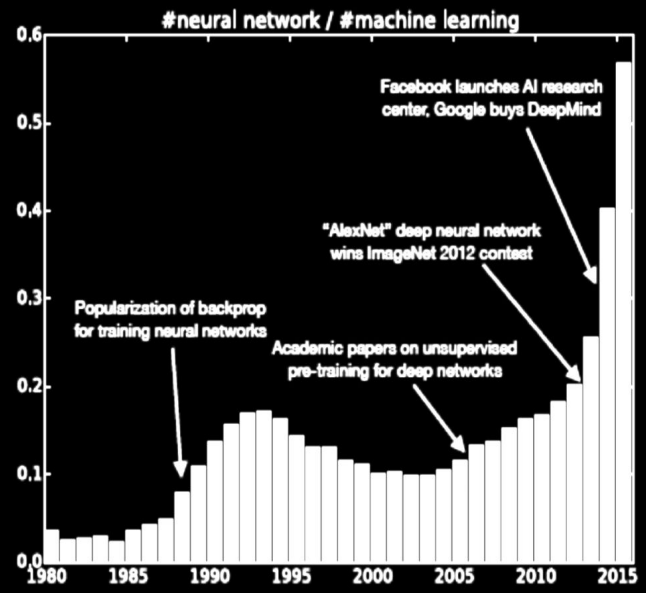
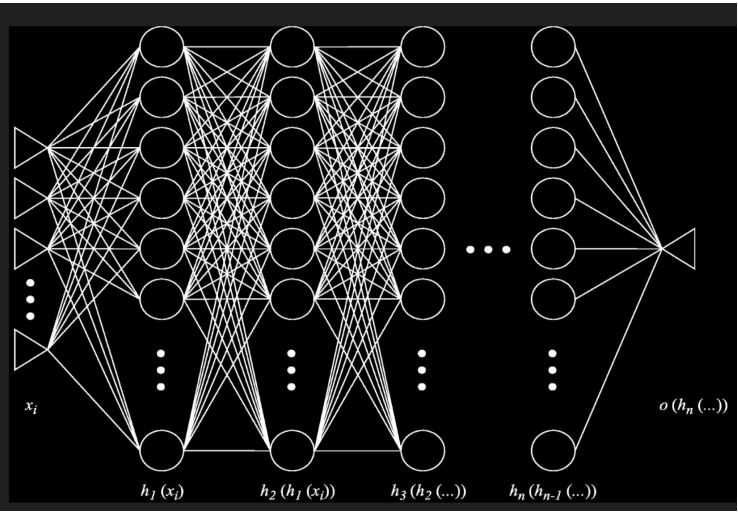


Figure 1.2: Number of NN research reports

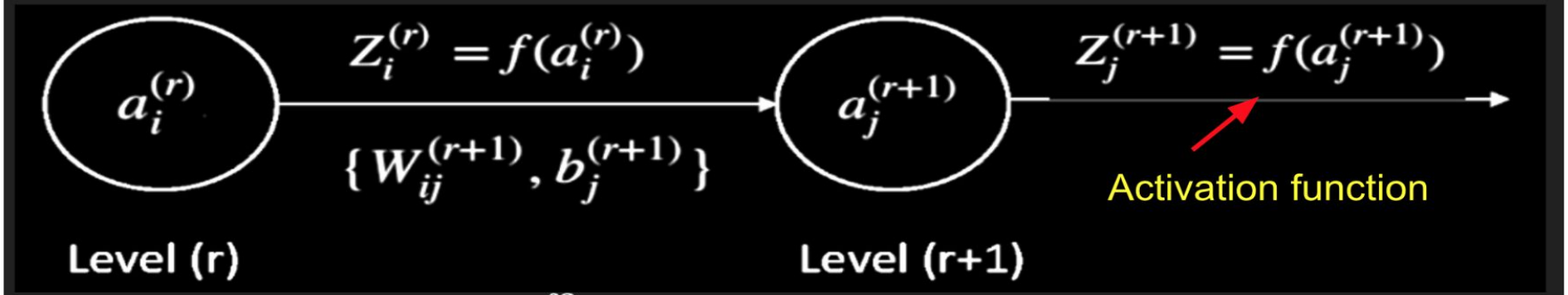
Reference Book:
http://srdas.github.io/DLBook



http://srdas.github.io/Presentations/ClassClust/DeepLearning_Introduction_Short.slides.html#/

Zooming In

Subset of the Net

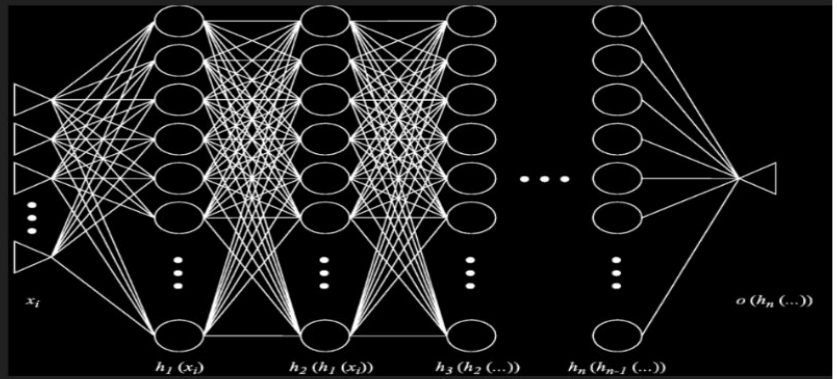


$$a_j = b_j + \sum_{I=1}^{n} w_{ij} z_i$$

Net input

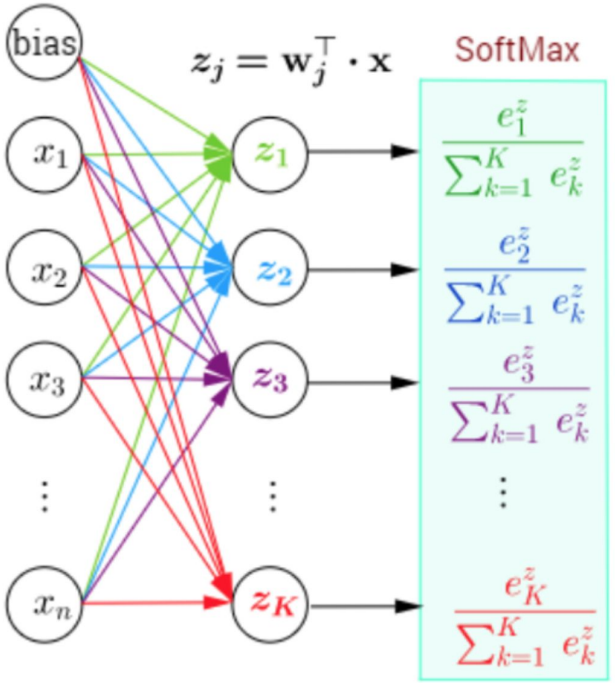
$$\sigma(a_j) = \frac{1}{1 + \exp(-a_j)}$$

Sigmoid function



Activation Functions

Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	Multi-layer NN	



Loss Function

Fitting the DLN is an exercise where the best weights $\{W, b\} = \{W_{ij}^{(r+1)}, b_j^{(r+1)}\}, \forall r$ for all layers are determined to minimize a loss function generally denoted as

$$\min_{W,b} \sum_{m=1}^{M} L_m[h(X^{(m)}), T^{(m)}]$$

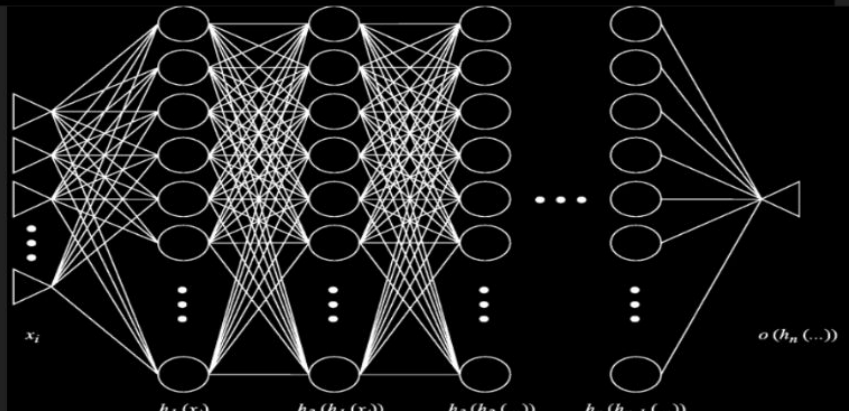
where M is the number of training observations (rows in the data set), $T^{(m)}$ is the true value of the output, and $h(X^{(m)})$ is the model output from the DLN. The loss function L_m quantifies the difference between the model output and the true output.

Cross Entropy

$$C = -\frac{1}{n} \sum_x [y \ln a + (1 - y) \ln(1 - a)]$$

Quadratic Loss

$$C = \frac{(y - a)^2}{2}$$



Gradient Descent

Fitting the DLN requires getting the weights $\{W, b\}$ that minimize L_m. These are done using gradient descent, i.e.,

$$W_{ij}^{(r+1)} \leftarrow W_{ij}^{(r+1)} - \eta \cdot \frac{\partial L_m}{\partial W_{ij}^{(r+1)}}$$

$$b_j^{(r+1)} \leftarrow b_j^{(r+1)} - \eta \cdot \frac{\partial L_m}{\partial b_j^{(r+1)}}$$

Here η is the learning rate parameter. We iterate on these functions until the gradients become zero, and the weights discontinue changing with each update, also known as an **epoch**.

Total Gradient

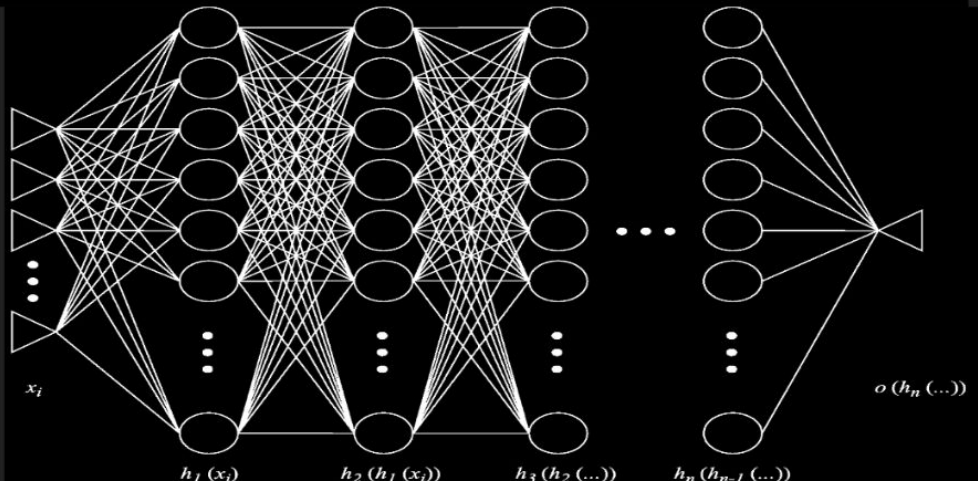
$$w_{ij}^{(r)} \leftarrow w_{ij}^{(r)} - \frac{\eta}{M} \sum_{m=1}^{M} \frac{\partial L_m(w, b; X(m), T(m))}{\partial w_{ij}^{(r)}}$$

$$b_i^{(r)} \leftarrow b_i^{(r)} - \frac{\eta}{M} \sum_{m=1}^{M} \frac{\partial L_m(w, b; X(m), T(m))}{\partial b_i^{(r)}}$$

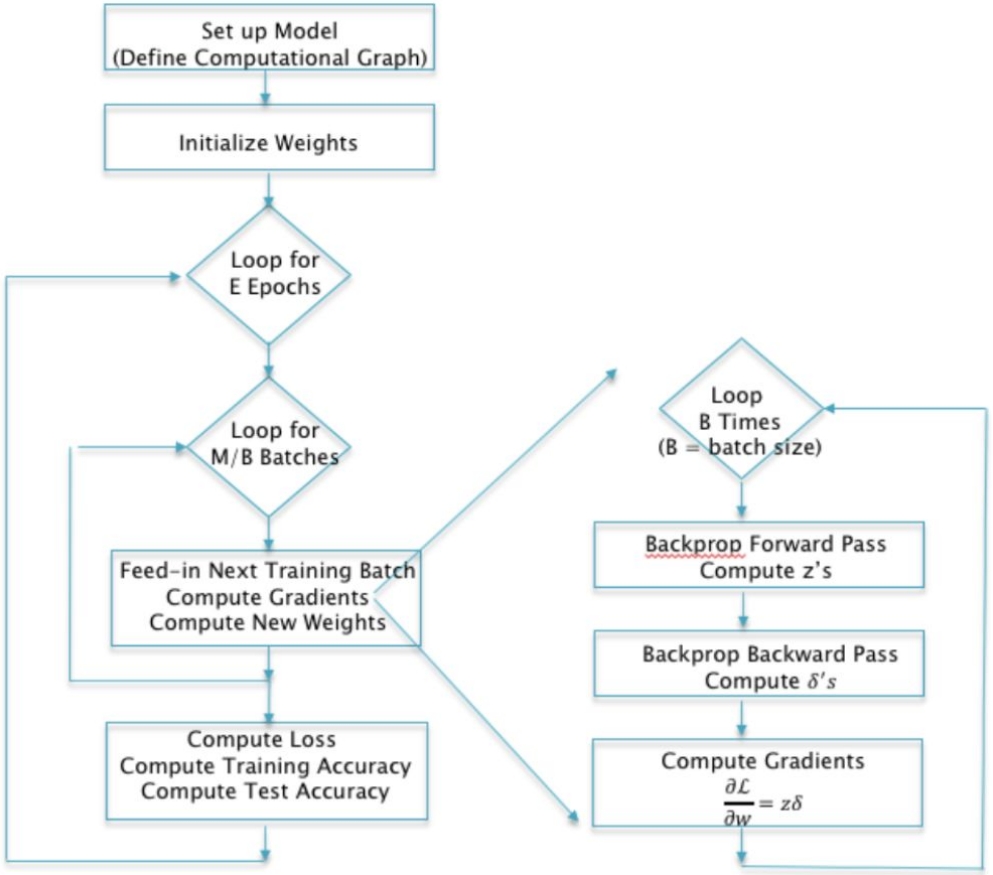
Stochastic Batch Gradient

$$w_{ij}^{(r)} \leftarrow w_{ij}^{(r)} - \frac{\eta}{s} \sum_{m=1}^{s} \frac{\partial L_m(w, b; X(m), T(m))}{\partial w_{ij}^{(r)}}$$

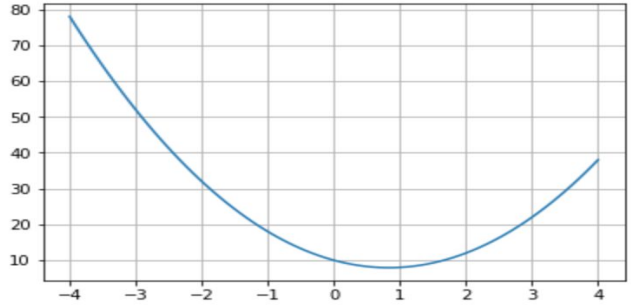
$$b_i^{(r)} \leftarrow b_i^{(r)} - \frac{\eta}{s} \sum_{m=1}^{s} \frac{\partial L_m(w, b; X(m), T(m))}{\partial b_i^{(r)}}$$



Batch Gradient Descent



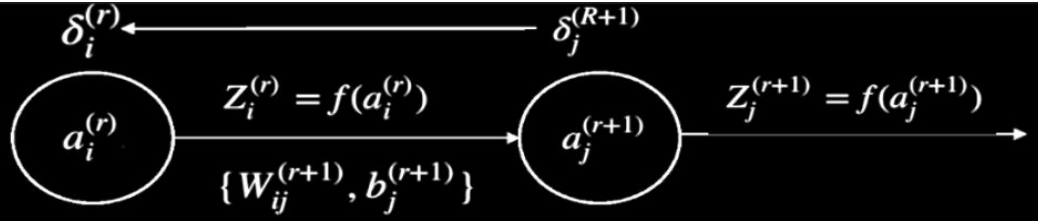

```
def f(x):
    return 3*x**2 -5*x + 10

x = linspace(-4,4,100)
plot(x,f(x))
grid()
```




```
dx = 0.001
eta = 0.05  #learning rate
x = -3
for j in range(20):
    df_dx = (f(x+dx)-f(x))/dx
    x = x - eta*df_dx
    print(x,f(x))
```

```
-1.8501500000001698 29.519915067502733
-1.0452550000002532 18.503949045077853
-0.4818285000003115 13.105618610239208
-0.08742995000019249 10.460081738472072
0.18864903499989083 9.163520200219716
0.3819043244999847 8.528031116715445
0.5171830271499616 8.216519714966186
0.6118781190049631 8.06379390252634
0.6781646833034642 7.988898596522943
0.7245652783124417 7.95215813604575
0.7570456948186948 7.934126078037087
0.7797819863730542 7.925269906950447
0.7956973904611502 7.920916059254301
0.8068381733227685 7.918772647178622
0.8146367213259147 7.917715356568335
0.8200957049281836 7.917192371084045
0.8239169934497053 7.916932669037079
0.8265918954147882 7.9168030076222955
0.8284643267903373 7.916737788340814
0.8297750287532573 7.916704651261121
```

Magic of Backpropagation



$$\min_{W,b} \sum_{m=1}^{M} L_m[h(X^{(m)}), T^{(m)}]$$

Loss Function

$$a_j^{(r+1)} = \sum_{i=1}^{n_r} W_{ij}^{(r+1)} Z_i^r + b_j^{(r+1)} = \sum_{i=1}^{n_r} W_{ij}^{(r+1)} f(a_i^{(r)}) + b_j^{(r+1)} \Rightarrow \frac{\partial a_j^{(r+1)}}{\partial a_i^{(r)}} = W_{ij}^{(r+1)} \cdot f'(a_i^{(r)})$$

Gradients

$$\frac{\partial L_m}{\partial W_{ij}^{(r+1)}} = \frac{\partial L_m}{\partial a_j^{(r+1)}} \cdot \frac{\partial a_j^{(r+1)}}{\partial W_{ij}^{(r+1)}} = \delta_j^{(r+1)} \cdot Z_i^{(r)}$$

$$\frac{\partial L_m}{\partial b_j^{(r+1)}} = \frac{\partial L_m}{\partial a_j^{(r+1)}} \cdot \frac{\partial a_j^{(r+1)}}{\partial b_j^{(r+1)}} = \delta_j^{(r+1)} \cdot 1 = \delta_j^{(r+1)}$$

$$\begin{aligned}
\delta_i^{(r)} &= \frac{\partial L_m}{\partial a_i^{(r)}} \\
&= \sum_{j=1}^{n_{r+1}} \frac{\partial L_m}{\partial a_j^{(r+1)}} \cdot \frac{\partial a_j^{(r+1)}}{\partial a_i^{(r)}} \\
&= \sum_{j=1}^{n_{r+1}} \delta_j^{(r+1)} \cdot W_{ij}^{(r+1)} \cdot f'(a_i^{(r)}) \\
&= f'(a_i^{(r)}) \cdot \sum_{j=1}^{n_{r+1}} \delta_j^{(r+1)} \cdot W_{ij}^{(r+1)}
\end{aligned}$$

Deltas

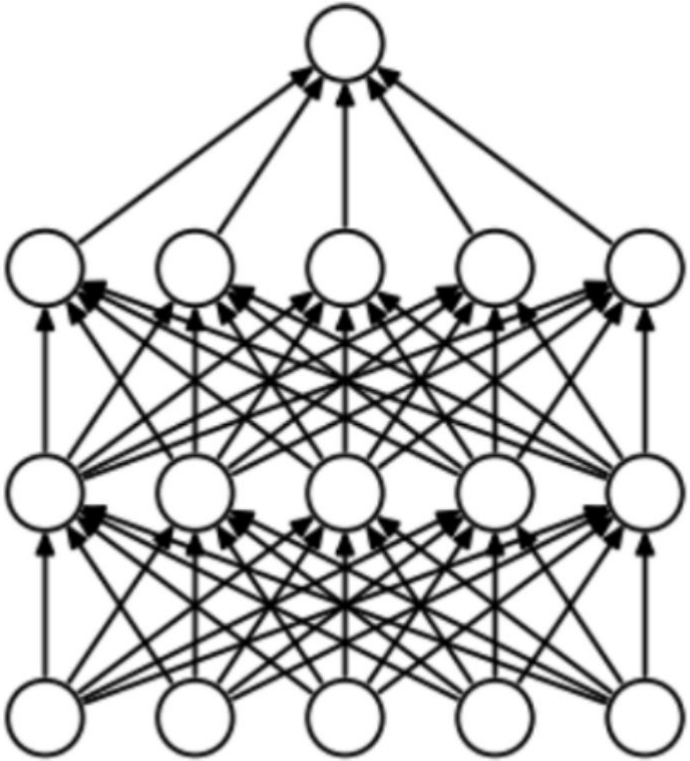
$$\delta_j^{(R+1)} = \frac{\partial L_m}{\partial a_j^{(R+1)}} = h'(a_j^{(R+1)})$$

Gradient Descent

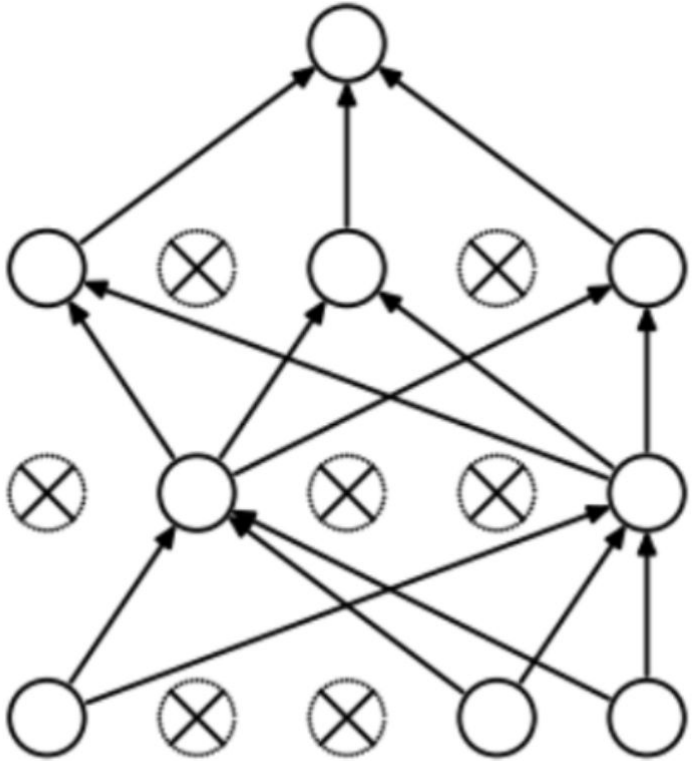
$$W_{ij}^{(r+1)} \leftarrow W_{ij}^{(r+1)} - \eta \cdot \frac{\partial L_m}{\partial W_{ij}^{(r+1)}}$$

$$b_j^{(r+1)} \leftarrow b_j^{(r+1)} - \eta \cdot \frac{\partial L_m}{\partial b_j^{(r+1)}}$$

Dropout Regularization



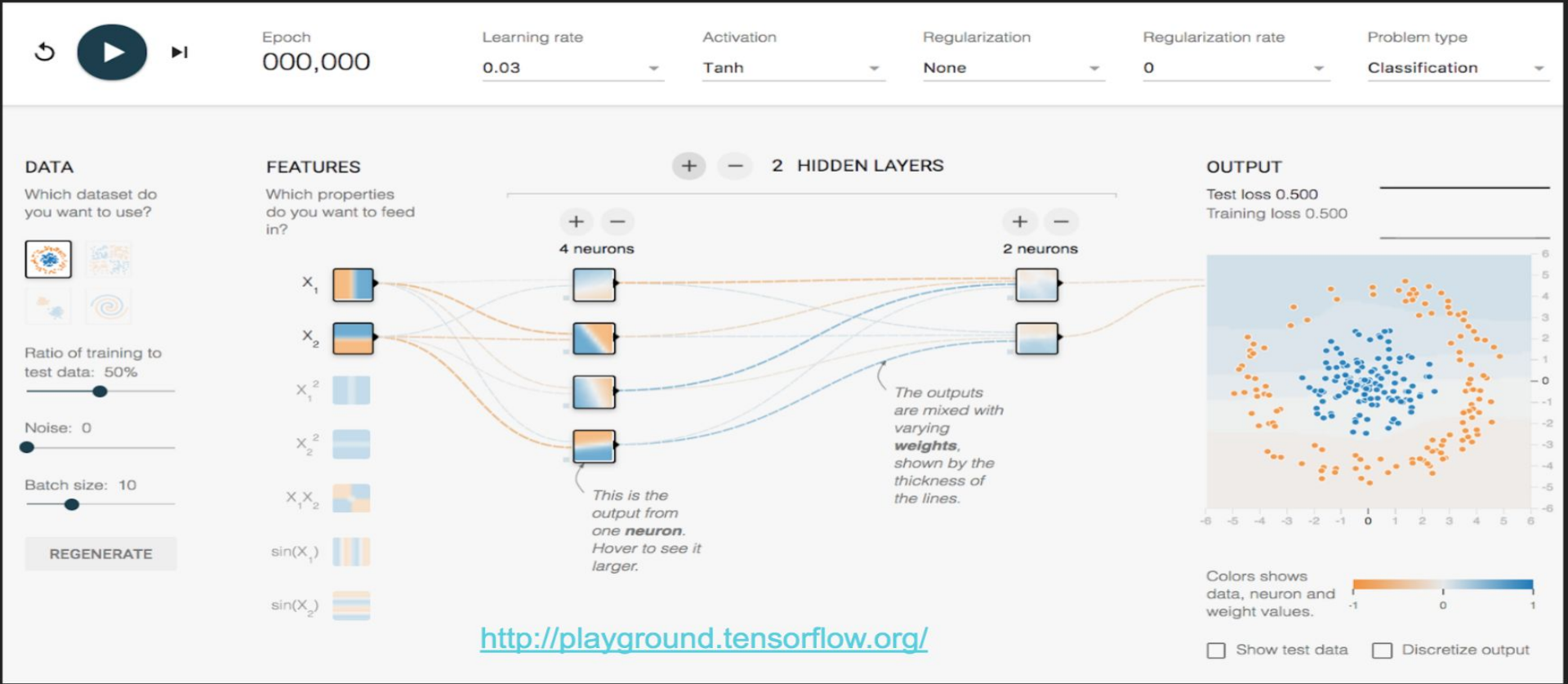
(a) Standard Neural Net



(b) After applying dropout.

http://playground.tensorflow.org/

TensorFlow Playground



Cancer Detection

```
## Read in the data set
data = pd.read_csv("data/BreastCancer.csv")
data.head()
```

	Id	Cl.thickness	Cell.size	Cell.shape	Marg.adhesion	Epith.c.size	Bare.nuclei	Bl.cromatin	Normal.nucleoli	Mitoses	Class
0	1000025	5	1	1	1	2	1	3	1	1	benign
1	1002945	5	4	4	5	7	10	3	2	1	benign
2	1015425	3	1	1	1	2	2	3	1	1	benign
3	1016277	6	8	8	1	3	4	3	7	1	benign
4	1017023	4	1	1	3	2	1	3	1	1	benign

```
## Define the neural net and compile it
from keras.models import Sequential
from keras.layers import Dense, Activation

model = Sequential()
model.add(Dense(32, activation='relu', input_dim=9))
model.add(Dense(32, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(optimizer='rmsprop',
              loss='binary_crossentropy',
              metrics=['accuracy'])
```

http://srdas.github.io/Presentations/ClassClust/DeepLearning_Introduction_Short.slides.html#/21

Digit Recognition

https://www.cs.toronto.edu/~kriz/cifar.html

http://srdas.github.io/Presentations/ClassClust/DeepLearning_Introduction_Short.slides.html#/28

THE MNIST DATABASE

of handwritten digits

Yann LeCun, Courant Institute, NYU
Corinna Cortes, Google Labs, New York
Christopher J.C. Burges, Microsoft Research, Redmond

The MNIST database of handwritten digits, available from this page, has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size-normalized and centered in a fixed-size image.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0							0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	60,000 x 784						0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Showing 1 to 25 of 60,000 entries

Learning the Black-Scholes-Merton Model

```
from scipy.stats import norm
def BSM(S,K,T,sig,rf,dv,cp):  #cp = {+1.0 (calls), -1.0 (puts)}
    d1 = (math.log(S/K)+(rf-dv+0.5*sig**2)*T)/(sig*math.sqrt(T))
    d2 = d1 - sig*math.sqrt(T)
    return cp*S*math.exp(-dv*T)*norm.cdf(d1*cp) - cp*K*math.exp(-rf*T)*norm.cdf(d2
*cp)

df = pd.read_csv('data/training.csv')
```

C is homogeneous degree one, so

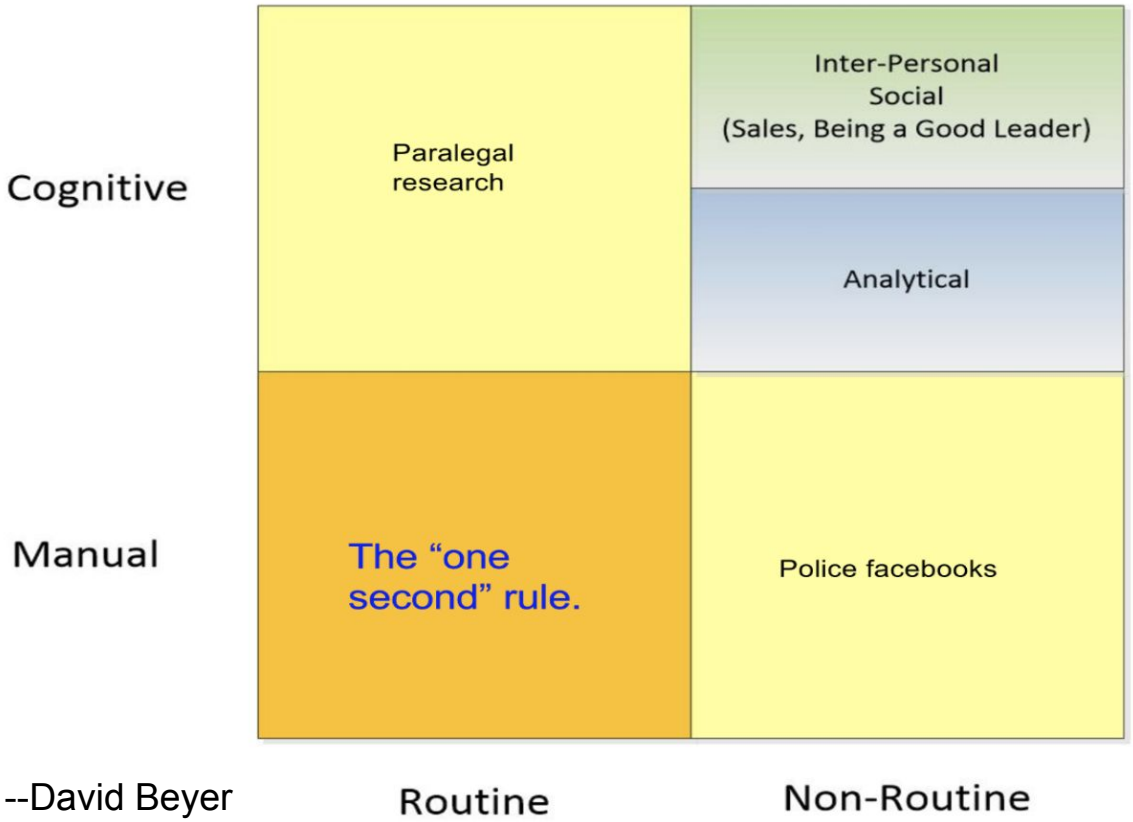
$$aC(S,K) = C(aS,aK)$$

This means we can normalize spot and call prices and remove a variable by dividing by K.

$$\frac{C(S,K)}{K} = C(S/K,1)$$

http://srdas.github.io/Presentations/ClassClust/DeepLearning_Introduction_Short.slides.html#/36

Machine Learning and the Future of Work



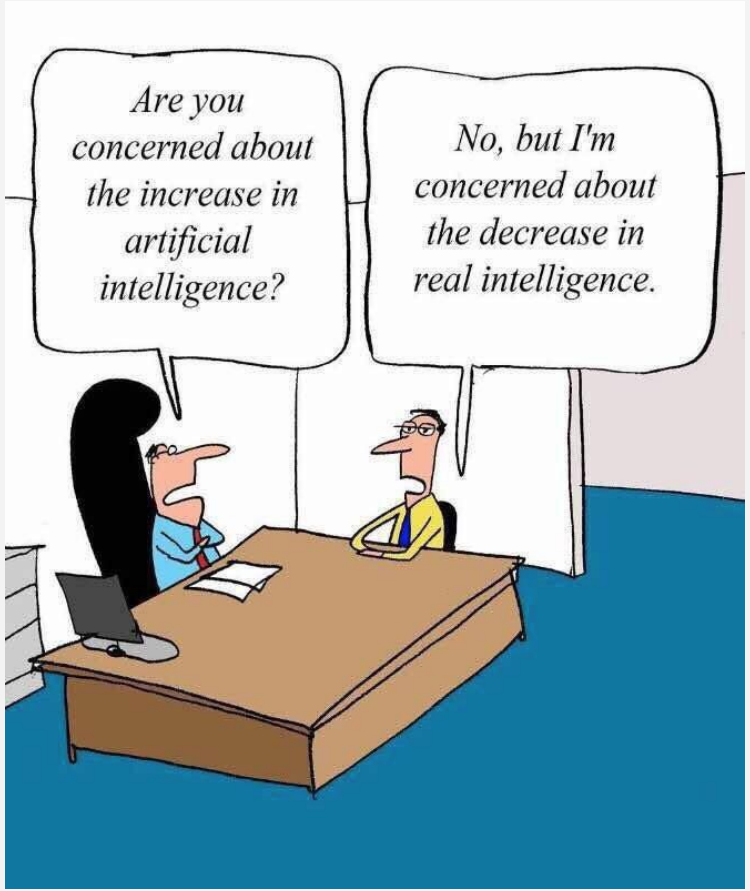
--David Beyer

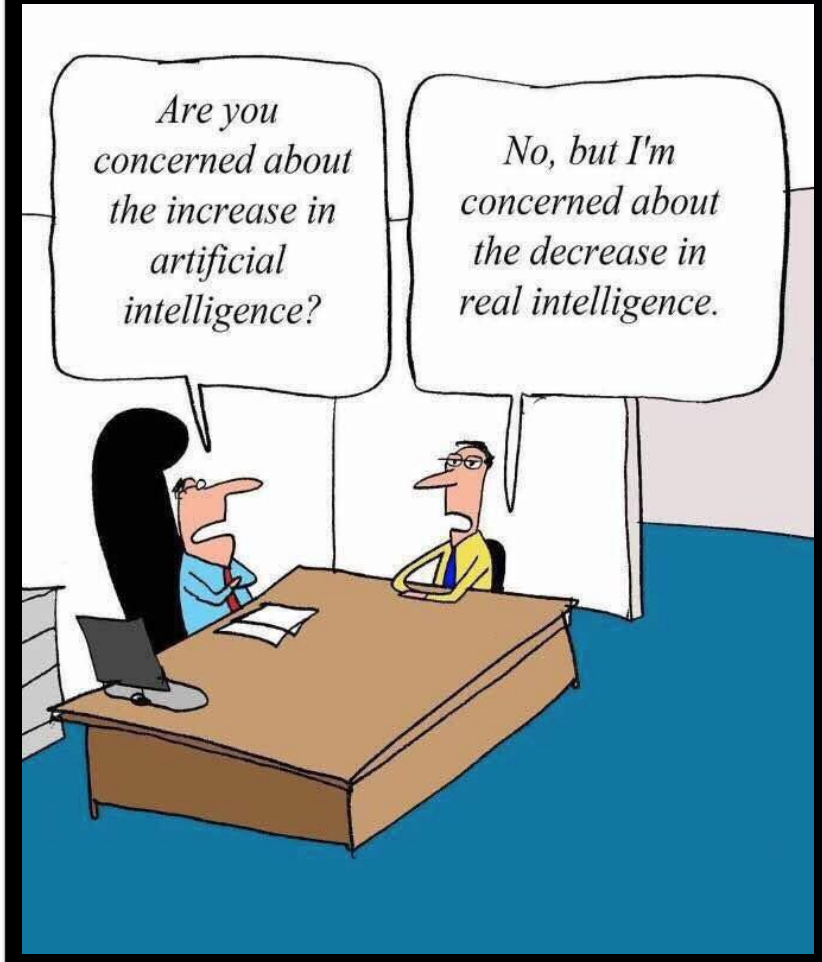
AI researchers salaries go through the roof:
https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html?hp&action=click&pgtype=Homepage&clickSource=story-heading&module=second-column-region&region=top-news&WT.nav=top-news

(Roy) Amara's Law: "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run."

Arthur C. Clarke's Three Laws:
1. When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.
2. The only way of discovering the limits of the possible is to venture a little way past them into the impossible.
3. Any sufficiently advanced technology is indistinguishable from magic.

Thank you.


Are you concerned about the increase in artificial intelligence?
No, but I'm concerned about the decrease in real intelligence.



http://srdas.github.io/Papers/fintech.pdf

Thank you!!

Bank of Indonesia – BIS/IFC Workshop on "Big Data for Central Bank Policies" July 2018

Professor Sanjiv Ranjan Das
Santa Clara University

(The presentation follows the links below. The slides will open in the browser automatically. Press "F" to get full-screen slides. "Esc" returns you from full screen mode.)

Table of Contents

Overview slides for all the content below are here. The slides contain links to all the models and program code below.

14.00 – 15.30
Session 3a: Extracting Knowledge From Large Quantitative Datasets: Classification and Clustering

1. Machine Learning
2. Linear Models
3. Logistic Regression
4. Discriminant Analysis
5. Bayes Classifier
6. Support Vector Machines
7. Nearest Neighbors (kNN)
8. Decision Trees
9. Random Forest

15.45 – 17.15
Session 3b: Extracting Knowledge From Large Quantitative Datasets: Classification and Clustering

1. Dimension Reduction
2. Clustering
3. Neural Networks and Deep Learning

PDFs of both sessions' program code are here:

- Session 1
- Session 2

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Introduction to text mining[1]

Stephen Hansen,

University of Oxford

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Introduction to Text Mining

Bank Indonesia—IFC Workshop

Stephen Hansen
University of Oxford

Introduction

Most empirical work in economics relies on inherently quantitative data: prices, demand, votes, etc.

But a large amount of unstructured text is also generated in economic environments: company reports, policy committee deliberations, media articles, political speeches, etc.

One can use such data qualitatively, but increasing interest in treating text quantitatively.

My lectures will review how economists have done this until recently, and also more modern machine learning approaches.

Textual Databases

A single observation in a textual database is called a *document*.

The set of documents that make up the dataset is called a *corpus*.

We often have covariates associated with each document that are sometimes called *metadata*.

Example

In "Transparency and Deliberation" we use a corpus of verbatim FOMC transcripts from the era of Alan Greenspan:

- 149 meetings from August 1987 through January 2006.
- A document is a single statement by a speaker in a meeting (46,502).
- Associated metadata: speaker biographical information, macroeconomic conditions, etc.

Data Sources

There are many potential sources for text data, such as:

1. PDF files or other non-editable formats
2. Word documents or other editable formats
3. Web pages
4. Application Programming Interfaces (API) for web applications.

From Files to Databases

Turning raw text files into structured databases is often a challenge:

1. Separate metadata from text
2. Identify relevant portions of text (paragraphs, sections, etc)
3. Remove graphs and charts

First step for non-editable files is conversion to editable format, usually with optical character recognition software.

With raw text files, we can use regular expressions to identify relevant patterns.

HTML and XML pages provide structure through tagging.

If all else fails, relatively cheap and reliable services exist for manual extraction.

What is Text?

At an abstract level, text is simply a string of characters.

Some of these may be from the Latin alphabet—'a', 'A', 'p' and so on—but there may also be:

1. Decorated Latin letters (e.g. ö)
2. Non-Latin alphabetic characters (e.g. Chinese and Arabic)
3. Punctuation (e.g. '!')
4. White spaces, tabs, newlines
5. Numbers
6. Non-alphanumeric characters (e.g. '@')

Key Question: How can we obtain an informative, quantitative representation of these character strings? This is the goal of text mining.

First step is to *pre-process* strings to obtain a cleaner representation.

Pre-Processing I: Tokenization

Tokenization is the splitting of a raw character string into individual elements of interest.

Often these elements are words, but we may also want to keep numbers or punctuation as well.

Simple rules work well, but not perfectly. For example, splitting on white space and punctuation will separate hyphenated phrases as in ‘risk-averse agent’ and contractions as in ‘aren’t’.

In practice, you should (probably) use a specialized library for tokenization.

Pre-Processing II: Stopword Removal

The frequency distribution of words in natural languages is highly skewed, with a few dozen words accounting for the bulk of text.

These *stopwords* are typically stripped out of the tokenized representation of text as they take up memory but do not help distinguish one document from another.

Examples from English are ‘a’, ‘the’, ‘to’, ‘for’ and so on.

No definitive list, but example on `http://snowball.tartarus.org/algorithms/english/stop.txt`.

Pre-Processing II: Stopword Removal

The frequency distribution of words in natural languages is highly skewed, with a few dozen words accounting for the bulk of text.

These *stopwords* are typically stripped out of the tokenized representation of text as they take up memory but do not help distinguish one document from another.

Examples from English are ‘a’, ‘the’, ‘to’, ‘for’ and so on.

No definitive list, but example on `http://snowball.tartarus.org/algorithms/english/stop.txt`.

Also common to drop rare words, for example those that appear is less than some fixed percentage of documents.

Pre-Processing III: Linguistic Roots

For many applications, the relevant information in tokens is their linguistic root, not their grammatical form. We may want to treat 'prefer', 'prefers', 'preferences' as equivalent tokens.

Two options:

- *Stemming*: Deterministic algorithm for removing suffixes. Porter stemmer is popular.
  Stem need not be an English word: Porter stemmer maps 'inflation' to 'inflat'.
- *Lemmatizing*: Tag each token with its part of speech, then look up each (word, POS) pair in a dictionary to find linguistic root.
  E.g. 'saw' tagged as verb would be converted to 'see', 'saw' tagged as noun left unchanged.

A related transformation is *case-folding* each alphabetic token into lowercase. Not without ambiguity, e.g. 'US' and 'us' each mapped into same token.

Pre-Processing IV: Multi-Word Phrases

Sometimes groups of individual tokens like "Bank Indonesia" or "text mining" have a specific meaning.

One ad-hoc strategy is to tabulate the frequency of all unique two-token (bigram) or three-token (trigram) phrases in the data, and convert the most common into a single token.

In FOMC data, most common bigrams include 'interest rate', 'labor market', 'basi point'; most common trigrams include 'feder fund rate', 'real interest rate', 'real gdp growth', 'unit labor cost'.

More Systematic Approach

Some phrases have meaning because they stand in for specific names, like "Bank Indonesia". One can used named-entity recognition software applied to raw, tokenized text data to identify these.

Other phrases have meaning because they denote a recurring concept, like "housing bubble". To find these, one can apply part-of-speech tagging, then tabulate the frequency of the following tag patterns:

$$\text{AN/NN/AAN/ANN/NAN/NNN/NPN.}$$

See chapter on collocations in Manning and Schütze's *Foundations of Statistical Natural Language Processing* for more details.

Example from NYT Corpus

$C(w^1\ w^2)$	w^1	w^2	tag pattern
11487	New	York	A N
7261	United	States	A N
5412	Los	Angeles	N N
3301	last	year	A N
3191	Saudi	Arabia	N N
2699	last	week	A N
2514	vice	president	A N
2378	Persian	Gulf	A N
2161	San	Francisco	N N
2106	President	Bush	N N
2001	Middle	East	A N
1942	Saddam	Hussein	N N
1867	Soviet	Union	A N
1850	White	House	A N
1633	United	Nations	A N
1337	York	City	N N
1328	oil	prices	N N

Pre-Processing of FOMC Corpus

	All terms	Alpha terms	No stopwords	Stems
# total terms	6249776	5519606	2505261	2505261
# unique terms	26030	24801	24611	13734

Notation

The corpus is composed of D documents indexed by d.

After pre-processing, each document is a finite, length-N_d list of terms $\mathbf{w}_d = (w_{d,1}, \ldots, w_{d,N_d})$ with generic element $w_{d,n}$.

Let $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_D)$ be a list of all terms in the corpus, and let $N \equiv \sum_d N_d$ be the total number of terms in the corpus.

Suppose there are V **unique** terms in $\mathbf{w}$, where $1 \leq V \leq N$, each indexed by v.

We can then map each term in the corpus into this index, so that $w_{d,n} \in \{1, \ldots, V\}$.

Let $x_{d,v} \equiv \sum_n \mathbb{1}(w_{d,n} = v)$ be the count of term v in document d.

Example

Consider three documents:

1. 'stephen is nice'
2. 'john is also nice'
3. 'george is mean'

We can consider the set of unique terms as $\{\text{stephen}, \text{is}, \text{nice}, \text{john}, \text{also}, \text{george}, \text{mean}\}$ so that $V = 7$.

Construct the following index:

stephen	is	nice	john	also	george	mean
1	2	3	4	5	6	7

We then have $\mathbf{w}_1 = (1, 2, 3)$; $\mathbf{w}_2 = (4, 2, 5, 3)$; $\mathbf{w}_3 = (6, 2, 7)$.

Moreover $x_{1,1} = 1$, $x_{2,1} = 0$, $x_{3,1} = 0$, etc.

Document-Term Matrix

A popular quantitative representation of text is the *document-term matrix* $\mathbf{X}$, which collects the counts $x_{d,v}$ into a $D \times V$ matrix.

In the previous example, we have

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The key characteristics of the document-term matrix are its:

1. High dimensionality
2. Sparsity

Ngram Models

In the above example, we made an explicit choice to count individual terms, which destroys all information on word order.

In some contexts, this may be sufficient for our information needs, but in others we might lose valuable information.

We could alternatively have counted all adjacent two-term phrases, called bigrams or, more generally, all adjacent N-term phrases, called Ngrams.

This is perfectly consistent with the model described above, where v now indexes unique bigrams rather than unique unigrams:

stephen.is	is.nice	john.is	is.also	also.nice	george.is	is.mean
1	2	3	4	5	6	7

We then have $\mathbf{w}_1 = (1, 2)$; $\mathbf{w}_2 = (3, 4, 5)$; $\mathbf{w}_3 = (6, 7)$.

Dimensionality Reduction through Keywords

The first approach to handling $\mathbf{X}$ is to limit attention to a subset of columns of interest.

In the natural language context, this is equivalent to representing text using the distribution of keywords across documents.

One can either look at the incidence of keywords (Boolean search), or else their frequency (dictionary methods).

The researcher must decide in advance which are the keywords of interest.

Application

The recent work of Baker, Bloom, and Davis on measuring economic policy uncertainty (`http://www.policyuncertainty.com/`) is largely based on a media index constructed via Boolean searches of US and European newspapers.

For each paper on each day since 1985, identify articles that contain:

1. "uncertain" OR "uncertainty", AND
2. "economic" OR "economy", AND
3. "congress" OR "deficit" OR "federal reserve" OR "legislation" OR "regulation" OR "white house"

Normalize resulting article counts by total newspaper articles that month.

Results

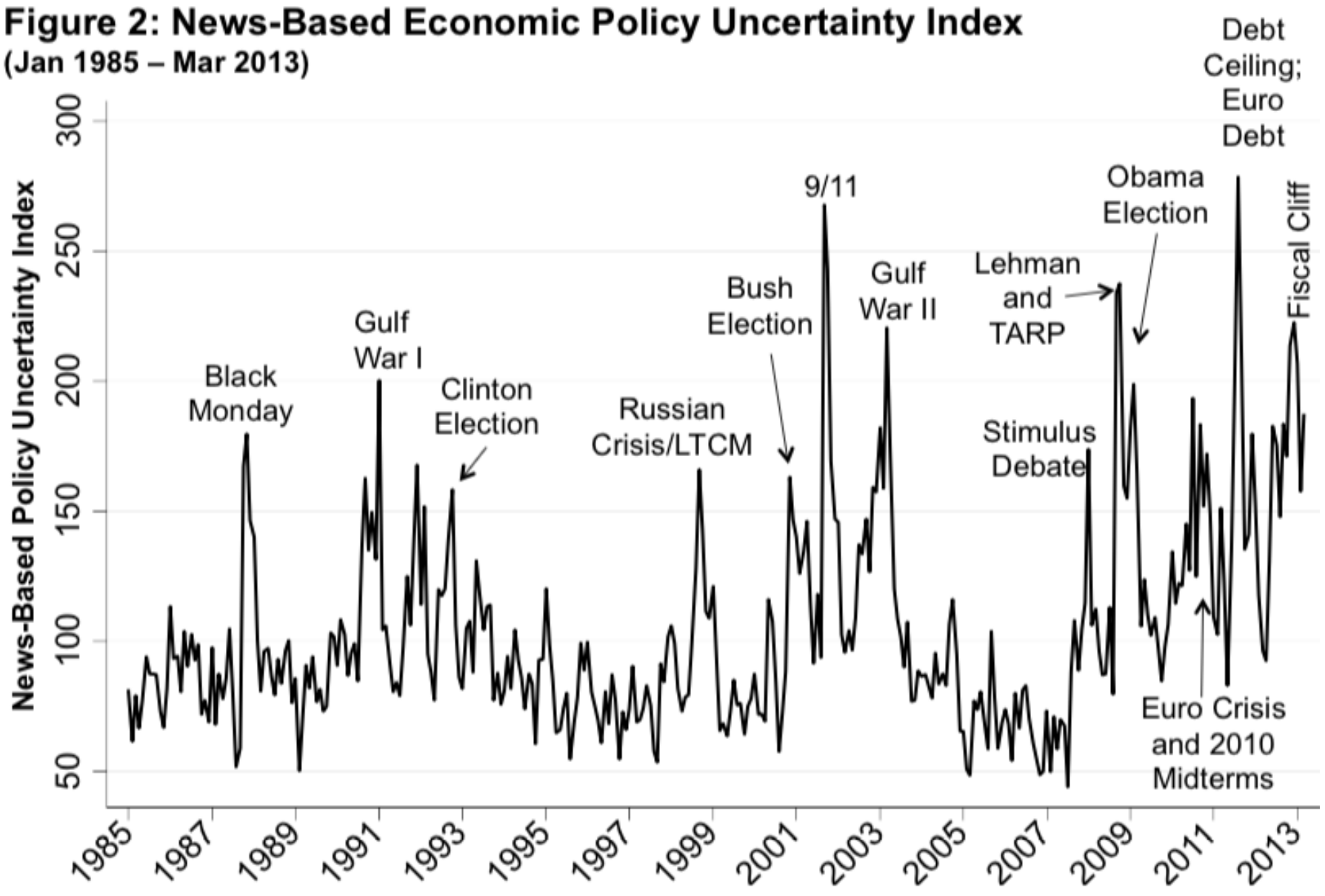


Figure 2: News-Based Economic Policy Uncertainty Index
(Jan 1985 – Mar 2013)
News-Based Policy Uncertainty Index
300
250
200
150
100
50
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
Black Monday
Gulf War I
Clinton Election
Russian Crisis/LTCM
Bush Election
9/11
Gulf War II
Lehman and TARP
Stimulus Debate
Obama Election
Euro Crisis and 2010 Midterms
Debt Ceiling; Euro Debt
Fiscal Cliff

Why Text?

VIX is an asset-based measure of uncertainty: implied S&P 500 volatility at 30-day horizon using option prices.

So what does text add to this?

1. Focus on broader type of uncertainty besides equity prices.
2. Much richer historical time series.
3. Cross-country measures.

Term Weighting

Dictionary methods are based on raw counts of words.

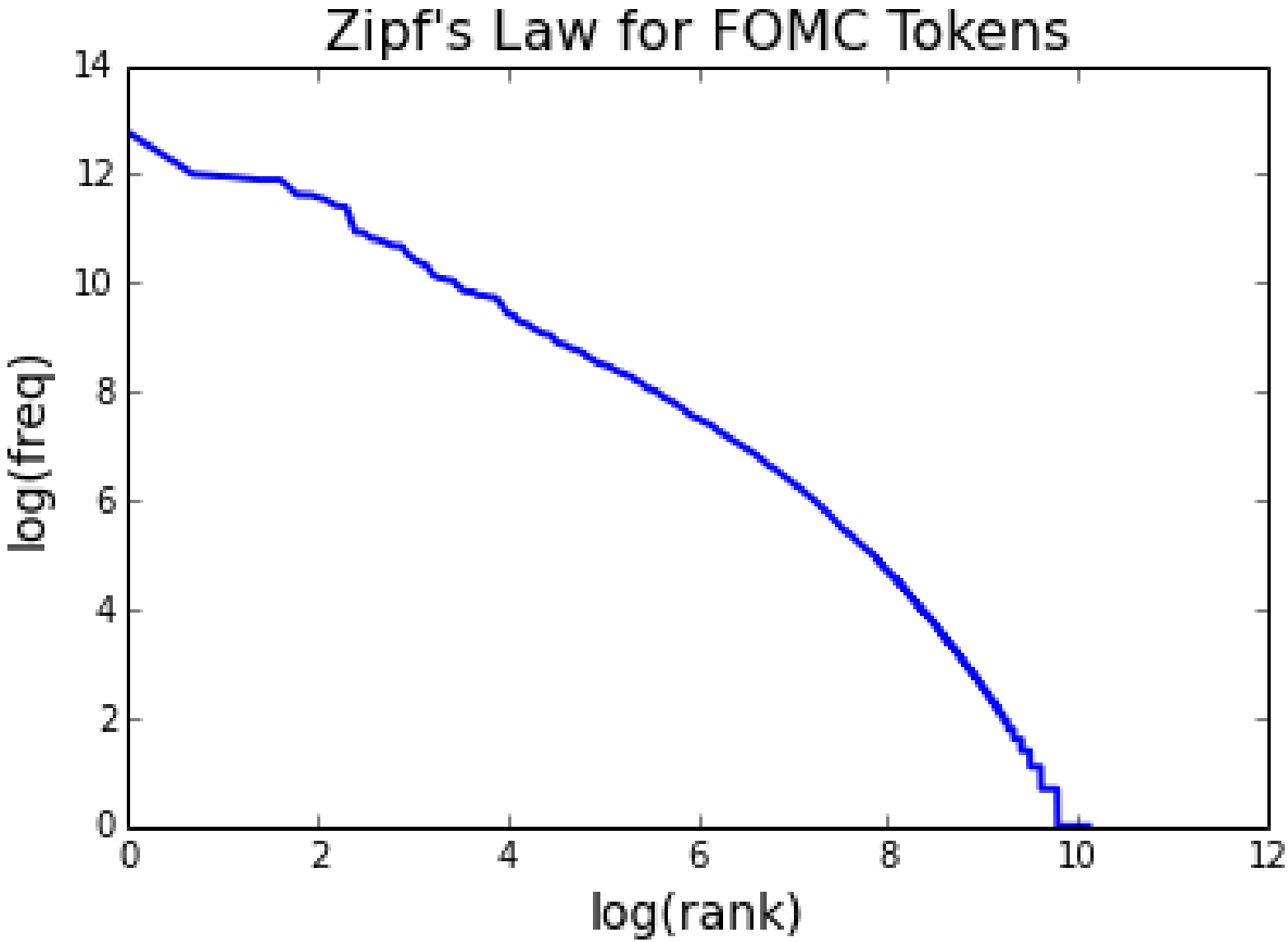
But the frequency of words in natural language can distort raw counts.

Zipf's Law is an empirical regularity for most natural languages that maintains that the frequency of a particular term is inversely proportional to its rank.

Means that a few terms will have very large counts, many terms have small counts.

Example of a *power law*.

Zipf's Law in FOMC Transcript Data



Rescaling Counts

Let $x_{d,v}$ be the count of the vth term in document d.

To dampen the power-law effect can express counts as

$$tf_{d,v} = \begin{cases} 0 & \text{if } x_{d,v} = 0 \\ 1 + \log(x_{d,v}) & \text{otherwise} \end{cases}$$

which is the *term frequency* of v in d.

Thought Experiment

Consider a two-term dictionary $\mathfrak{D} = \{v', v''\}$.

Suppose two documents d' and d'' are such that:

$$x_{d',v'} > x_{d'',v'} \text{ and } x_{d',v''} < x_{d'',v''}.$$

Now suppose that no other document uses term v' but every other document uses term v''.

Which document is "more about" the theme the dictionary captures?

Inverse Document Frequency

Let df_v be the number of documents that contain the term v.

The *inverse document frequency* is

$$\mathrm{idf}_v = \log\left(\frac{D}{df_v}\right),$$

where D is the number of documents.

Properties:

1. Higher weight for words in fewer documents.
2. Log dampens effect of weighting.

TF-IDF Weighting

Combining the two observations from above allows us to express the *term frequency - inverse document frequency* of term v in document d as

$$\text{tf-idf}_{d,v} = tf_{d,v} \times idf_v.$$

Gives prominence to words that occur many times in few documents.

Can now score each document as $s_d = \sum_{v \in \mathfrak{D}} \text{tf-idf}_{d,v}$ and then compare.

In practice, this can provide better results than simple counts.

Data-Driven Stopwords

Stopword lists are useful for generic language, but there are also context-specific frequently used words.

For example, in a corpus of court proceedings, words like 'lawyer', 'law', 'justice' will show up a lot.

Can also define term-frequency across entire corpus as

$$tf_v = 1 + \log\left(\sum_d x_{d,v}\right).$$

One can then rank each term in the corpus according to $tf_v \times idf_v$, and choose a threshold below which to drop terms.

This provides a means for data-driven stopword selection.

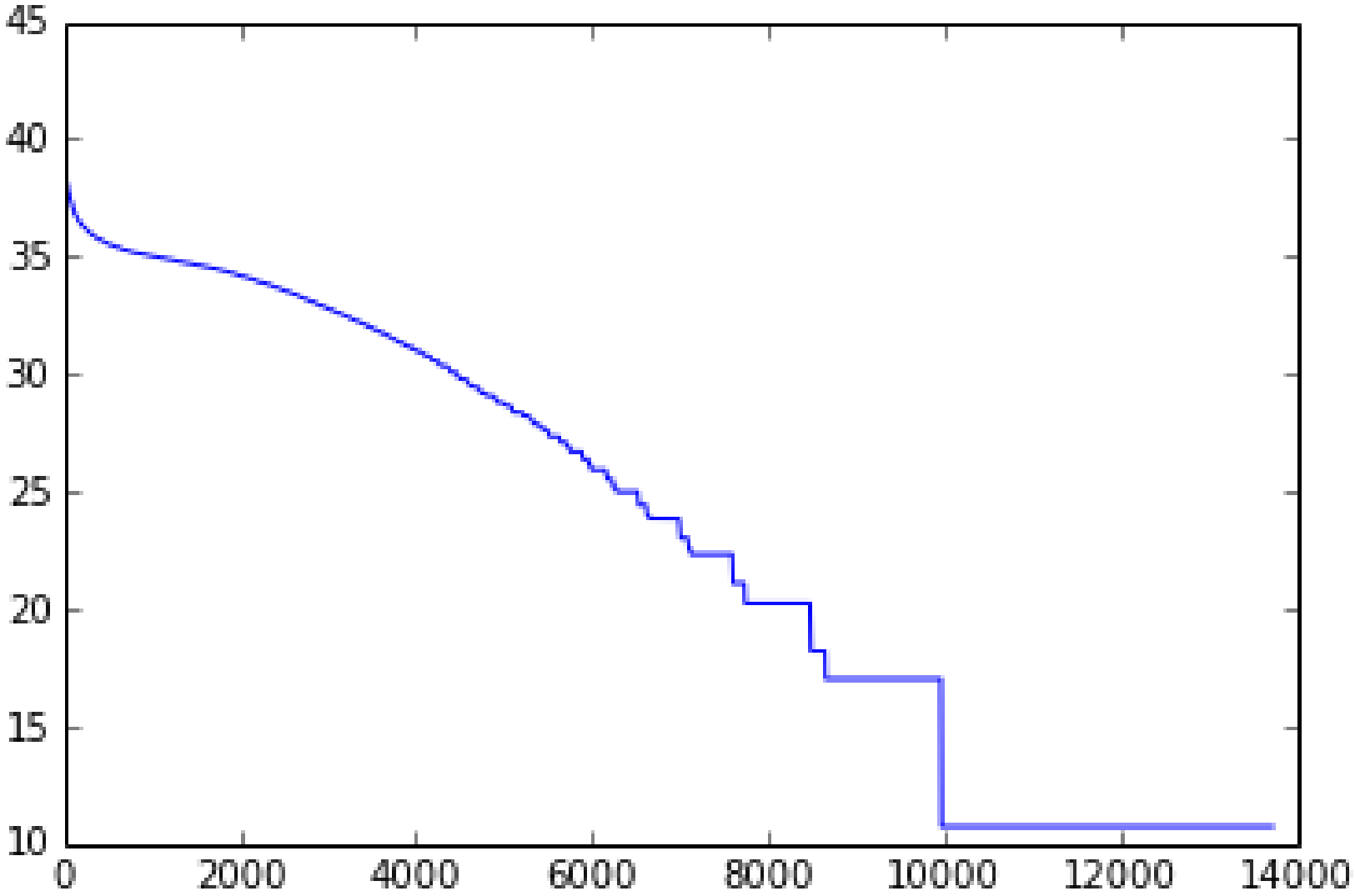
Stem Rankings in FOMC Transcript Data

R1 = collection frequency ranking

R2 = tf-idf-weighted ranking

Rank	1	2	3	4	5	6	7	8	9
R1	rate	think	year	will	market	growth	inflat	price	percent
R2	panel	katrina	graph	fedex	wal	mart	mbs	mfp	euro

Ranking of All FOMC Stems



Vector Space Model

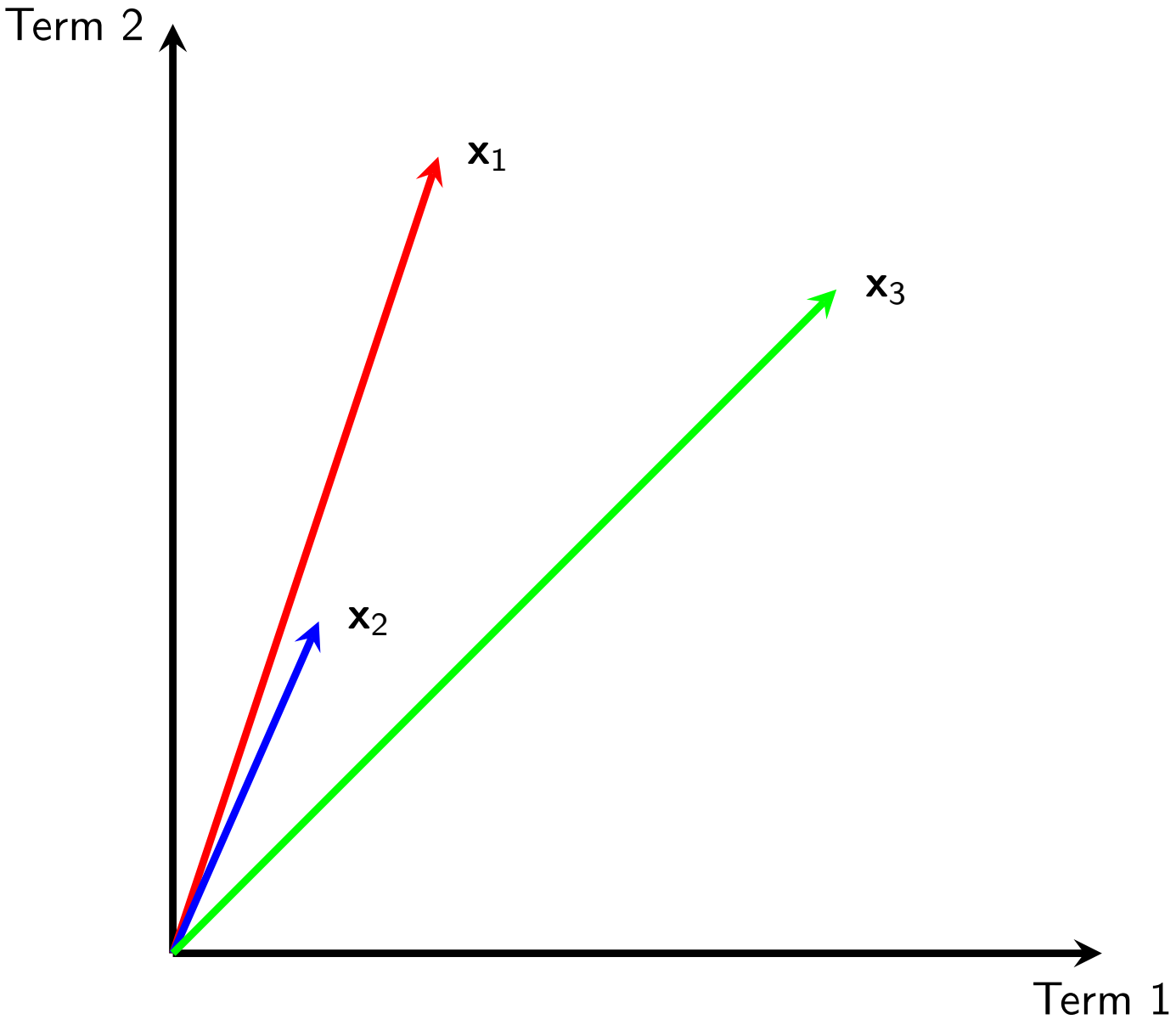
One can also view rows of document-term matrix as vectors lying in a V-dimensional space.

Tf-idf weighting usually used, but not necessary.

The question of interest is how to measure the similarity of two documents in the vector space.

Initial instinct might be to use Euclidean distance $\sqrt{\sum_v (x_{i,v} - x_{j,v})^2}$.

Three Documents



Problem with Euclidean Distance

Semantically speaking, documents 1 and 2 are very close, and document 3 is an outlier.

But the Euclidean distance between 1 and 2 is high due to differences in document length.

What we really care about is whether vectors point in same direction.

Cosine Similarity

Define the cosine similarity between documents i and j as

$$CS(i,j) = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \, \|\mathbf{x}_j\|}$$

1. Since document vectors have no negative elements $CS(i,j) \in [0,1]$.
2. $\mathbf{x}_i / \|\mathbf{x}_i\|$ is unit-length, correction for different distances.

Application

An important theoretical concept in industrial organization is location on a product space.

Industry classification measures are quite crude proxies of this.

Hoberg and Phillips (2010) take product descriptions from 49,408 10-K filings and use the vector space model (with bit vectors defined by dictionaries) to compute similarity between firms.

Data available from `http://alex2.umd.edu/industrydata/`.

Towards Machine Learning

Dictionary methods focus on variation across observations along a limited number of dimensions and ignore the rest.

Ideally we would use variation across *all* dimensions to describe documents.

This obviously provides a richer description of the data, but a deeper point relevant for many high-dimensional datasets is that economic theory does not tell us which dimensions are important.

At the same time, incorporating thousands of independent dimensions of variation in empirical work is difficult.

(Unsupervised) machine learning approaches exploit all dimensions of variation to estimate a lower-dimensional set of types for each documents.

Unsupervised Learning for Text

The implicit assumption of dictionary methods is that the set of words in the dictionary map back into an underlying theme of interest.

For example we might have that
$\mathfrak{D} = \{\text{school}, \text{university}, \text{college}, \text{teacher}, \text{professor}\} \rightarrow \text{education}.$

Latent variable models formalize the idea that documents are formed by hidden variables that generate correlations among observed words.

In a natural language context, these variables can be thought of as topics; other applications will have other interpretations.

Information Retrieval

Latent variable representations can more accurately identify document similarity.

The problem of *synonomy* is that several different words can be associated with the same topic. Cosine similarity between following documents?

school	university	college	teacher	professor
0	5	5	0	2
school	university	college	teacher	professor
10	0	0	4	0

The problem of *polysemy* is that the same word can have multiple meanings. Cosine similarity between following documents?

tank	seal	frog	animal	navy	war
10	10	3	2	0	0
tank	seal	frog	animal	navy	war
10	10	0	0	4	3

If we correctly map words into topics, comparisons become more accurate.

Mixture versus Mixed-Membership Models

An important distinction in modeling documents is whether they are associated with one or more latent variables.

In traditional cluster analysis, and in mixture models, observations are represented as coming from a single latent category.

In fact, we might imagine that documents cover more than one topic: monetary policy speeches discuss inflation and growth.

Models that associated observations with more than one latent variable are called *mixed-membership* models. Also relevant outside of text mining: in models of group formation, agents can be associated with different latent communities (sports team, workplace, church, etc).

Latent Semantic Analysis

One of the first mixed-membership models in text mining was the Latent Semantic Analysis/Indexing model of Deerwester et. al. (1990).

A linear algebra rather than probabilistic approach that applies a singular value decomposition to document-term matrix.

Closely related to classical principal components analysis.

Examples in economics: Boukus and Rosenberg (2006); Hendry and Madeley (2010); Acosta (2014); Waldinger et. al. (2018).

Singular Value Decomposition

The document-term matrix $\mathbf{X}$ is not square, but we can decompose it using a generalization of the eigenvector decomposition called the *singular value decomposition.*

Proposition

The document-term matrix can be written $\mathbf{X} = \mathbf{A}\mathbf{\Sigma}\mathbf{B}^T$ *where* $\mathbf{A}$ *is a* $D \times D$ *orthogonal matrix,* $\mathbf{B}$ *is a* $V \times V$ *orthogonal matrix, and* $\mathbf{\Sigma}$ *is a* $D \times V$ *matrix where* $\mathbf{\Sigma}_{ji} = \sigma_i$ *with* $\sigma_i \geq \sigma_{i+1}$ *and* $\mathbf{\Sigma}_{ij} = 0$ *for all* $i \neq j$.

Approximating the Document-Term Matrix

We can obtain a rank k approximation of the document-term matrix $\mathbf{X}_k$ by constructing $\mathbf{X}_k = \mathbf{A}\mathbf{\Sigma}_k\mathbf{B}^T$, where $\mathbf{\Sigma}_k$ is the diagonal matrix formed by replacing $\mathbf{\Sigma}_{ii} = 0$ for $i > k$.

The idea is to keep the "content" dimensions that explain common variation across terms and documents and drop "noise" dimensions that represent idiosyncratic variation.

Often k is selected to explain a fixed portion p of variance in the data. In this case k is the smallest value that satisfies $\sum_{i=1}^{k} \sigma_i^2 / \sum_i \sigma_i^2 \geq p$.

We can then perform the same operations on $\mathbf{X}_k$ as on $\mathbf{X}$, e.g. cosine similarity.

Example

Suppose the document-term matrix is given by

$$\mathbf{X} = \begin{array}{c} \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{array} \begin{array}{c} \begin{array}{cccc} \text{car} & \text{automobile} & \text{ship} & \text{boat} \end{array} \\ \begin{bmatrix} 10 & 0 & 1 & 0 \\ 5 & 5 & 1 & 1 \\ 0 & 14 & 0 & 0 \\ 0 & 2 & 10 & 5 \\ 1 & 0 & 20 & 21 \\ 0 & 0 & 2 & 7 \end{bmatrix} \end{array}$$

Matrix of Cosine Similarities

	d_1	d_2	d_3	d_4	d_5	d_6
d_1	1	·	·	·	·	·
d_2	0.70	1	·	·	·	·
d_3	0.00	0.69	1	·	·	·
d_4	0.08	0.30	0.17	1	·	·
d_5	0.10	0.21	0.00	0.92	1	·
d_6	0.02	0.17	0.00	0.66	0.88	1

SVD

The singular values are $(31.61, 15.14, 10.90, 5.03)$.

$$\mathbf{A} = \begin{bmatrix} 0.0381 & 0.1435 & -0.8931 & -0.02301 & 0.3765 & 0.1947 \\ 0.0586 & 0.3888 & -0.3392 & 0.0856 & -0.7868 & -0.3222 \\ 0.0168 & 0.9000 & 0.2848 & 0.0808 & 0.3173 & 0.0359 \\ 0.3367 & 0.1047 & 0.0631 & -0.7069 & -0.2542 & 0.5542 \\ 0.9169 & -0.0792 & 0.0215 & 0.1021 & 0.1688 & -0.3368 \\ 0.2014 & -0.0298 & 0.0404 & 0.6894 & -0.2126 & 0.6605 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0.0503 & 0.2178 & -0.9728 & 0.0595 \\ 0.0380 & 0.9739 & 0.2218 & 0.0291 \\ 0.7024 & -0.0043 & -0.0081 & -0.7116 \\ 0.7088 & -0.0634 & 0.0653 & 0.6994 \end{bmatrix}$$

Rank-2 Approximation

$$\mathbf{X}_2 = \begin{array}{c} \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{array} \begin{array}{c} \begin{array}{cccc} \text{car} & \text{automobile} & \text{ship} & \text{boat} \end{array} \\ \begin{bmatrix} 0.5343 & 2.1632 & 0.8378 & 0.7169 \\ 1.3765 & 5.8077 & 1.2765 & 0.9399 \\ 2.9969 & 13.2992 & 0.3153 & 0.4877 \\ 0.8817 & 1.9509 & 7.4715 & 7.4456 \\ 1.1978 & 0.0670 & 20.3682 & 20.6246 \\ 0.2219 & 0.1988 & 4.4748 & 4.5423 \end{bmatrix} \end{array}$$

Matrix of Cosine Similarities

	d_1	d_2	d_3	d_4	d_5	d_6
d_1	1	·	·	·	·	·
d_2	0.97	1	·	·	·	·
d_3	0.91	0.97	1	·	·	·
d_4	0.60	0.43	0.23	1	·	·
d_5	0.45	0.26	0.05	0.98	1	·
d_6	0.47	0.29	0.07	0.98	0.99	1

Application: Transparency

How transparent should a public organization be?

Benefit of transparency: accountability.

Costs of transparency:

1. Direct costs
2. Privacy
3. Security
4. Worse behavior $\rightarrow$ "chilling effect"

Transparency and Monetary Policy

Mario Draghi (2013): "It would be wise to have a richer communication about the rationale behind the decisions that the governing council takes."

Table: Disclosure Policies as of 2014

	Fed	BoE	ECB
Minutes?	✓	✓	X
Transcripts?	✓	X	X

Natural Experiment

FOMC meetings were recorded and transcribed from at least the mid-1970's in order to assist with the preparation of the minutes.

Committee members unaware that transcripts were stored prior to October 1993.

Greenspan then acknowledged the transcripts' existence to the Senate Banking Committee, and the Fed agreed:

1. To begin publishing them with a five-year lag.
2. To publish the back data.

"All the News That's Fit to Print"

The New York Times

VOL. CLXIII . . . No. 56,420 SATURDAY, FEBRUARY 22, 2014

Fed Misread Fiscal Crisis, Records Show

After Caution in 2008, Series of Bold Steps

By BINYAMIN APPELBAUM

WASHINGTON — On the morning after Lehman Brothers filed for bankruptcy in 2008, most Federal Reserve officials still believed that the American economy would keep growing despite the metastasizing financial crisis.

The Fed's policy-making committee voted unanimously against bolstering the economy by cutting interest rates, and several officials praised what they described as the decision to let Lehman fail, saying it would help to restore a sense of accountability on Wall Street.

James Bullard, president of the Federal Reserve Bank of St. Louis, urged his colleagues "to wait for some time to assess the impact of the Lehman bankruptcy filing, if any, on the national economy," according to transcripts of the Fed's 2008 meetings that it published on Friday.

DETROIT OUTLINES MAP TO SOLVENCY, STRESSING REPAIR

WAY OUT OF BANKRUPTCY

Balancing Act Worries Banks and Angers Retirees in City

By MONICA DAVEY and MARY WILLIAMS WALSH

DETROIT — Seven months after this city entered bankruptcy, its leaders on Friday presented a federal judge with the first official road map to Detroit's future — documents designed to show how it aims to settle its $18 billion debt to creditors and make itself livable again.

But the proposal is less a vision for a brand-new city than a repair estimate for the old one. It is a document designed by lawyers and bankruptcy experts to find ways to pay off more than 100,000 creditors and then budget money over a period of years to create a

Deal Signed in Ukraine, but Shows Strai

Greenspan's View on Transparency

"**A considerable amount of free discussion and probing questioning by the participants of each other and of key FOMC staff members takes place.** In the wide-ranging debate, new ideas are often tested, many of which are rejected ... **The prevailing views of many participants change as evidence and insights emerge.** This process has proven to be a very effective procedure for gaining a consensus ... It could not function effectively if participants had to be concerned that their half-thought-through, but nonetheless potentially valuable, notions would soon be made public. **I fear in such a situation the public record would be a sterile set of bland pronouncements scarcely capturing the necessary debates which are required of monetary policymaking.**"

Measuring Disagreement

Acosta (2014) uses LSA to measure disagreement before and after transparency.

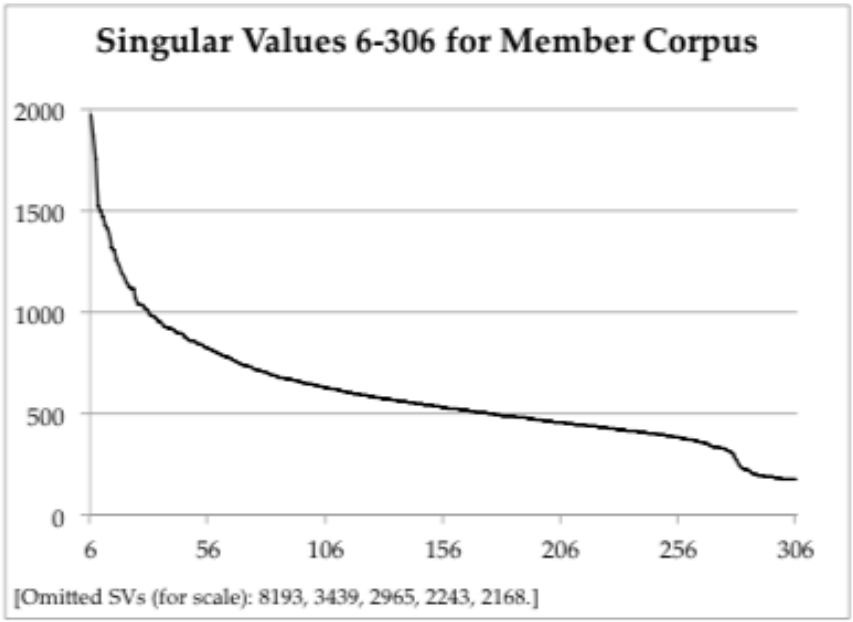
For each member i in each meeting t, let $\vec{d}_{it}$ be member i's words.

Let $\vec{d}_{-i,t} = \sum_i \vec{d}_{it} - \vec{d}_{it}$ be all other members' words.

Quantity of interest is the similarity between $\vec{d}_{it}$ and $\vec{d}_{-i,t}$.

Total set of documents—$\vec{d}_{it}$ and $\vec{d}_{-i,t}$ for all meetings and speakers—is 6,152.

Singular Values



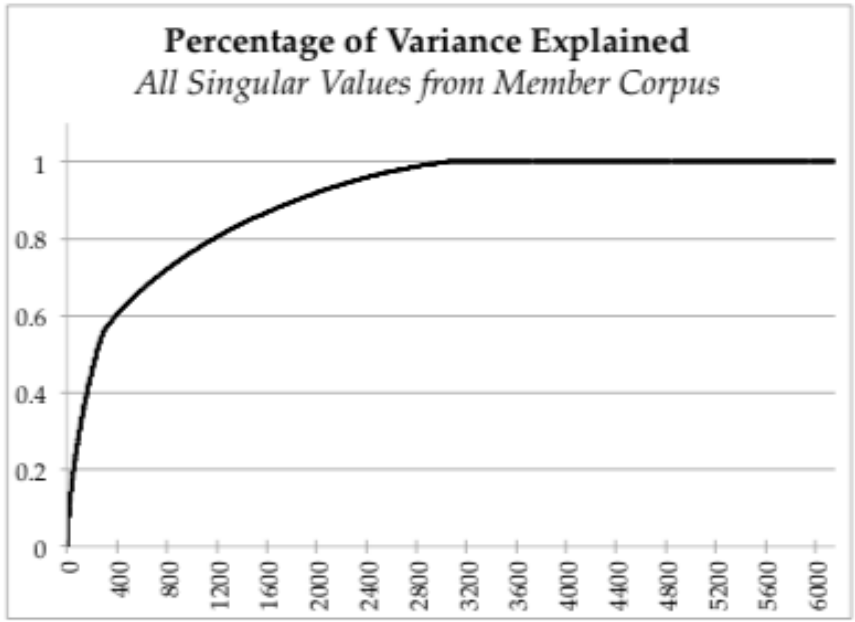
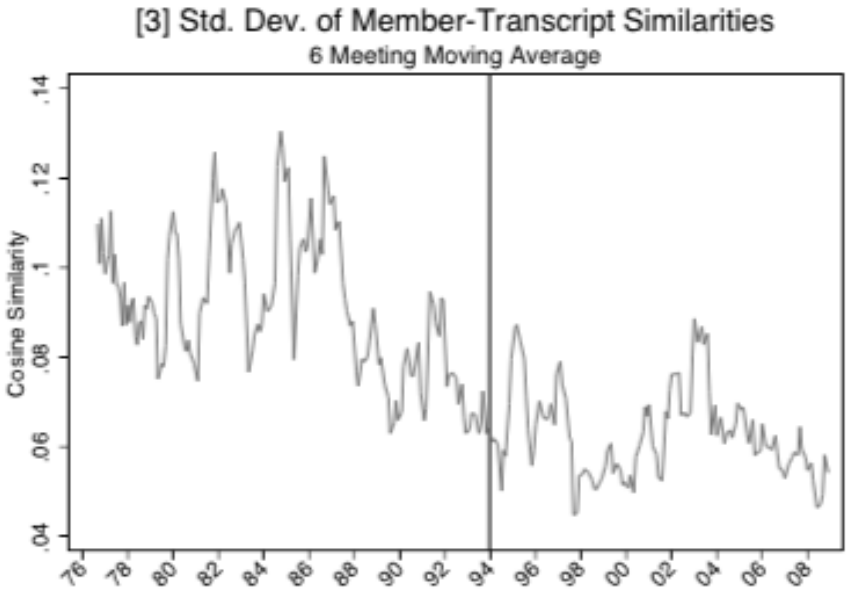
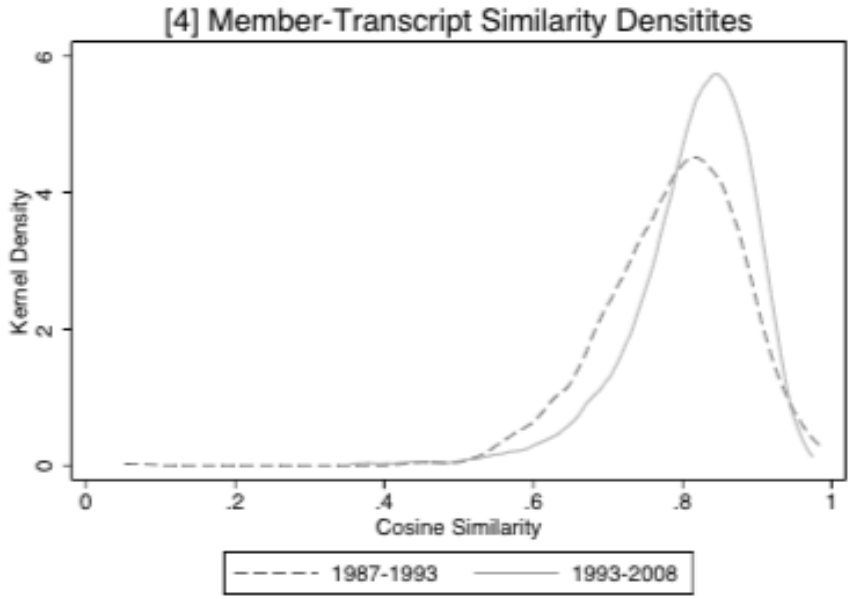


Figure 11: The left hand side shows the 6^{th} through 306^{th} singular values (the elements $\sigma_i \in \mathbf{\Sigma}$ from the SVD) from the member corpus. The right hand side graph show percentage of the variance explained by all 6152 singular values for the member corpus.

Results


[3] Std. Dev. of Member-Transcript Similarities
6 Meeting Moving Average
Cosine Similarity
.04
.06
.08
.1
.12
.14
76
78
80
82
84
86
88
90
92
94
96
98
00
02
04
06
08


[4] Member-Transcript Similarity Densitites
Kernel Density
0
2
4
6
0
.2
.4
.6
.8
1
Cosine Similarity
1987-1993
1993-2008

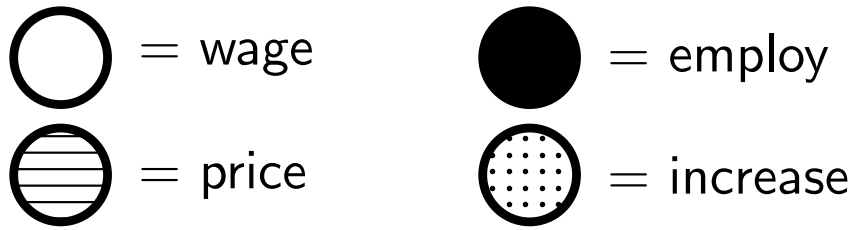
Probabilistic Models

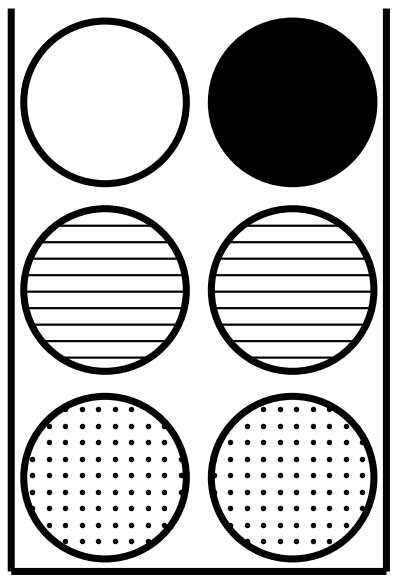
LSA is an important development in machine learning approaches to text, but has some important limitations:

1. SVD is a linear algebra approach to dimensionality reduction, no underlying probability model.
2. The statistical foundations that do exist for SVD are not appropriate for text.
3. Difficult to interpret the components.
4. Difficult to extend LSA to incorporate additional dependencies of interest.

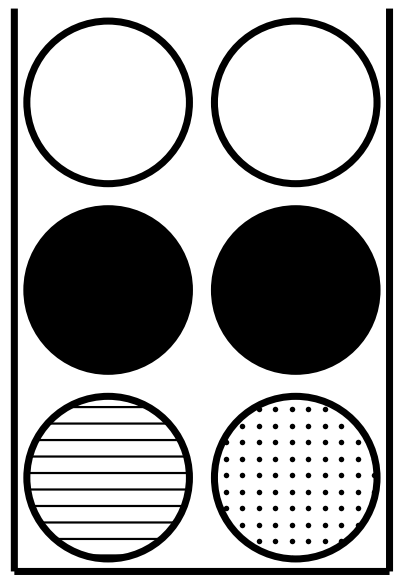
Explicit probability models for text address all of these.

Topics as Urns



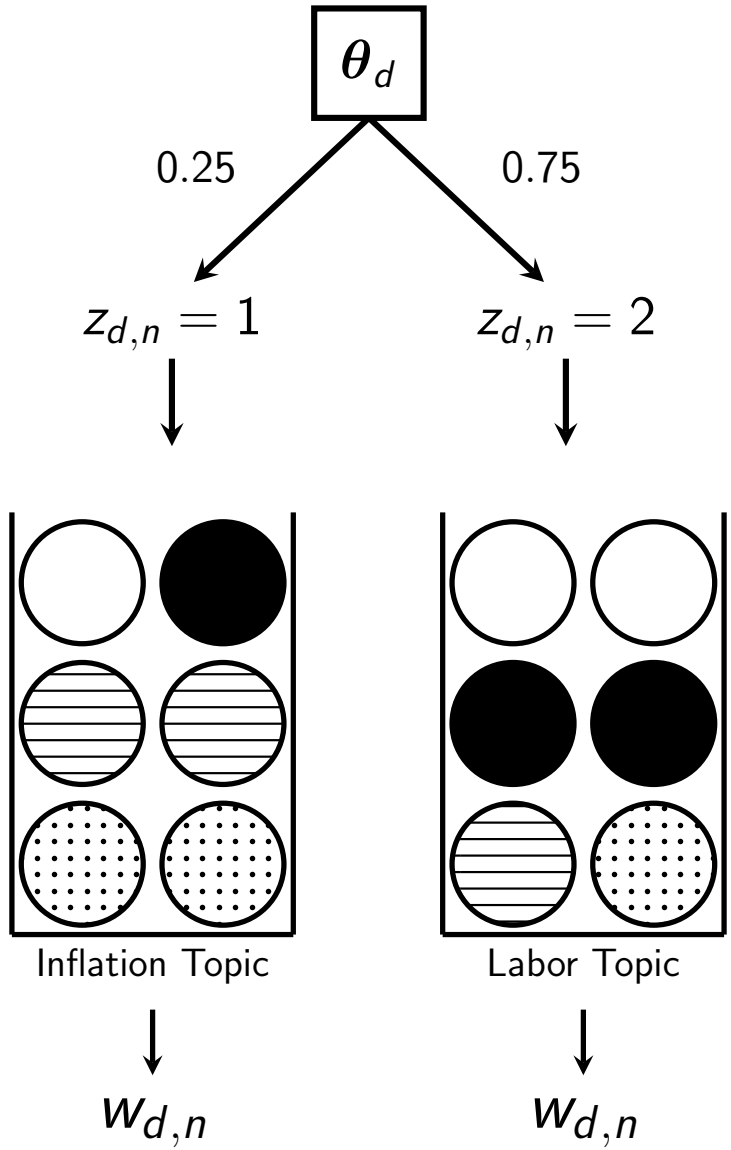


"Inflation" Topic



"Labor" Topic

Mixed-Membership Model for Document



Latent Dirichlet Allocation

The preceding model of documents has a huge number of parameters, so maximum likelihood estimation risks overfitting.

One solution is to adopt a Bayesian approach; the preceding model with a prior distribution on the document-specific mixing probabilities is called Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003).

LDA is widely used within computer science and, increasingly, social sciences.

LDA forms the basis of many, more complicated mixed-membership models.

Latent Dirichlet Allocation—Formal Model

1. Draw $\boldsymbol{\beta}_k$ independently for $k = 1, \ldots, K$ from Dirichlet(η).
2. Draw $\boldsymbol{\theta}_d$ independently for $d = 1, \ldots, D$ from Dirichlet(α).
3. Each word $w_{d,n}$ in document d is generated from a two-step process:
   - 3.1 Draw topic assignment $z_{d,n}$ from θ_d.
   - 3.2 Draw $w_{d,n}$ from $\boldsymbol{\beta}_{z_{d,n}}$.

Fix scalar values for η and α.

Example statement: Yellen, March 2006, #51

Raw Data $\rightarrow$ Remove Stop Words $\rightarrow$ Stemming $\rightarrow$ Multi-word tokens = Bag of Words

We have noticed a change in the relationship between the core CPI and the chained core CPI, which suggested to us that maybe something is going on relating to substitution bias at the upper level of the index. You focused on the nonmarket component of the PCE, and I wondered if something unusual might be happening with the core CPI relative to other measures.

Example statement: Yellen, March 2006, #51

Raw Data → Remove Stop Words → Stemming → Multi-word tokens = Bag of Words

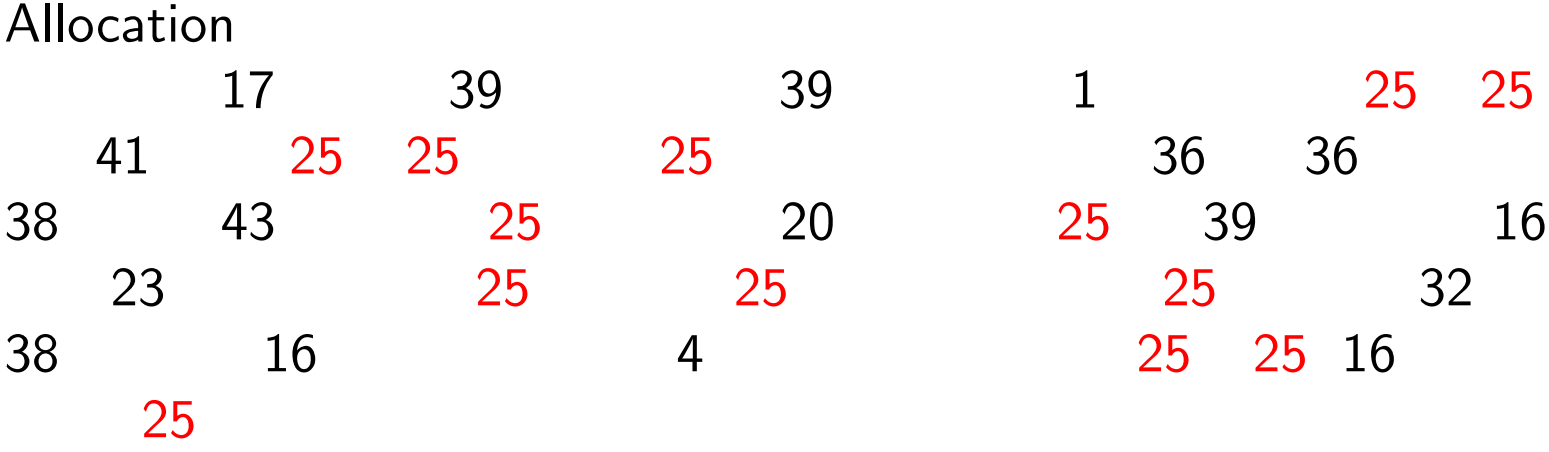
noticed change relationship between core CPI chained core CPI suggested maybe something going relating substitution bias upper level index focused nonmarket component PCE wondered something unusual happening core CPI relative measures

Example statement: Yellen, March 2006, #51

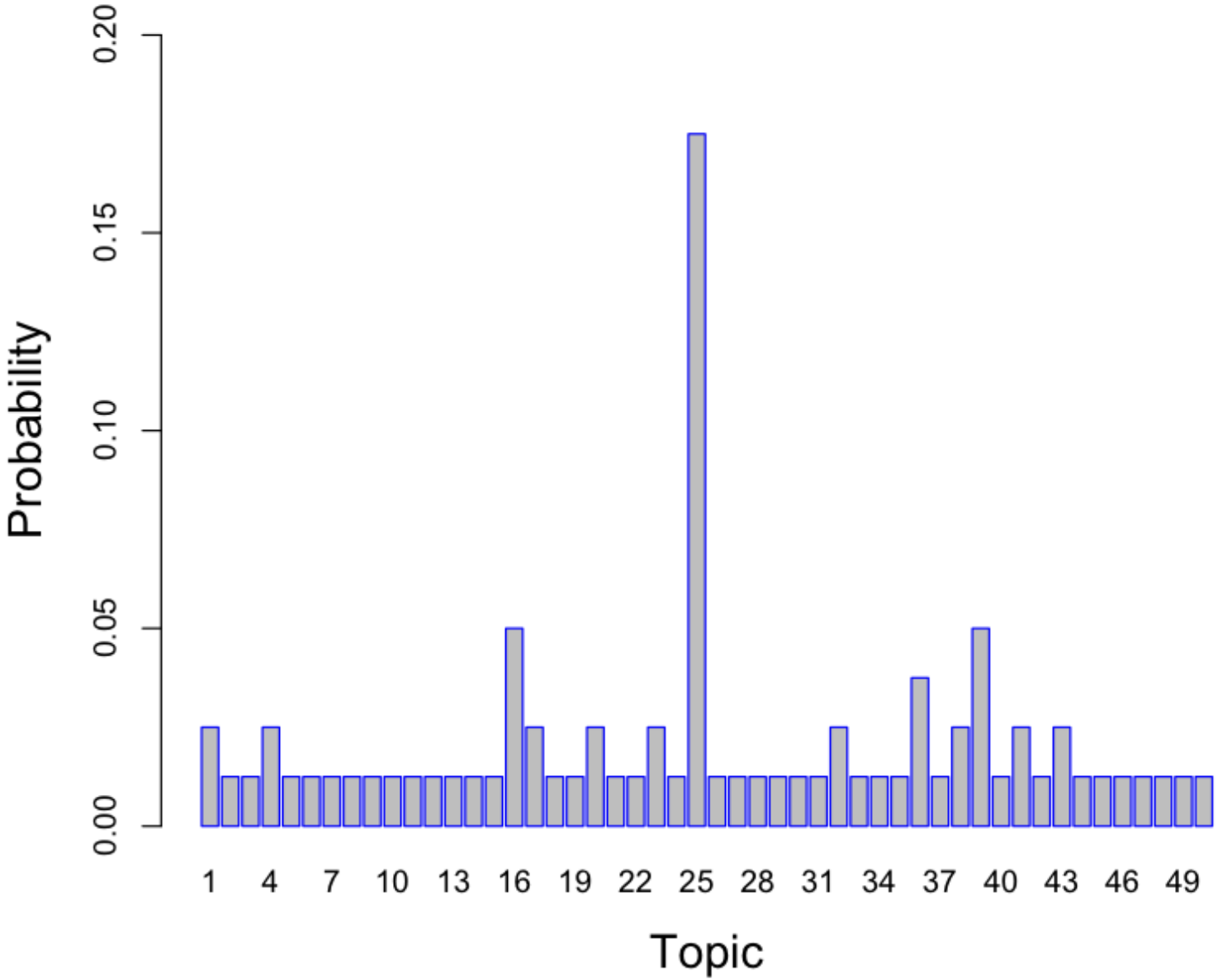
Raw Data → Remove Stop Words → Stemming → Multi-word tokens = Bag of Words

notic chang relationship between core CPI chain core CPI suggest mayb someth go relat substitut bia upper level index focus nonmarket compon PCE wonder someth unusu happen core CPI rel measur

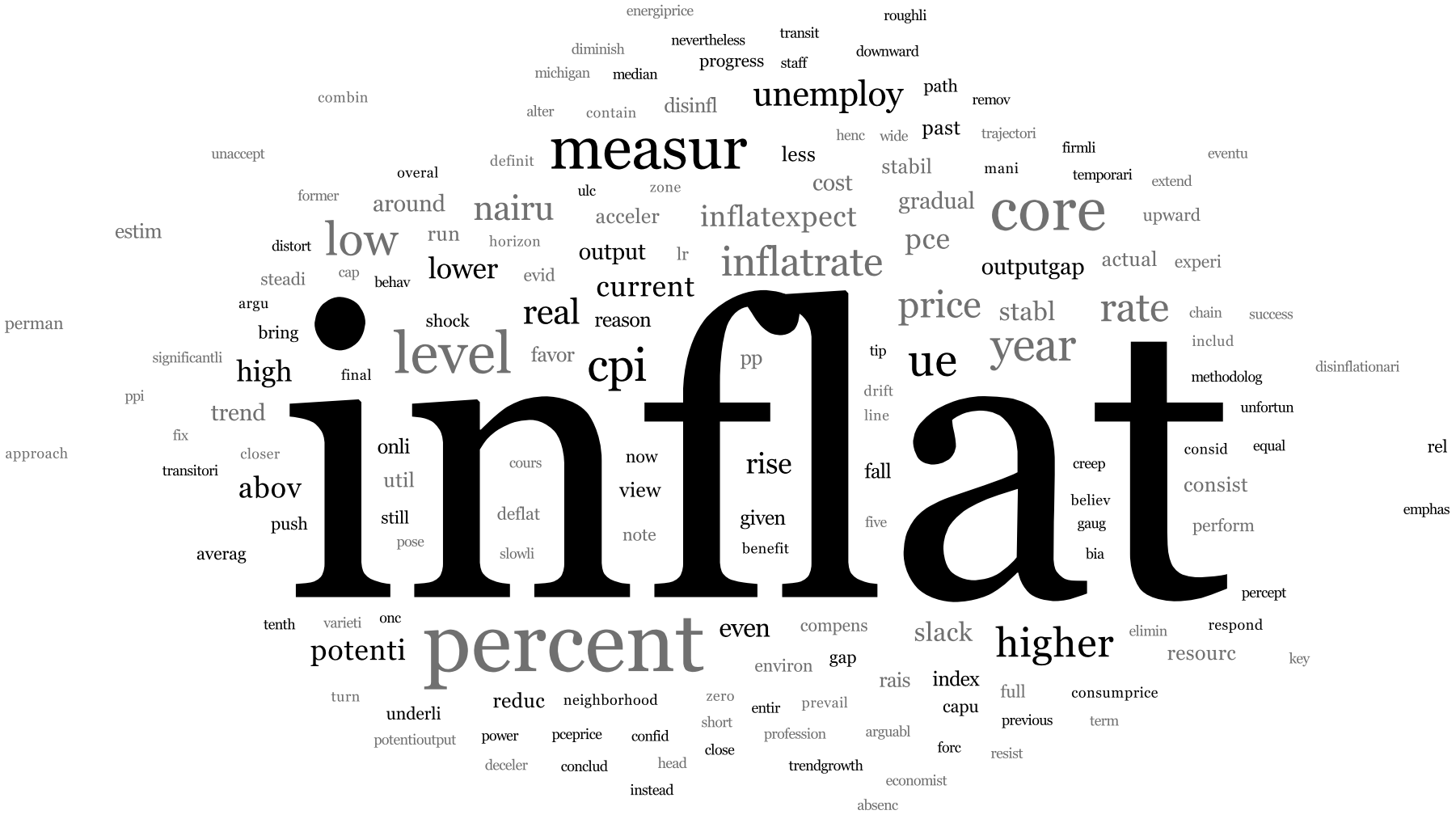
Example statement: Yellen, March 2006, #51



Distribution of Attention



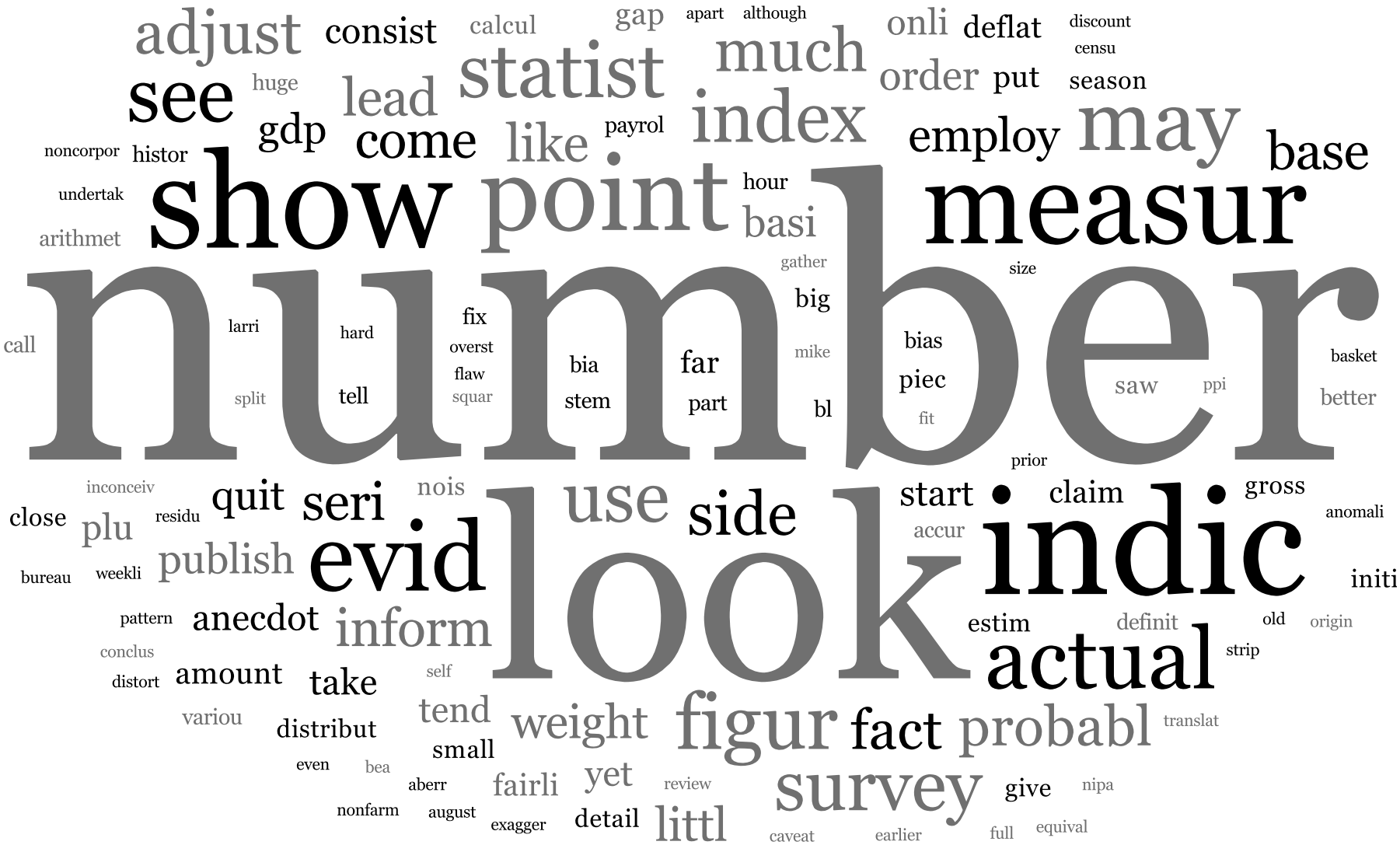
Topic 25


energiprice
roughli
nevertheless
transit
diminish
progress
staff
downward
michigan
median
combin
unemploy
path
alter
contain
disinfl
remov
measur
past
henc
wide
trajectori
unaccept
definit
less
firmli
eventu
overal
stabil
mani
temporari
former
ulc
zone
cost
extend
around
nairu
acceler
inflatexpect
gradual
core
upward
estim
low
run
horizon
pce
distort
output
lr
inflatrate
outputgap
actual
experi
steadi
cap
behav
lower
evid
current
argu
price
stabl
rate
chain
success
perman
bring
shock
real
reason
level
favor
cpi
pp
tip
ue
year
includ
significantli
high
final
methodolog
disinflationari
ppi
drift
line
trend
unfortun
fix
onli
consid
equal
rel
approach
closer
cours
now
rise
fall
creep
transitori
util
view
consist
abov
believ
deflat
given
five
gaug
emphas
push
still
note
perform
pose
benefit
averag
slowli
bia
percept
tenth
varieti
onc
even
compens
slack
higher
elimin
respond
potenti
percent
resourc
key
environ
gap
rais
index
full
consumprice
turn
reduc
neighborhood
zero
entir
prevail
capu
underli
short
profession
arguabl
previous
term
potentioutput
power
pceprice
confid
close
forc
resist
deceler
conclud
head
trendgrowth
economist
instead
absenc

Advantage of Flexibility

'measur' has probability 0.026 in topic 25, and probability 0.021 in topic 11

Topic 11


adjust
consist
calcul
gap
apart
although
onli
deflat
discount
censu
see
huge
lead
statist
much
order
put
season
gdp
come
like
payrol
index
employ
may
noncorpor
histor
base
undertak
show
point
hour
basi
measur
arithmet
number
gather
size
big
call
larri
hard
fix
overst
flaw
bia
far
mike
bias
basket
split
tell
squar
stem
part
bl
piec
saw
ppi
better
fit
prior
inconceiv
quit
seri
nois
use
side
start
claim
gross
close
plu
residu
look
accur
anomali
bureau
weekli
publish
evid
indic
initi
pattern
anecdot
inform
estim
definit
old
origin
conclus
actual
strip
distort
amount
take
self
variou
distribut
tend
weight
figur
fact
probabl
translat
small
even
bea
survey
aberr
fairli
yet
review
give
nipa
nonfarm
august
exagger
detail
littl
caveat
earlier
full
equival

Advantage of Flexibility

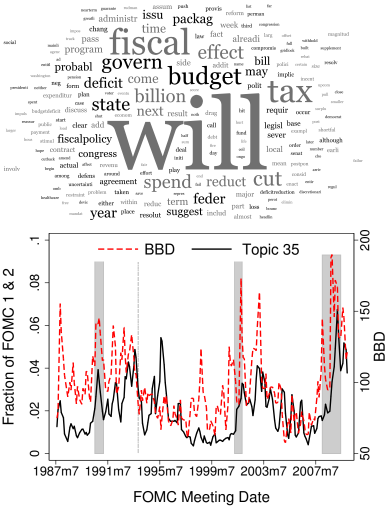
'measur' has probability 0.026 in topic 25, and probability 0.021 in topic 11.

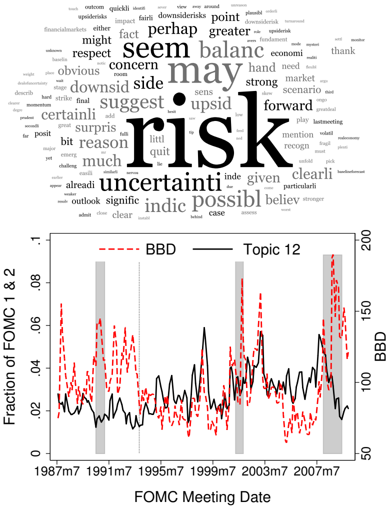
It gets assigned to 25 in this statement consistently due to the presence of other topic 25 words.

In statements containing words on evidence and numbers, it consistently gets assigned to 11.

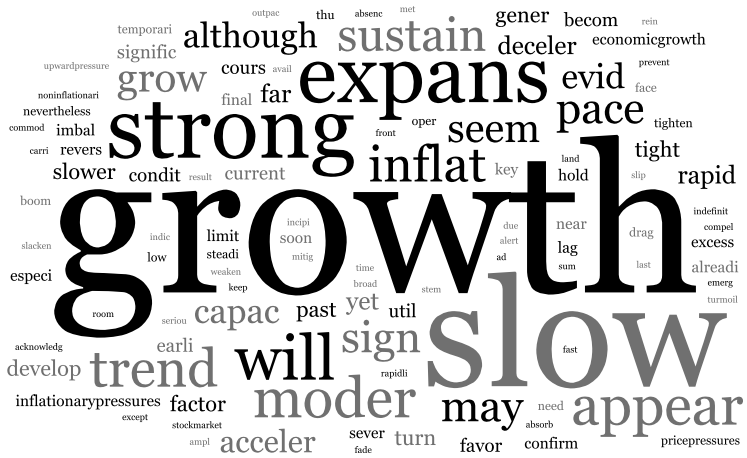
Sampling algorithm can help place words in their appropriate context.

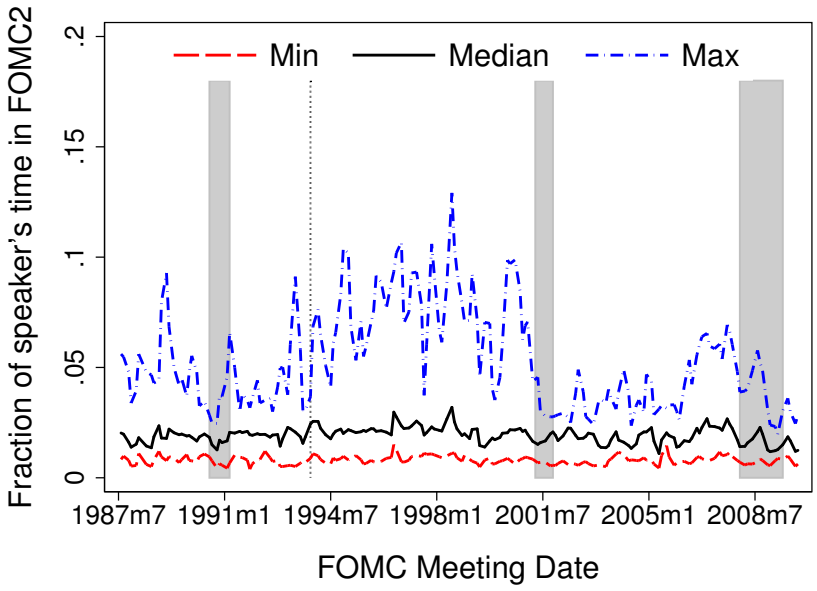
External Validation—BBD

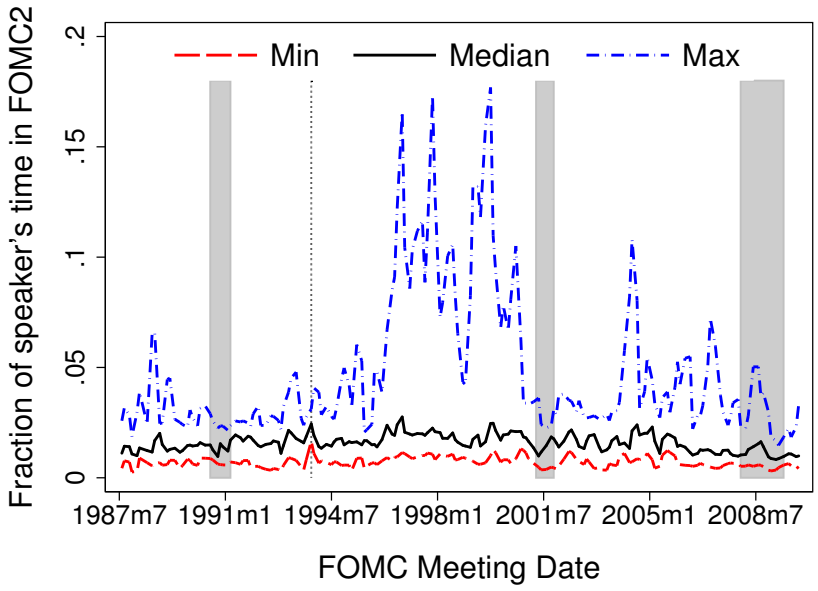




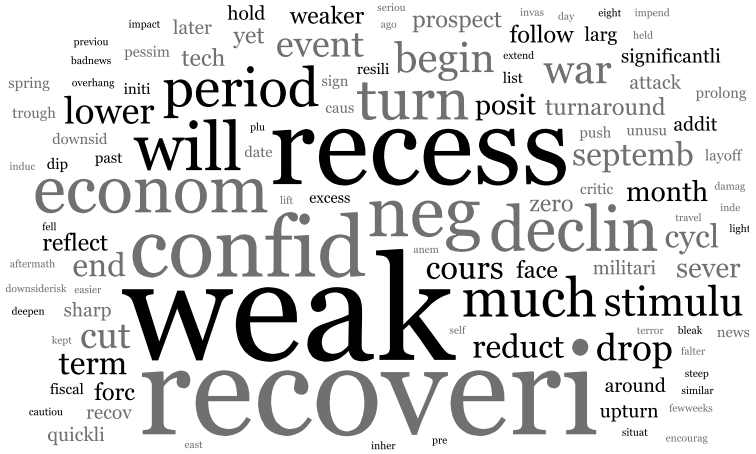
Pro-Cyclical Topics

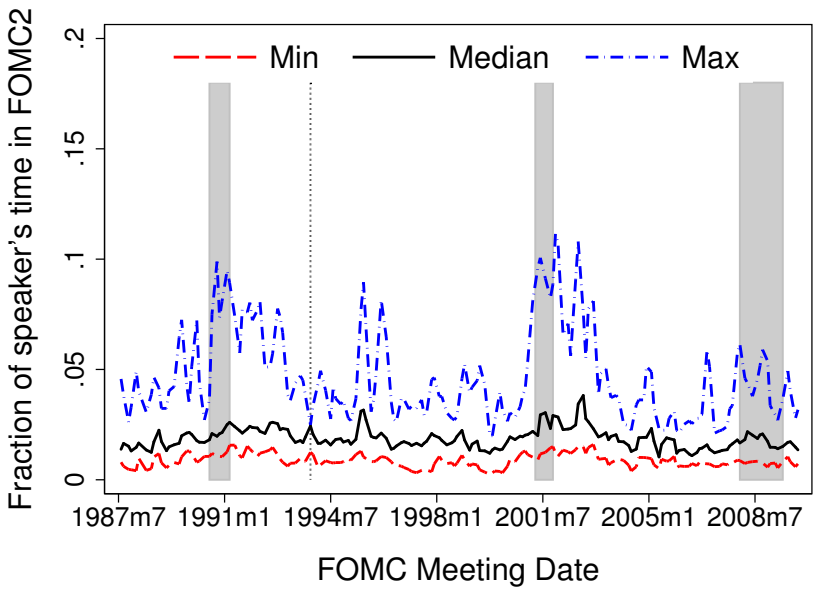


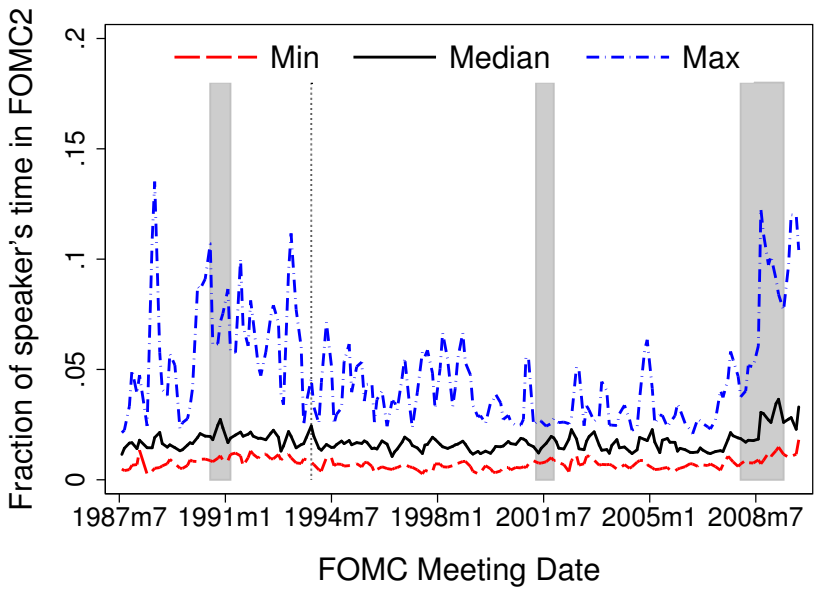




Counter-Cyclical Topics



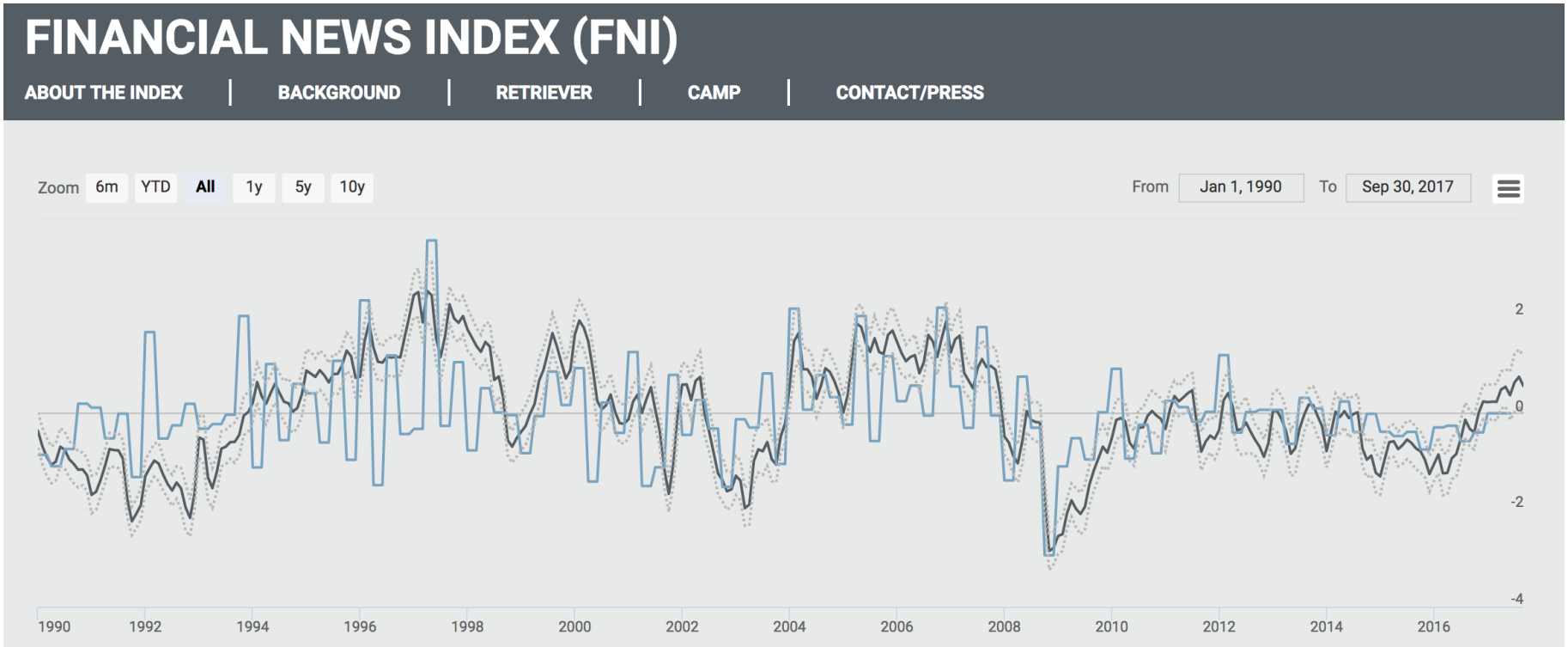




Applications of LDA

1. Forecasting: Mueller and Rauh (2017); Larsen and Thorsrud (2016).
2. Transparency: Hansen et. al. (2017).
3. Information Processing: Nimark and Pitschner (2017).
4. Basis for structural estimation?

Financial News Index



Intuition for Algorithm

Probability term n in document d is assigned to topic k is increasing in:

1. The number of other terms in document d that are currently assigned to k.
2. The number of other occurrences of the term $w_{d,n}$ in the entire corpus that are currently assigned to k.

Both mean that terms that regularly co-occur in documents will be grouped together to form topics.

Property 1 means that terms within a document will tend to be grouped together into few topics rather than spread across many separate topics.

Model Selection

There are three parameters to set to run the Gibbs sampling algorithm: number of topics K and hyperparameters α, η.

Priors don't receive too much attention in literature. Griffiths and Steyvers recommend $\eta = 200/V$ and $\alpha = 50/K$. Smaller values will tend to generate more concentrated distributions. (See also Wallach et. al. 2009).

Methods to choose K:

1. Predict text well $\rightarrow$ out-of-sample goodness-of-fit.
2. Information criteria.
3. Cohesion (focus on interpretability).

Cross Validation

Fit LDA on training data, obtain estimates of $\widehat{\beta}_1, \ldots, \widehat{\beta}_K$.

For test data, estimate θ_d distributions, or else use uniform distribution.

Compute log-likelihood of held-out data as

$$\ell\left(\mathbf{w} \mid \widehat{\Theta}\right) = \sum_{d=1}^{D} \sum_{v=1}^{V} x_{d,v} \log\left(\sum_{k=1}^{K} \widehat{\theta}_{d,k} \widehat{\beta}_{k,v}\right)$$

Higher values indicate better goodness-of-fit.

Information Criteria

Information criteria trade off goodness-of-fit with model complexity.

There are various forms: AIC, BIC, DIC, etc.

Erosheva et. al. (2007) compare several in the context of an LDA-like model, and find that AICM is optimal.

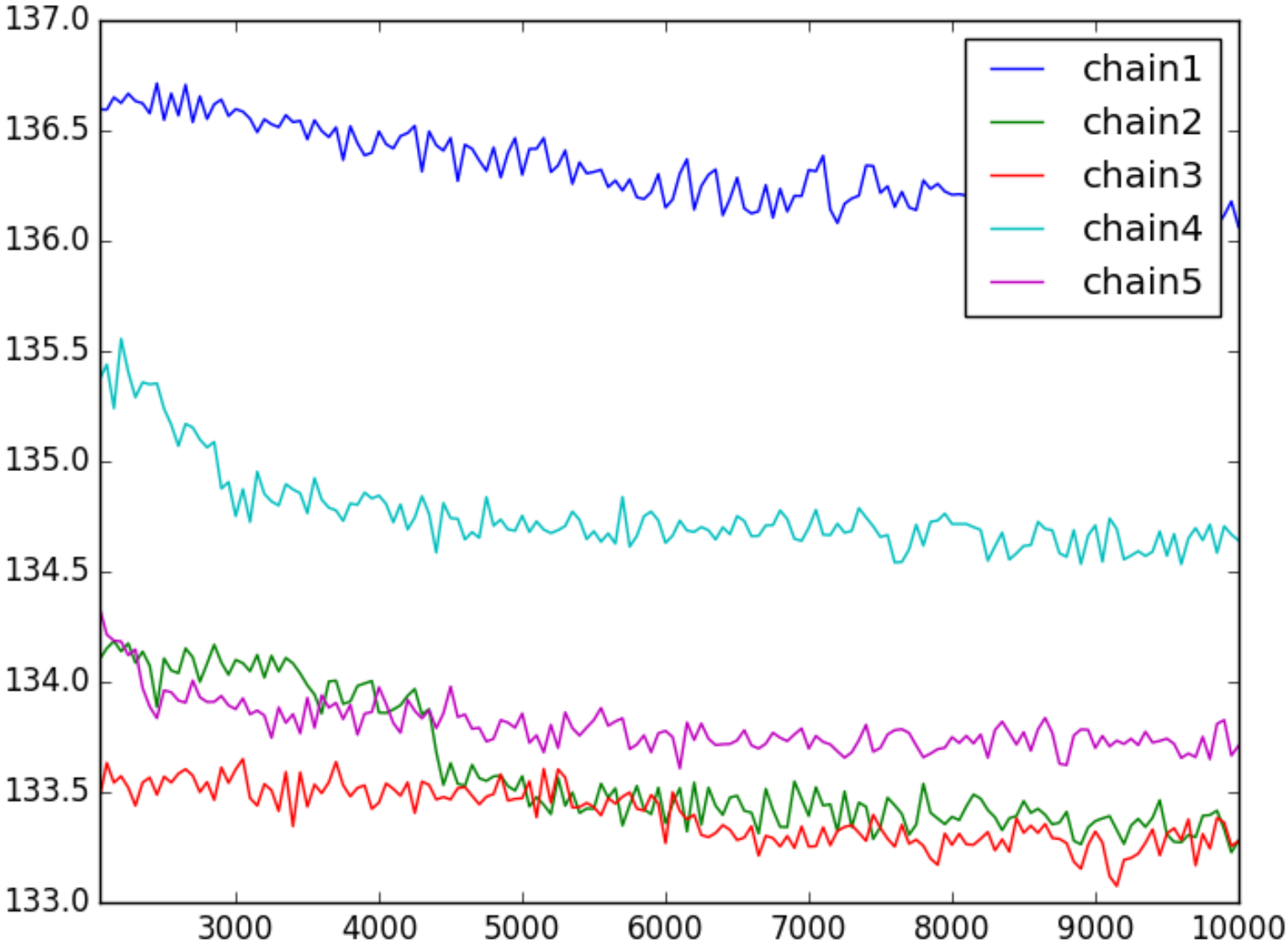
Let $\mu_\ell = \frac{1}{S}\sum_s \ell\left(\mathbf{w} \mid \widehat{\Theta}^s\right)$ be the average value of the log-likelihood across S draws of a Markov chain and

Let $\sigma^2_\ell = \frac{1}{S}\sum_s \left(\ell\left(\mathbf{w} \mid \widehat{\Theta}^s\right) - \mu_\ell\right)^2$ be the variance.

The AICM is $2(\mu_\ell - \sigma^2_\ell)$.

Goodness-of-Fit with $K = 2$

Goodness-of-Fit with $K = 10$



Formalizing Interpretablility

Chang et. al. (2009) propose an objective way of determining whether topics are interpretable.

Two tests:

1. *Word intrusion*. Form set of words consisting of top five words from topic k + word with low probability in topic k. Ask subjects to identify inserted word.
2. *Topic intrusion*. Show subjects a snippet of a document + top three topics associated to it + randomly drawn other topic. Ask to identify inserted topic.

Estimate LDA and other topic models on NYT and Wikipedia articles for $K = 50, 100, 150$.

Results

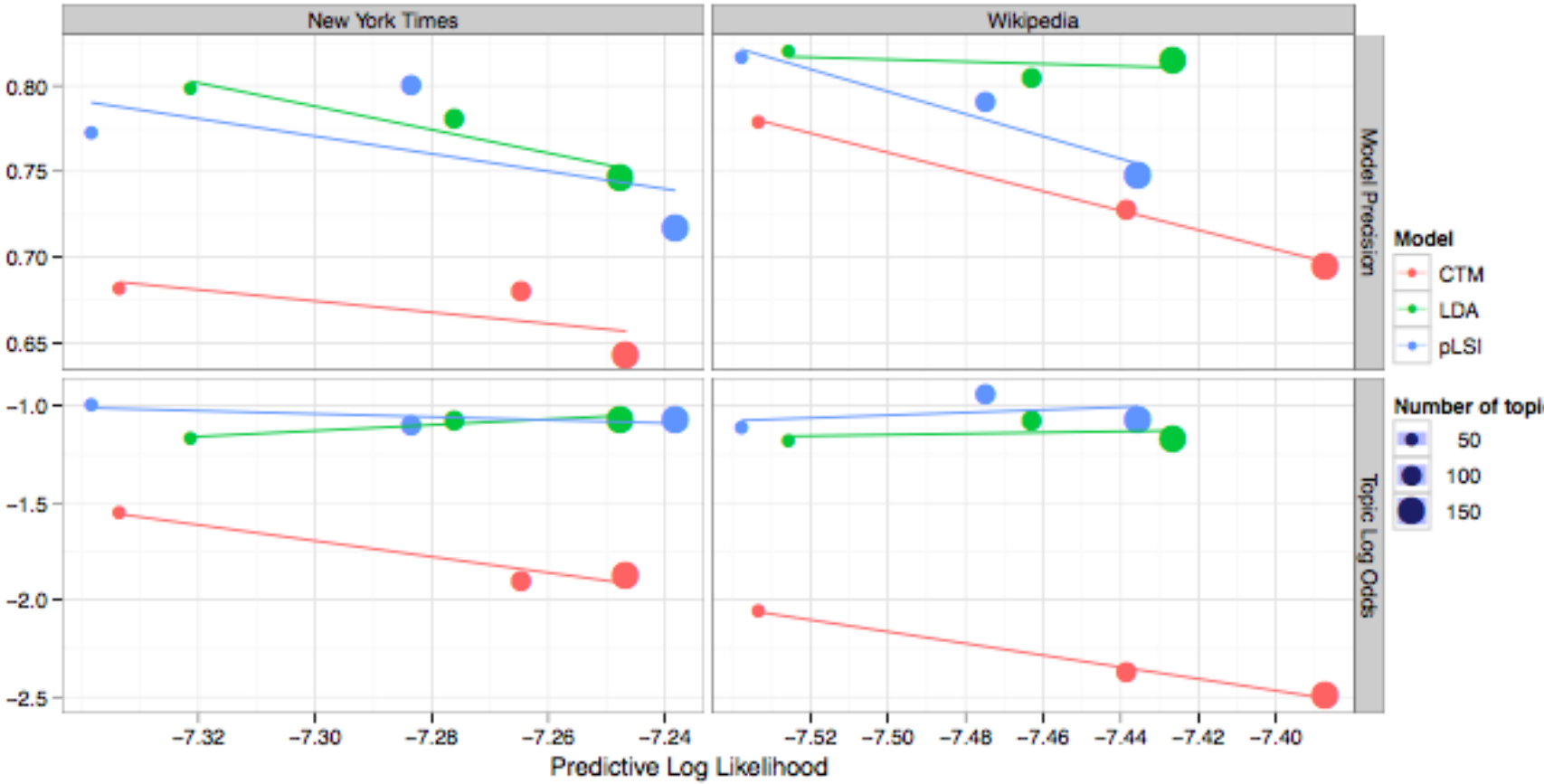


Figure 5: A scatter plot of model precision (top row) and topic log odds (bottom row) vs. predictive log likelihood. Each point is colored by model and sized according to the number of topics used to fit the model. Each model is accompanied by a regression line. Increasing likelihood does not increase the agreement between human subjects and the model for either task (as shown by the downward-sloping regression lines).

Takeaway

Topics seem objectively interpretable in many contexts.

Tradeoff between goodness-of-fit and interpretablility, which is generally more important in social science.

Active area of research assessing LDA models in terms of topic coherence.

Newman et. al. (2010) propose a method based on mutual pointwise information between top words in topics as computed via co-occurrence in Wikipedia.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Introduction to network science & visualisation[1]

Kimmo Soramäki,

Financial Network Analytics

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

FNA

Introduction to Network Science & Visualization I

Dr. Kimmo Soramäki
Founder & CEO, FNA

Agenda

Network Science

- Introduction
- Key concepts

Exposure Networks

- OTC Derivatives
- CCP Interconnectedness

Correlation Networks

- Housing Bubble and Crisis
- US Presidential Election

Network Science and Graphs Analytics

Is already powering the best known AI applications

Knowledge Graph

Social Graph

Product Graph

Economic Graph

Knowledge Graph

Payment Graph

Network Science and Graphs Analytics

“Goldman Sachs takes a DIY approach to graph analytics”

For enhanced compliance and fraud detection
(www.TechTarget.com, Mar 2015).

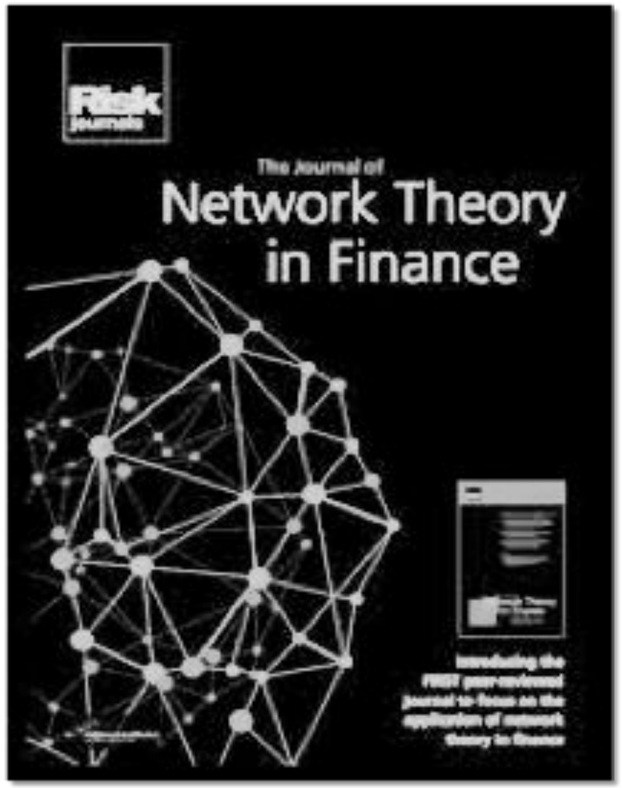
“PayPal relies on graph techniques to perform sophisticated fraud detection”

Saving them more than $700 million and enabling them to perform predictive fraud analysis, according to the IDC
(www.globalbankingandfinance.com, Jan 2016)

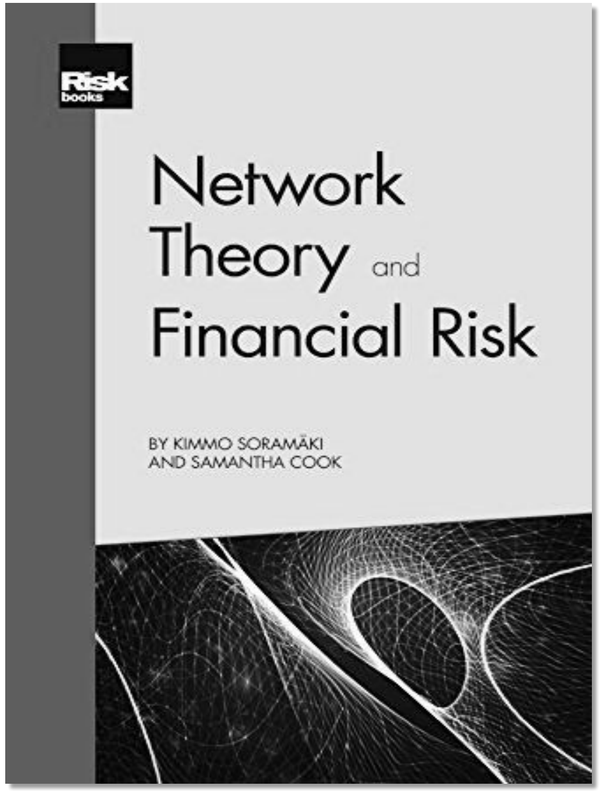
"Network diagnostics .. may displace atomised metrics such as VaR”

Regulators are increasing using network science for financial stability analysis.
(Andy Haldane, Bank of England Executive Director)

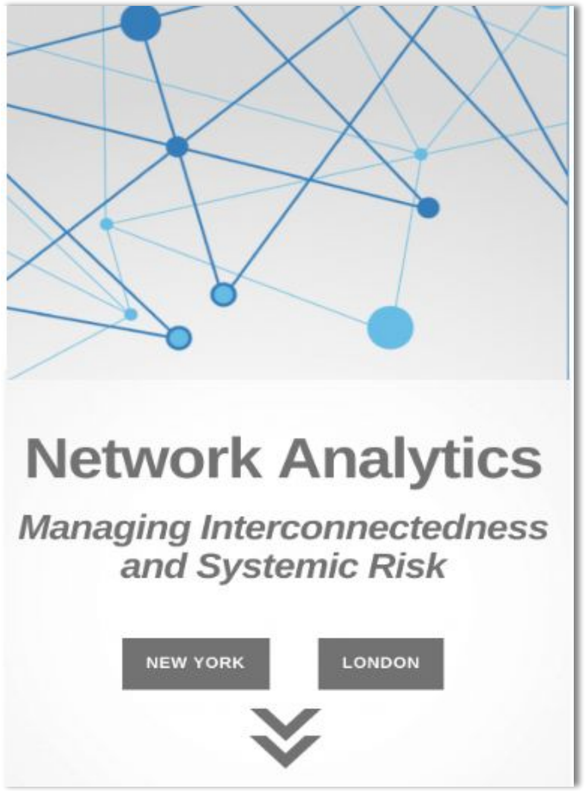
Further Resources on Network Analytics and Systemic Risk



Risk Journal founded by Kimmo Soramäki | link

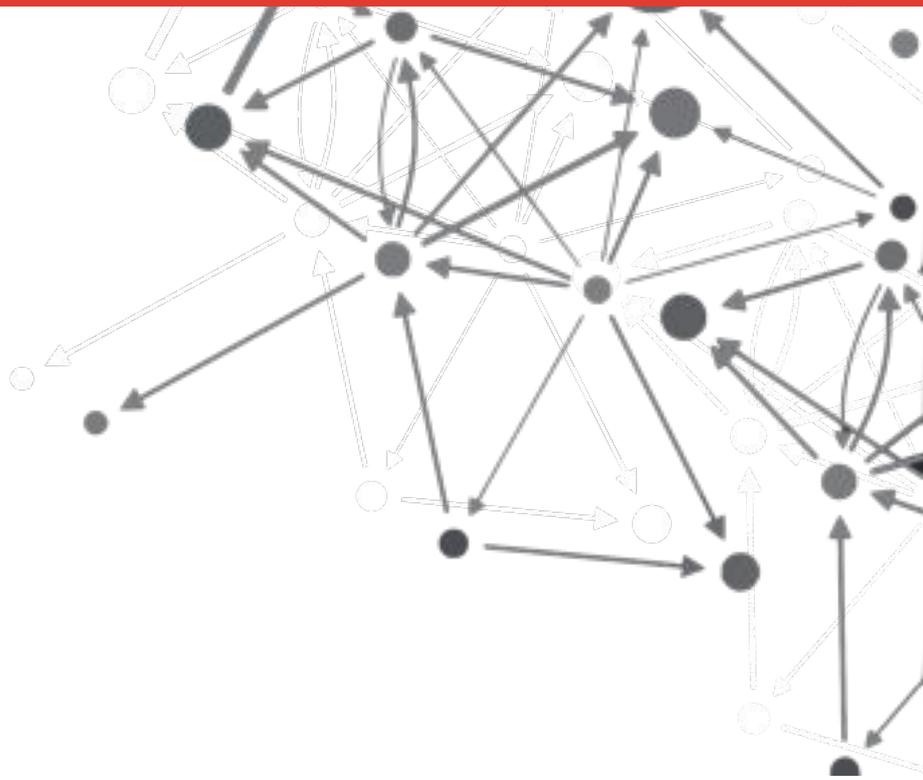


Risk Book by Kimmo Soramäki and Samantha Cook, FNA's Chief Scientist | link



Two-day Training course in London, New York and Singapore, instructed by Kimmo Soramäki | link

Network Theory is about



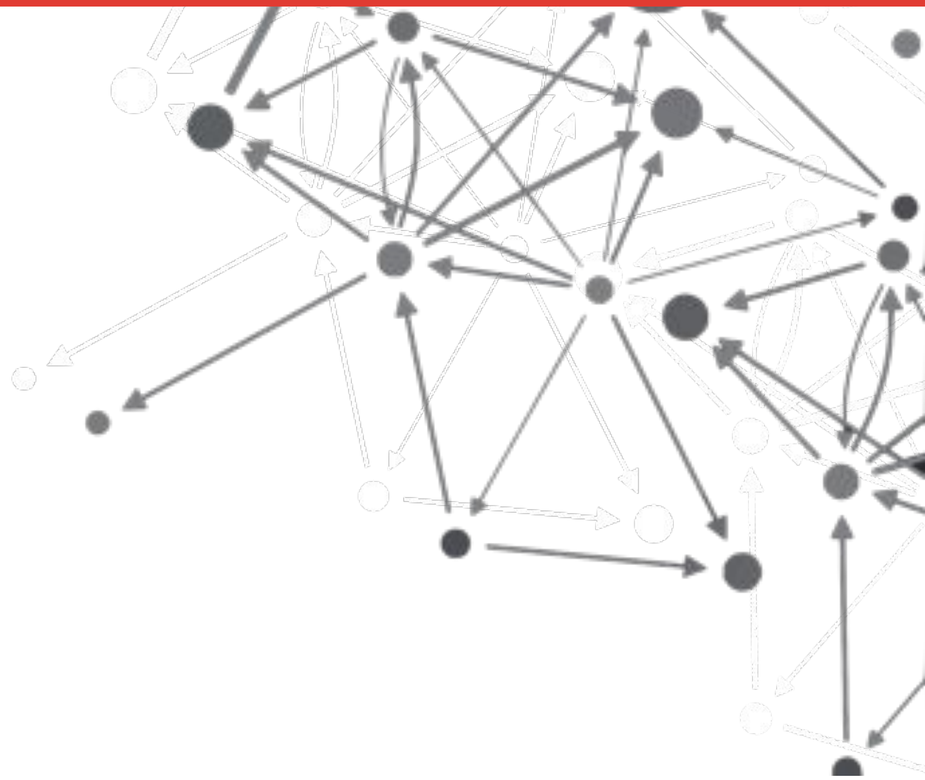
New Way of Looking at Data

- How is data connected with other data?
- How do these connections matter?
- How do complex systems move in time?

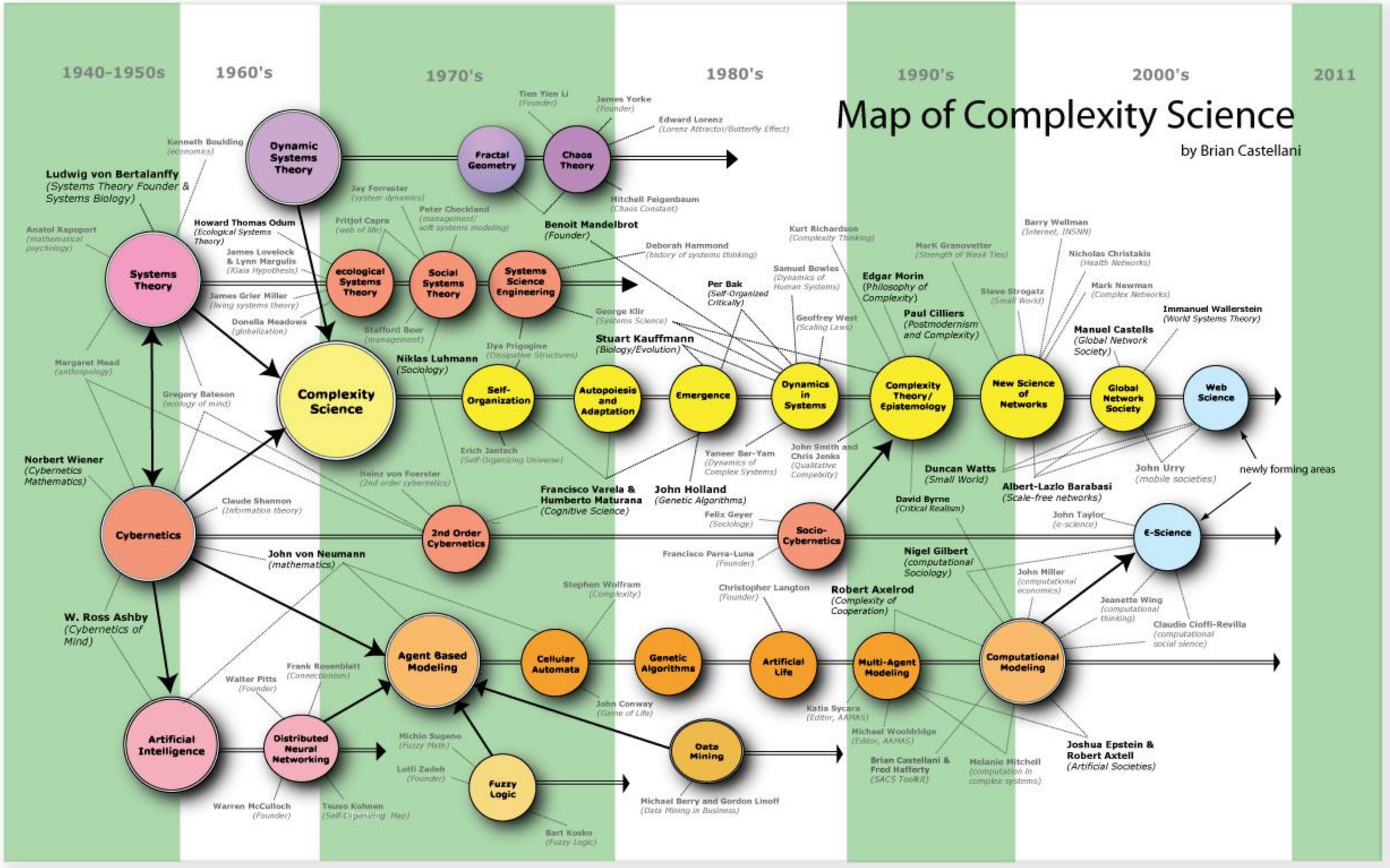
For the first time we are able to measure and model this!

Complex Systems

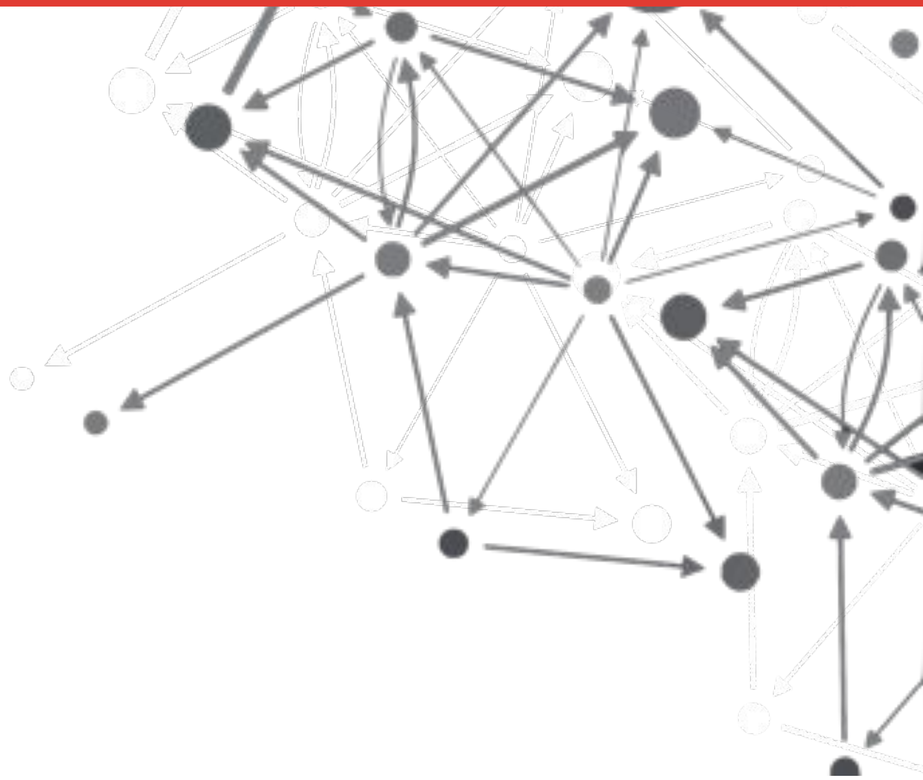
"Systems with rich interactions between the components of the system"



eg. financial markets, payment systems, road systems, friendship networks, ... almost **every** socio-economic system.

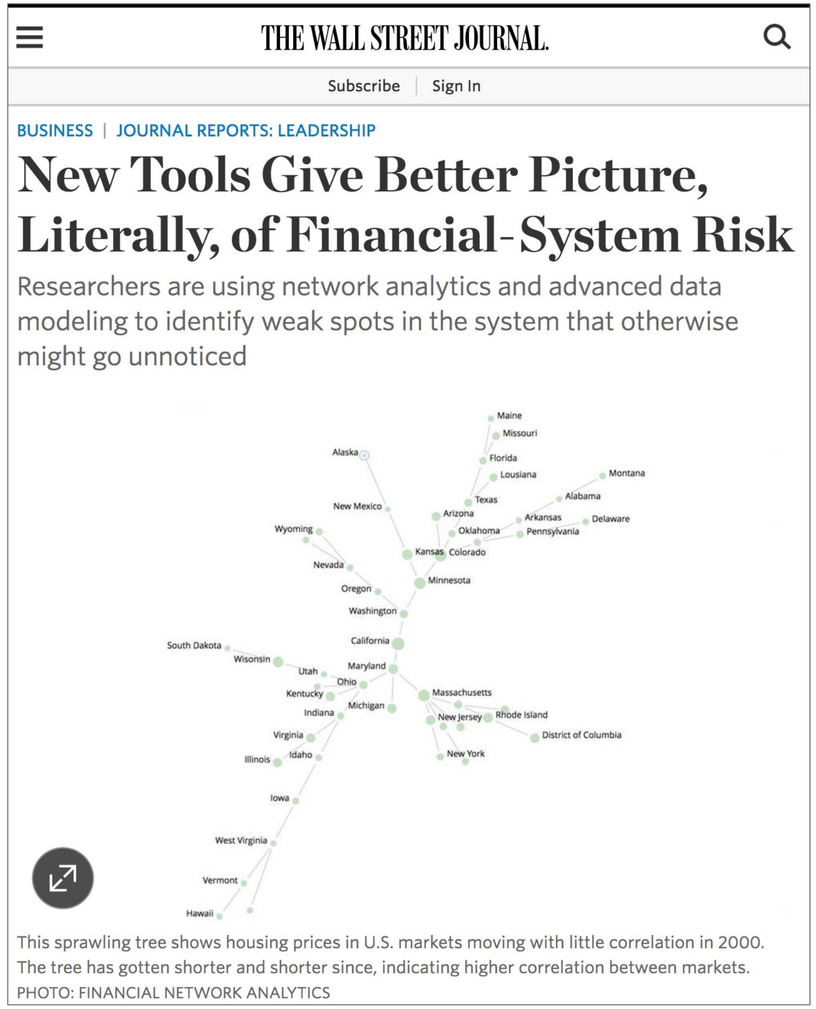

1940-1950s
1960's
1970's
1980's
1990's
2000's
2011
Map of Complexity Science
by Brian Castellani
Dynamic Systems Theory
Fractal Geometry
Chaos Theory
Systems Theory
ecological Systems Theory
Social Systems Theory
Systems Science Engineering
Complexity Science
Self-Organization
Autopoiesis and Adaptation
Emergence
Dynamics in Systems
Complexity Theory/ Epistemology
New Science of Networks
Global Network Society
Web Science
Cybernetics
2nd Order Cybernetics
Socio-Cybernetics
€-Science
Agent Based Modeling
Cellular Automata
Genetic Algorithms
Artificial Life
Multi-Agent Modeling
Computational Modeling
Artificial Intelligence
Distributed Neural Networking
Fuzzy Logic
Data Mining
Ludwig von Bertalanffy (Systems Theory Founder & Systems Biology)
Kenneth Boulding (economics)
Howard Thomas Odum (Ecological Systems Theory)
Anatol Rapoport (mathematical psychology)
James Lovelock & Lynn Margulis (Gaia Hypothesis)
James Grier Miller (living systems theory)
Donella Meadows (globalization)
Margaret Mead (anthropology)
Gregory Bateson (ecology of mind)
Norbert Wiener (Cybernetics Mathematics)
Claude Shannon (Information theory)
John von Neumann (mathematics)
W. Ross Ashby (Cybernetics of Mind)
Walter Pitts (Founder)
Frank Rosenblatt (Connectionism)
Warren McCulloch (Founder)
Teuvo Kohnen (Self-Organizing Map)
Michio Sugeno (Fuzzy Math)
Lotfi Zadeh (Founder)
Bart Kosko (Fuzzy Logic)
Tien Yien Li (Founder)
James Yorke (Founder)
Edward Lorenz (Lorenz Attractor/Butterfly Effect)
Mitchell Feigenbaum (Chaos Constant)
Jay Forrester (system dynamics)
Fritjof Capra (web of life)
Peter Checkland (management/ soft systems modeling)
Stafford Beer (management)
Benoit Mandelbrot (Founder)
Deborah Hammond (history of systems thinking)
George Klir (Systems Science)
Niklas Luhmann (Sociology)
Ilya Prigogine (Dissipative Structures)
Stuart Kauffmann (Biology/Evolution)
Erich Jantsch (Self-Organizing Universe)
Heinz von Foerster (2nd order cybernetics)
Francisco Varela & Humberto Maturana (Cognitive Science)
John Holland (Genetic Algorithms)
Yaneer Bar-Yam (Dynamics of Complex Systems)
Per Bak (Self-Organized Criticality)
Kurt Richardson (Complexity Thinking)
Samuel Bowles (Dynamics of Human Systems)
Geoffrey West (Scaling Laws)
John Smith and Chris Jonks (Qualitative Complexity)
Edgar Morin (Philosophy of Complexity)
Paul Cilliers (Postmodernism and Complexity)
Duncan Watts (Small World)
David Byrne (Critical Realism)
Felix Geyer (Sociology)
Francisco Parra-Luna (Founder)
Mark Granovetter (Strength of Weak Ties)
Steve Strogatz (Small World)
Barry Wellman (Internet, INSNN)
Nicholas Christakis (Health Networks)
Mark Newman (Complex Networks)
Manuel Castells (Global Network Society)
Immanuel Wallerstein (World Systems Theory)
Albert-Lazlo Barabasi (Scale-free networks)
John Urry (mobile societies)
newly forming areas
John Taylor (e-science)
Nigel Gilbert (computational Sociology)
John Miller (computational economics)
Jeanette Wing (computational thinking)
Claudio Cioffi-Revilla (computational social sience)
Joshua Epstein & Robert Axtell (Artificial Societies)
Robert Axelrod (Complexity of Cooperation)
Christopher Langton (Founder)
Stephen Wolfram (Complexity)
John Conway (Game of Life)
Katia Sycara (Editor, AAMAS)
Michael Wooldridge (Editor, AAMAS)
Brian Castellani & Fred Hafferty (SACS Toolkit)
Melanie Mitchell (computation in complex systems)
Michael Berry and Gordon Linoff (Data Mining in Business)

Main Modes of Analysis



- Top Down Analysis
- Bottom Up Analysis
- Features of Data
- Agent Based Models

Top Down Analysis



THE WALL STREET JOURNAL.

Subscribe | Sign In

BUSINESS | JOURNAL REPORTS: LEADERSHIP

New Tools Give Better Picture, Literally, of Financial-System Risk

Researchers are using network analytics and advanced data modeling to identify weak spots in the system that otherwise might go unnoticed

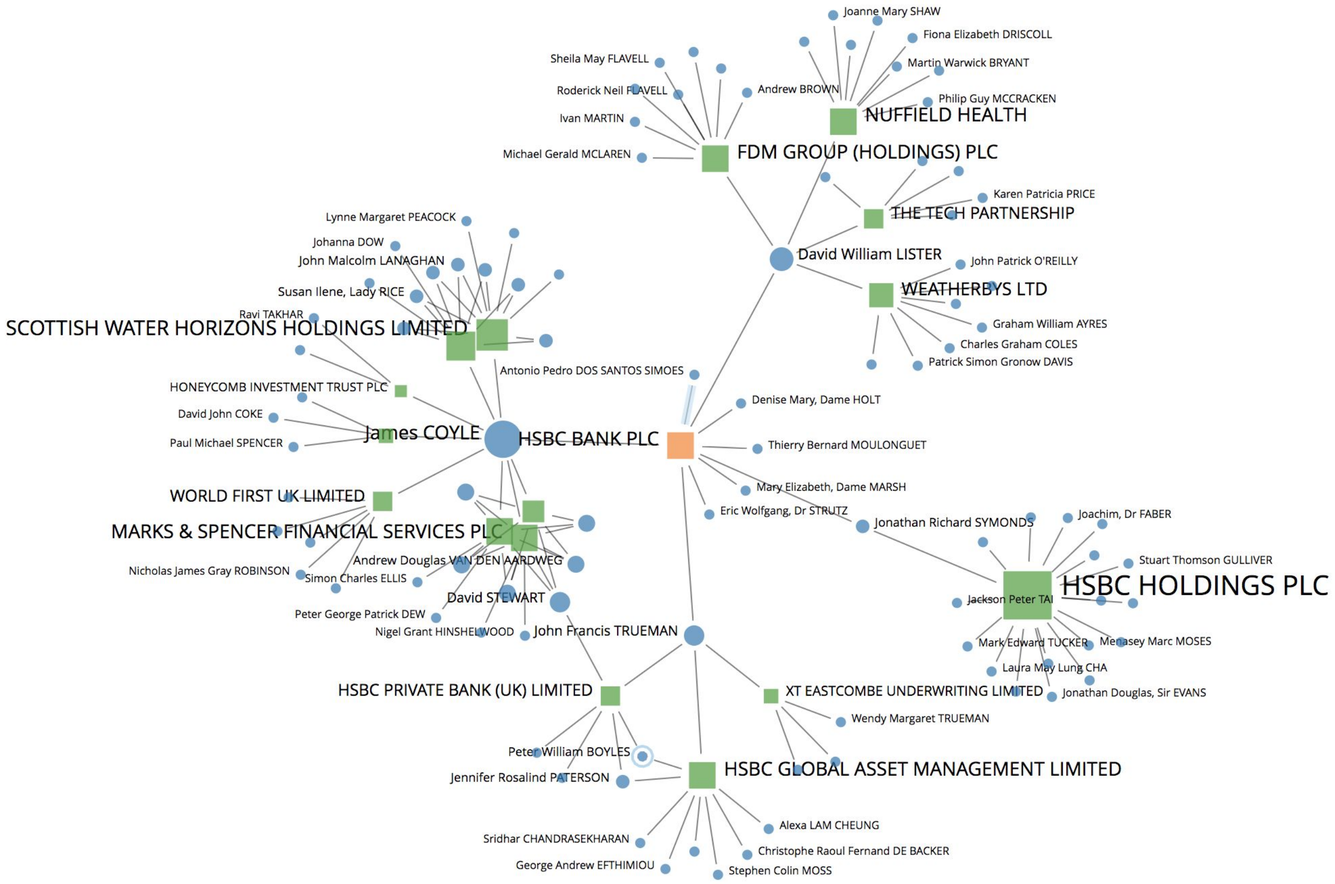
This sprawling tree shows housing prices in U.S. markets moving with little correlation in 2000. The tree has gotten shorter and shorter since, indicating higher correlation between markets.

PHOTO: FINANCIAL NETWORK ANALYTICS

Typical use cases:

- Systemic risk analysis
- System monitoring
- System design
- System stress testing
- Clustering/Classification
- Early warning
- Anomaly detection

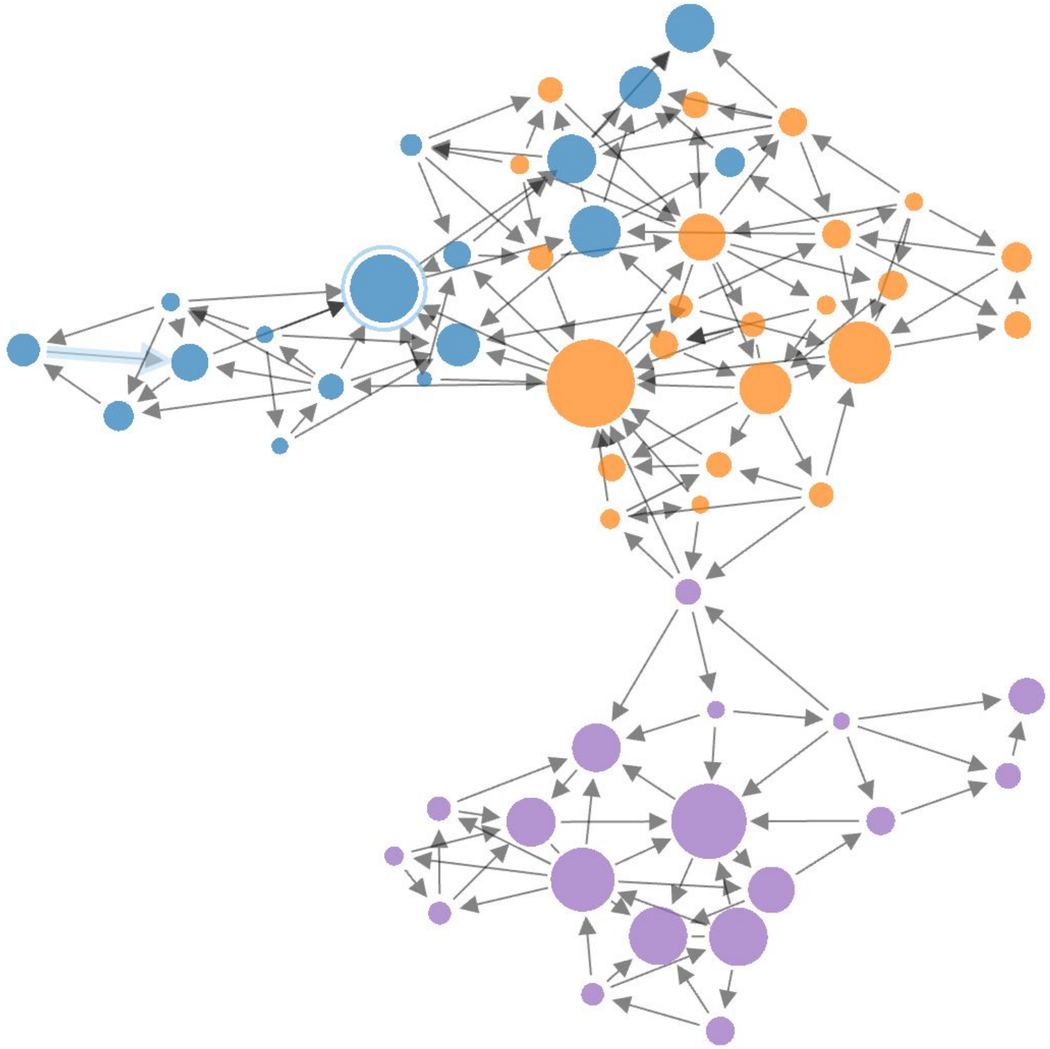
Bottom Up Analysis



Typical use cases:

- Criminal investigation
- Terrorist networks
- Money laundering
- KYC & KYCC
- Fundamental investment analysis
- Supply chain analysis

Network Features of Data

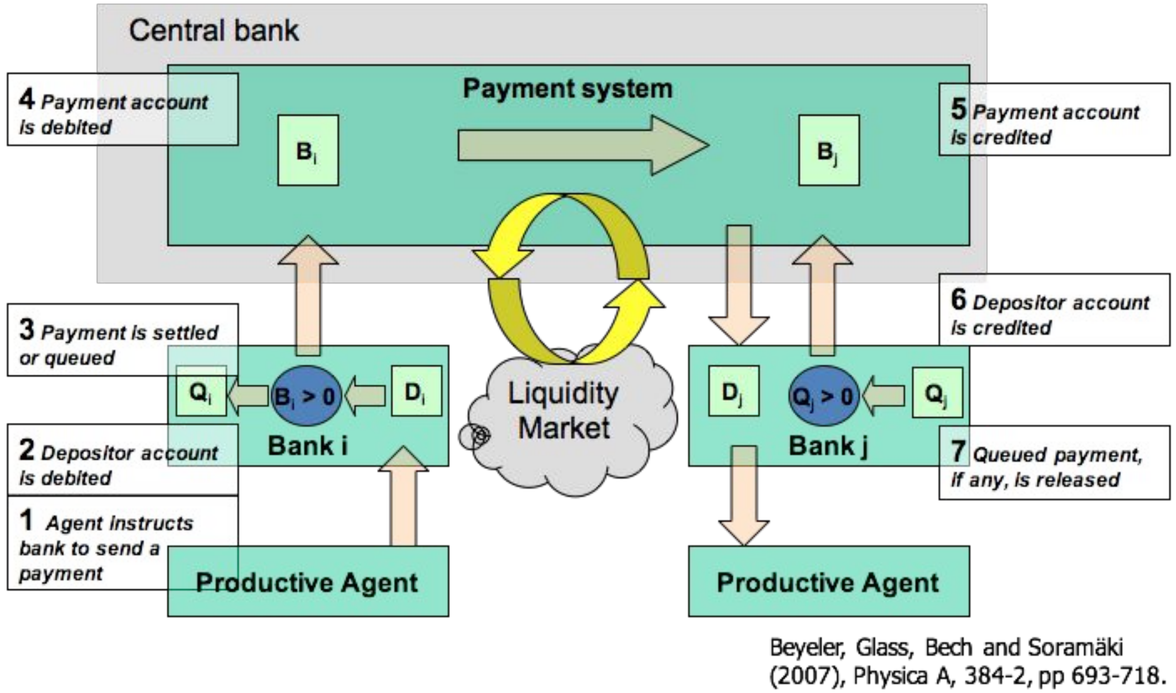


Typical use cases:

- AI/ML
- Fraud algorithms
- Recommendation engines
- Algorithmic investment

FNA Research: Comparison of Graph Computing Platform Performance

Agent Based Models



Beyeler, Glass, Bech and Soramäki (2007), Physica A, 384-2, pp 693-718.

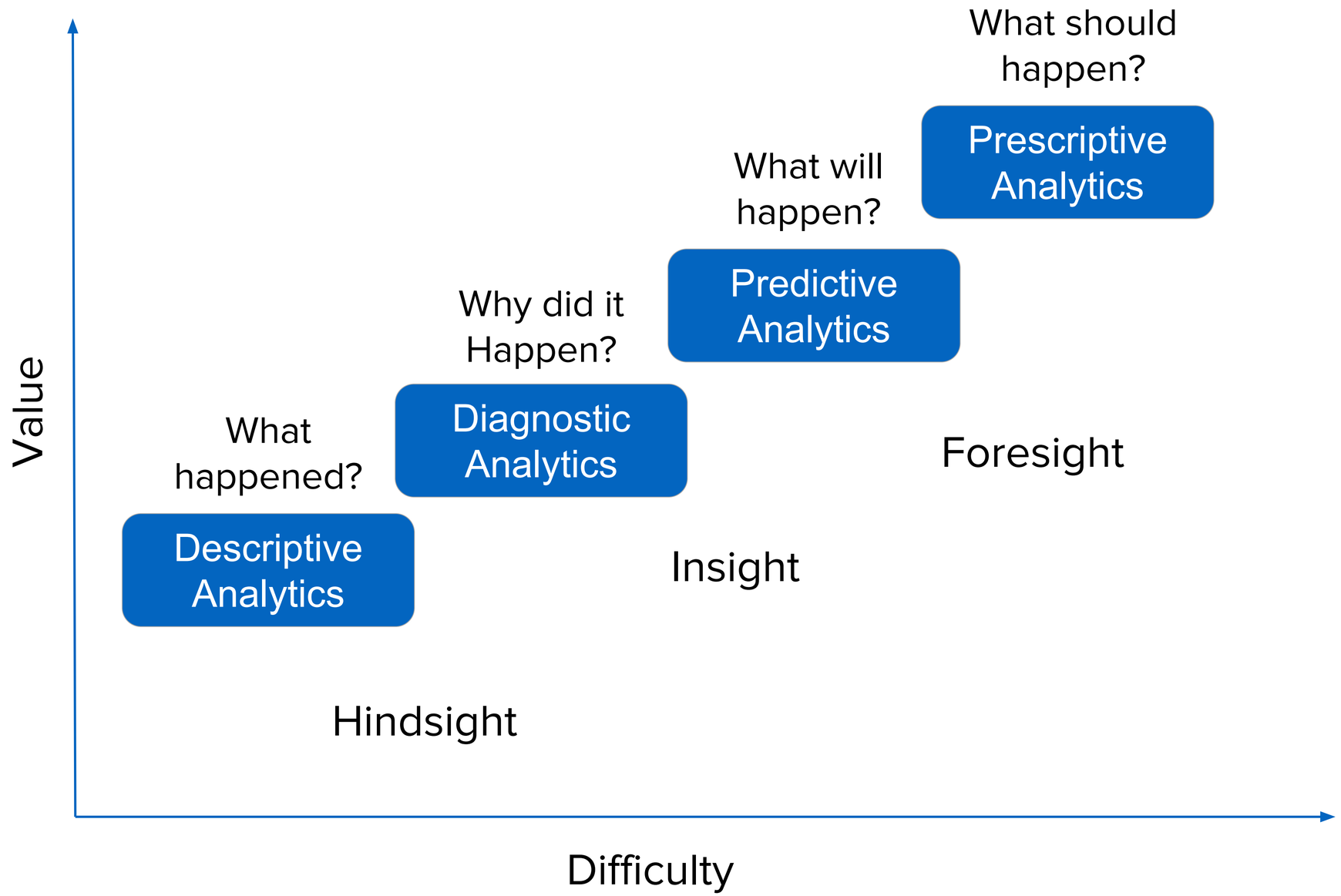
Typical use cases:

- Central Counterparty Clearing
- Payment Systems
- FX Settlement
- Financial Markets
- Housing Markets

Journey

Advanced Analytics	Monitoring	Simulation
1	2	3
Understand interconnectedness, data visualization	Detect risks and anomalies in real-time	What-if analysis, failure simulations & remediation scenarios
Identify Risks Concentrations	Provide Early Warning	Predict Outcomes

Analytical Framework



FNA

Types of Networks

www.fna.fi

Network Concepts

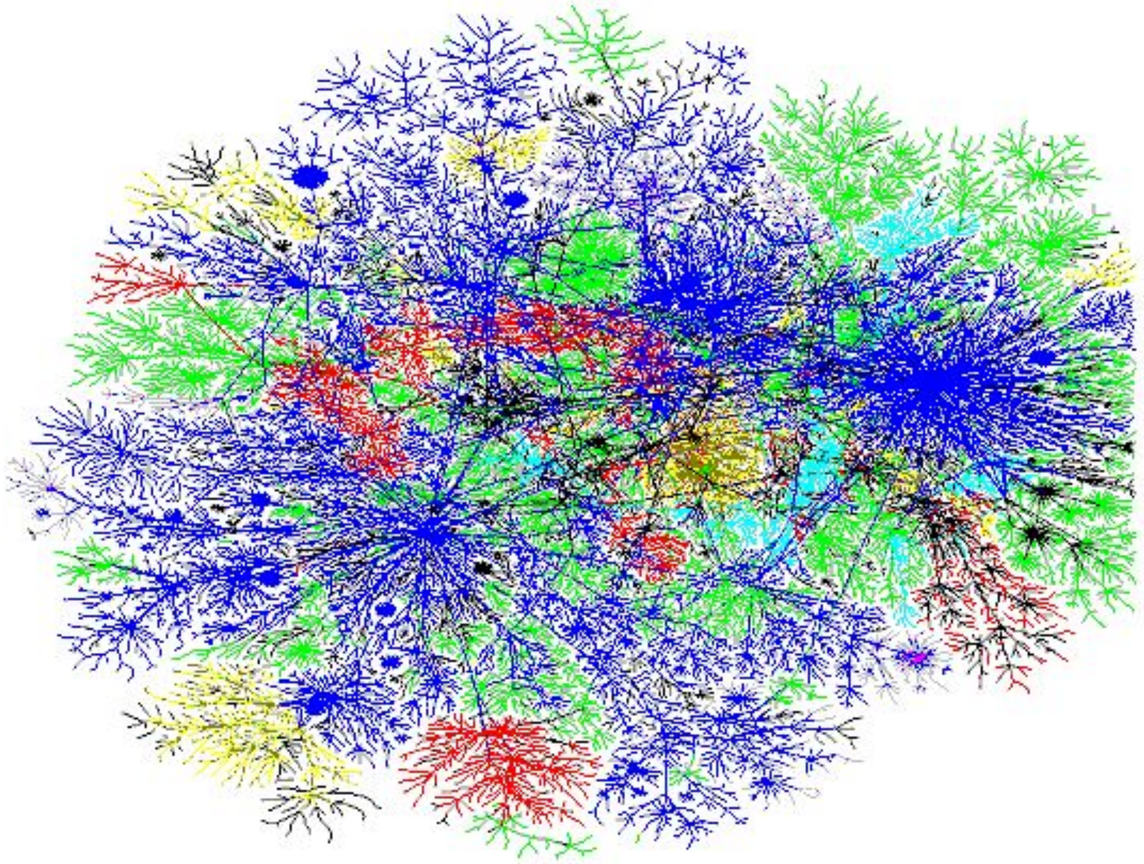
Constituents

- Networks (graphs)
- Nodes (vertices)
- Links (ties, edges or arcs)

Links can be

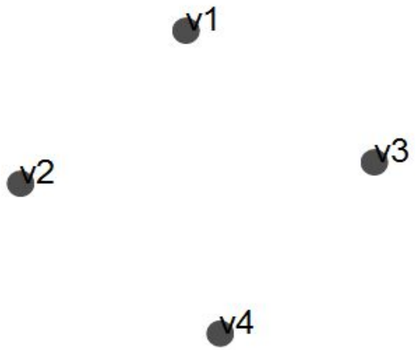
- Directed (arcs) vs undirected (edges, ties)
- Weighed vs unweighted

Graph + properties = Network

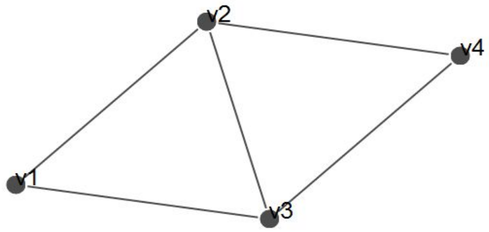


Some Graph Types

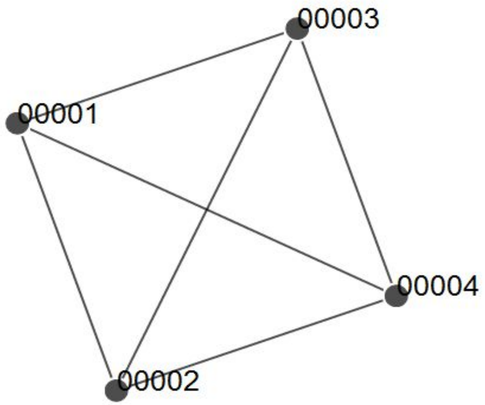
Trivial Graph



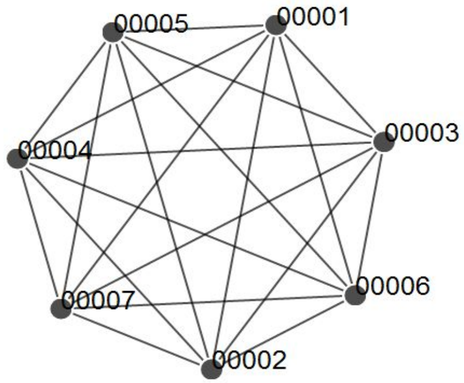
Empty Graph



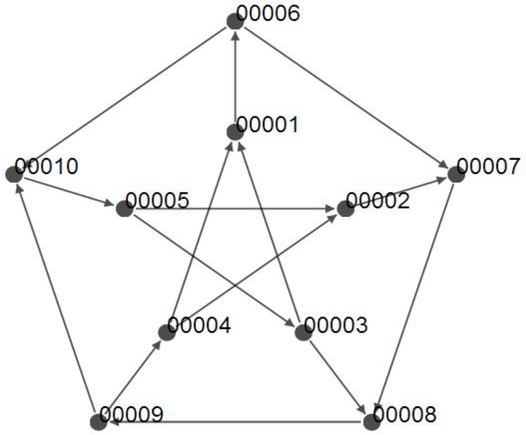
Undirected Graph



Complete graph, K4

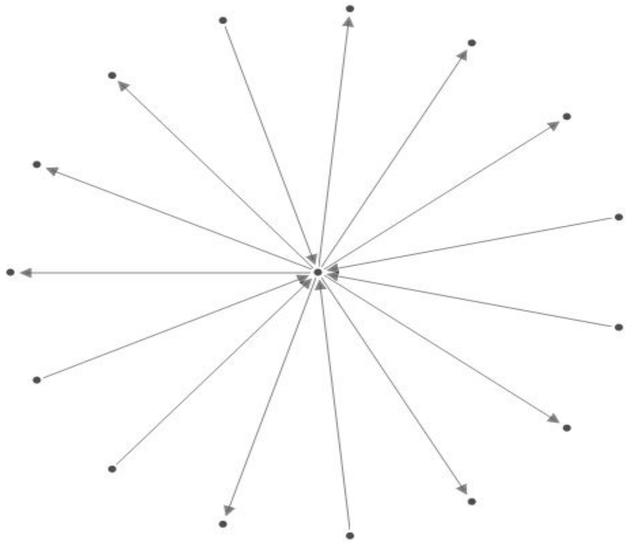


Complete graph, K7

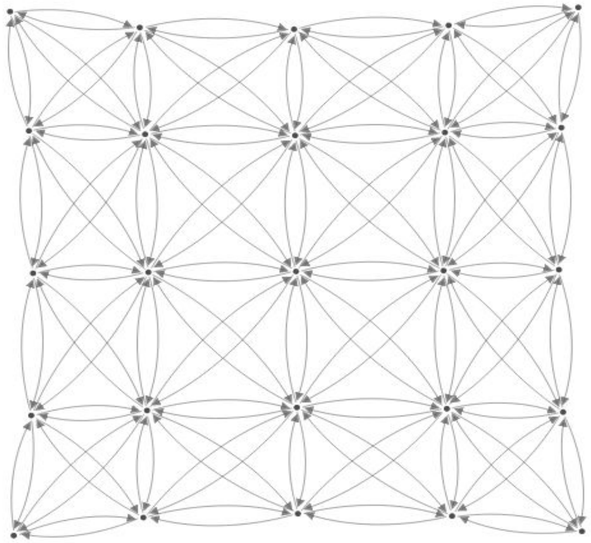


Directed Graph

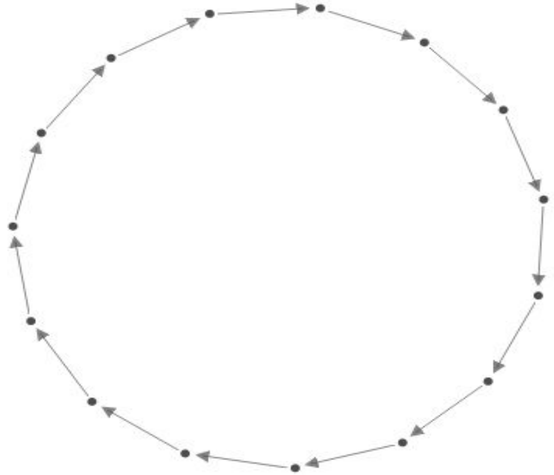
Simple Graphs / Non-random Graphs



Star

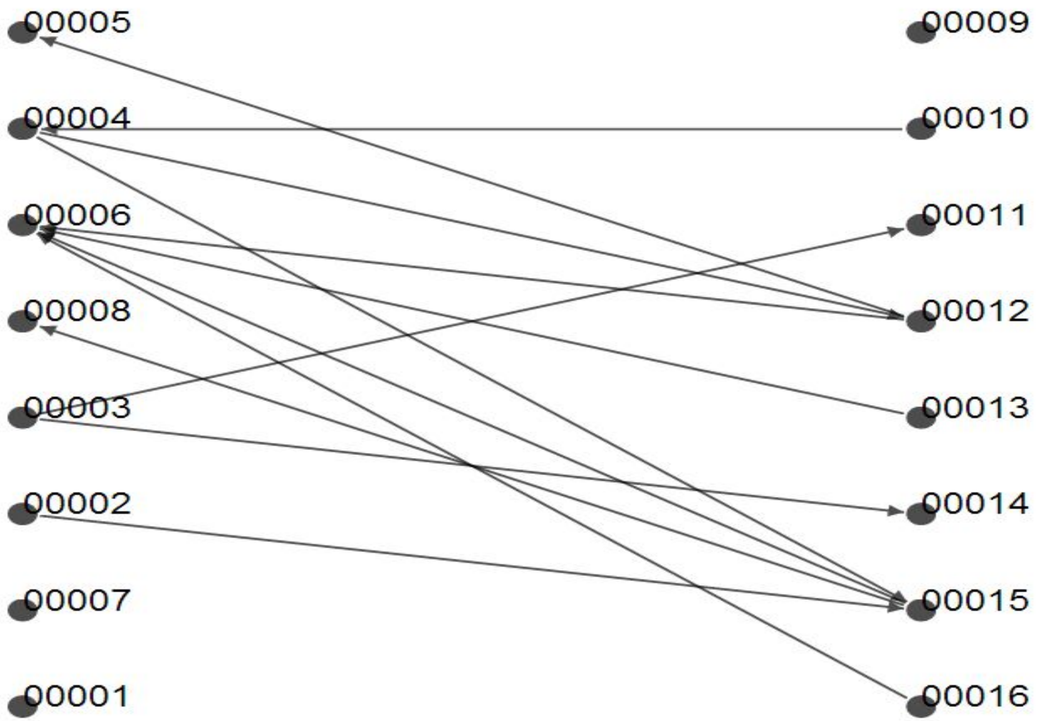


Lattice

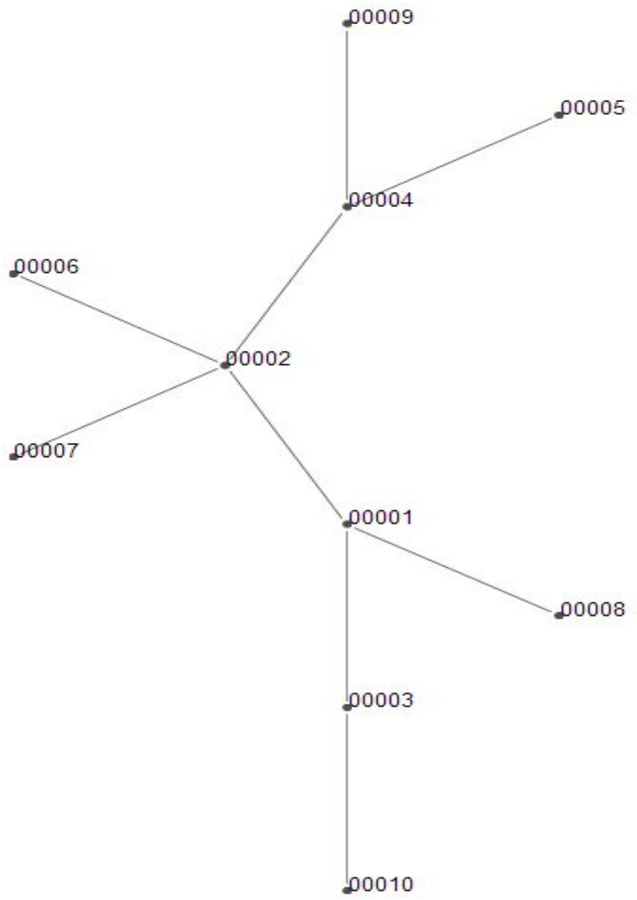


Ring

Simple Graphs / Non-random Graphs

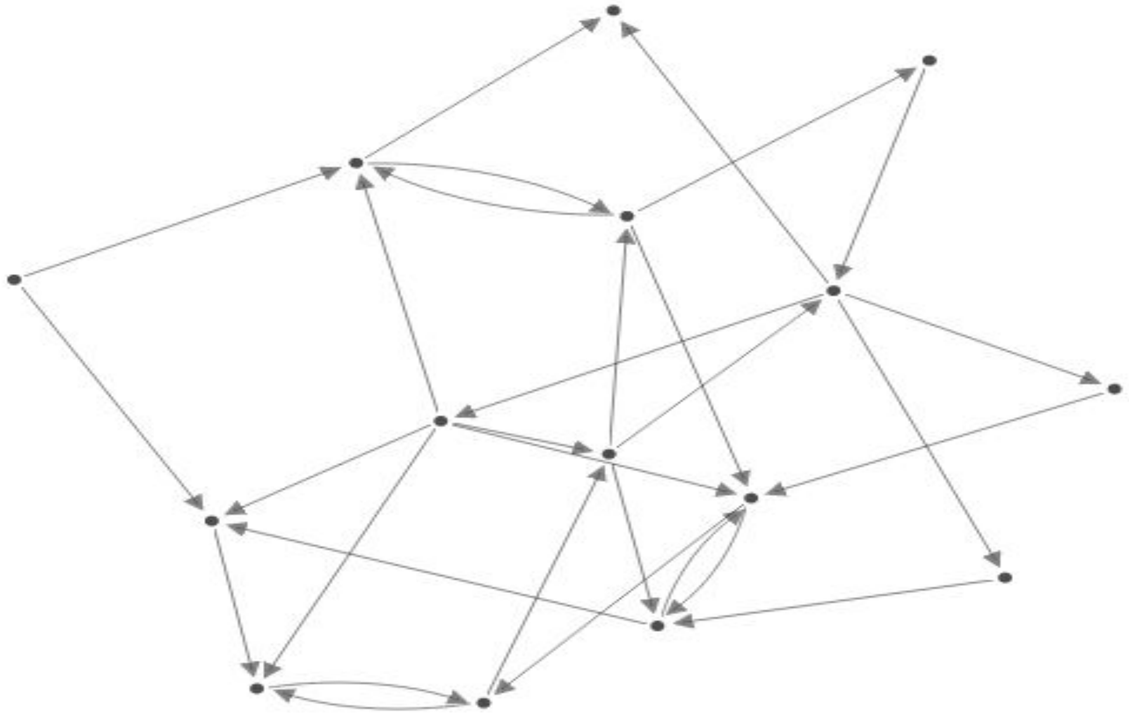


Bipartite

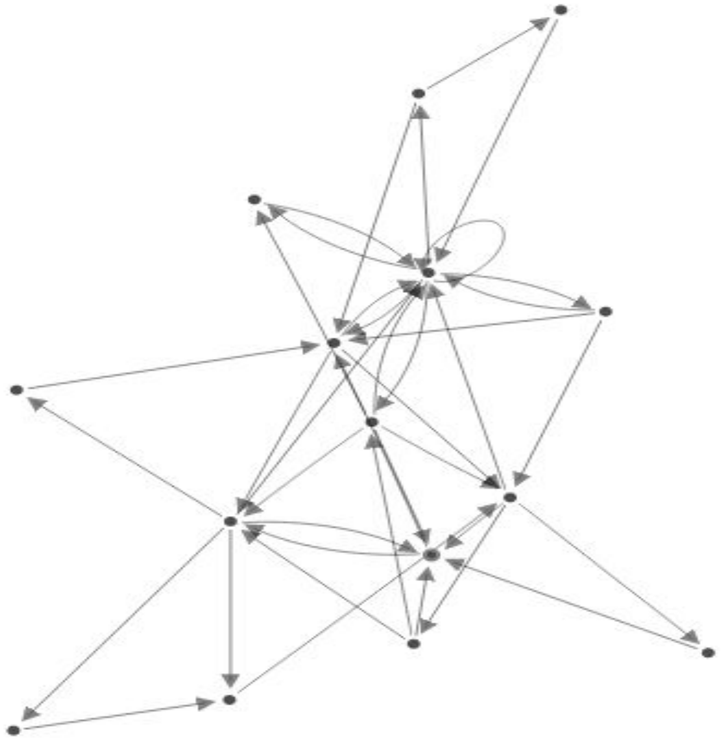


Tree

Random Graphs

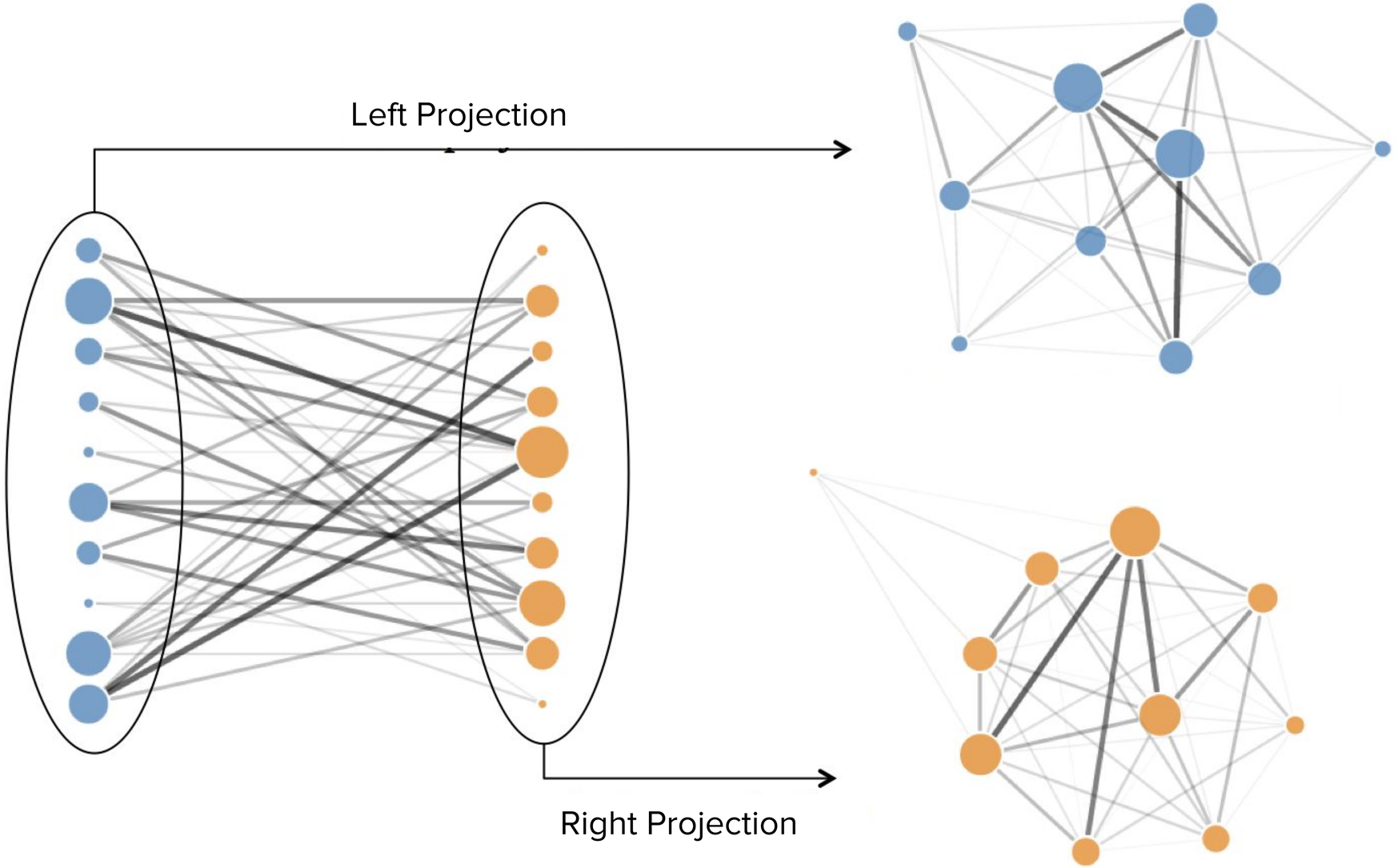


Random
(Erdos-Renyi)



Scale-free
(Barabasi-Albert)

Projection Networks



FNA

Key Concept: Centrality

Centrality

Centrality measures importance of nodes (or links) in a network.

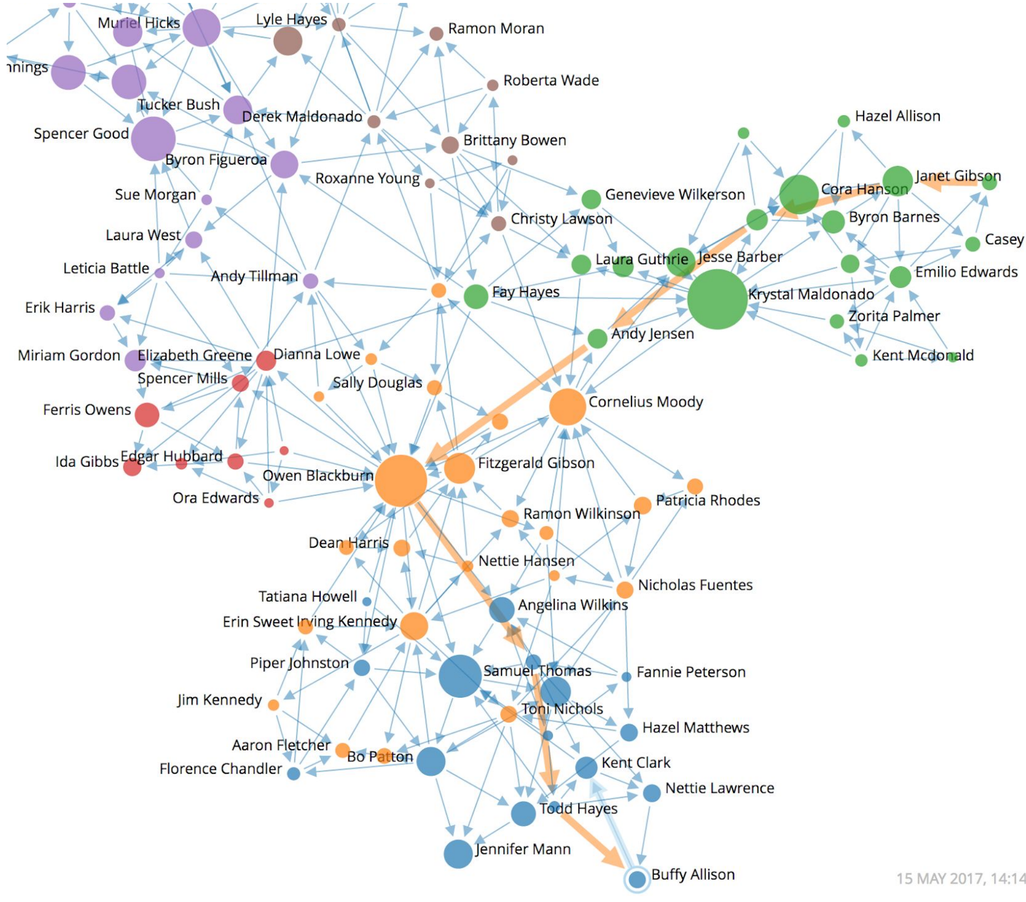
Depends on process that takes place in the network!

Trajectory

- Geodesic paths (shortest paths)
- Any path (visit no node twice)
- Trails (visit no link twice)
- Walks (free movement)

Transmission

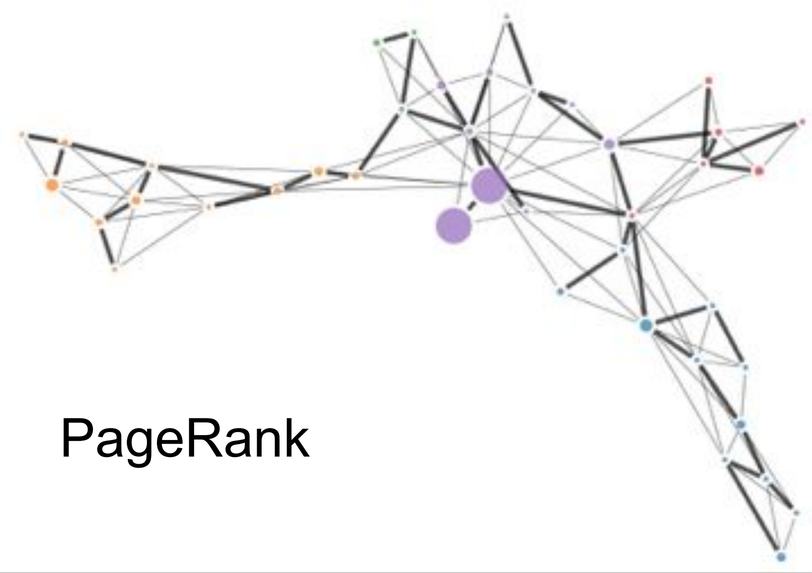
- Parallel duplication
- Serial duplication
- Transfer



DHL Package = Transfer via shortest path
Money = Transfer via random walks
Virus = Serial duplication via paths

etc.

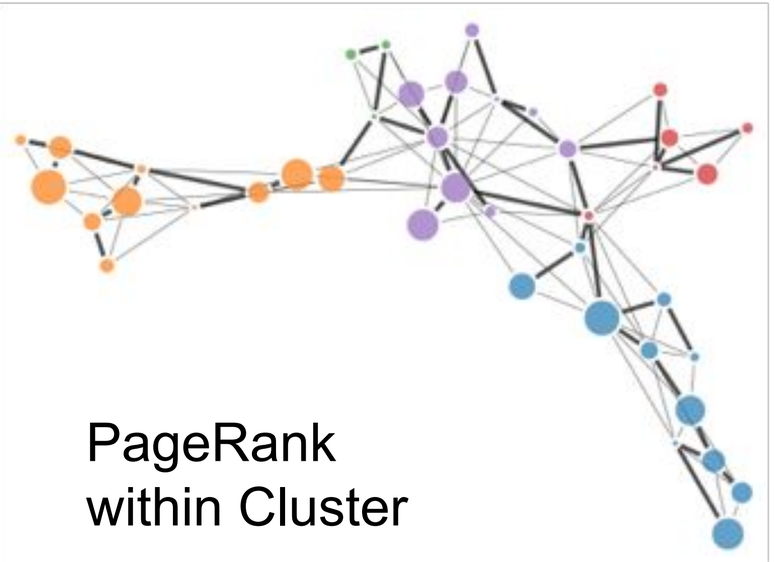
Common Centrality Metrics



PageRank

Betweenness

Closeness



PageRank
within Cluster

FNA

Communities

www.fna.fi

Community Detection

Often networks are large and complex and we want to simplify, categorize and label nodes into meaningful groups.

Community detection is an algorithmic way of doing this, and there are numerous methods available.

- Unsupervised learning, how do we know result is correct. What is correct?
- Which algorithm to choose?
- Some algorithms detect well large, but not small communities
- Is it a community or a cluster of several?
- What about overlapping communities?

Still more an art than a science. Try what works?

FNA Research: [Overview and Comparison of Community Detection Algorithms]

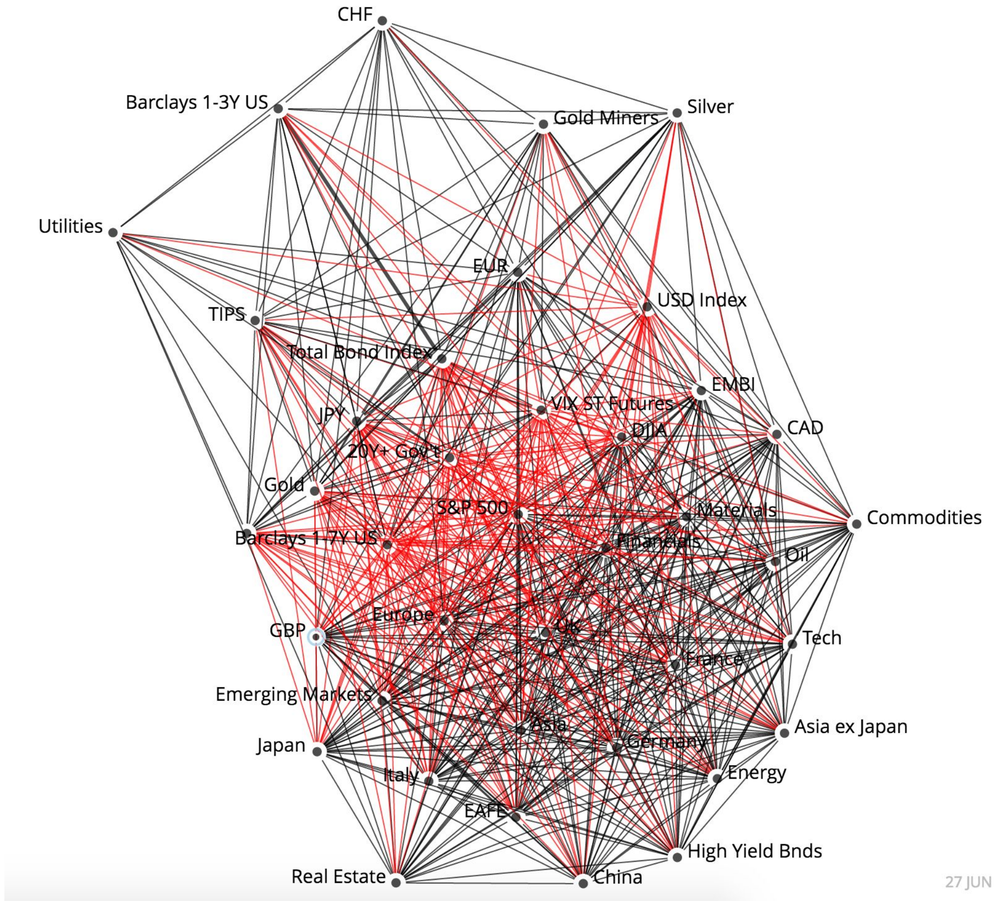
FNA

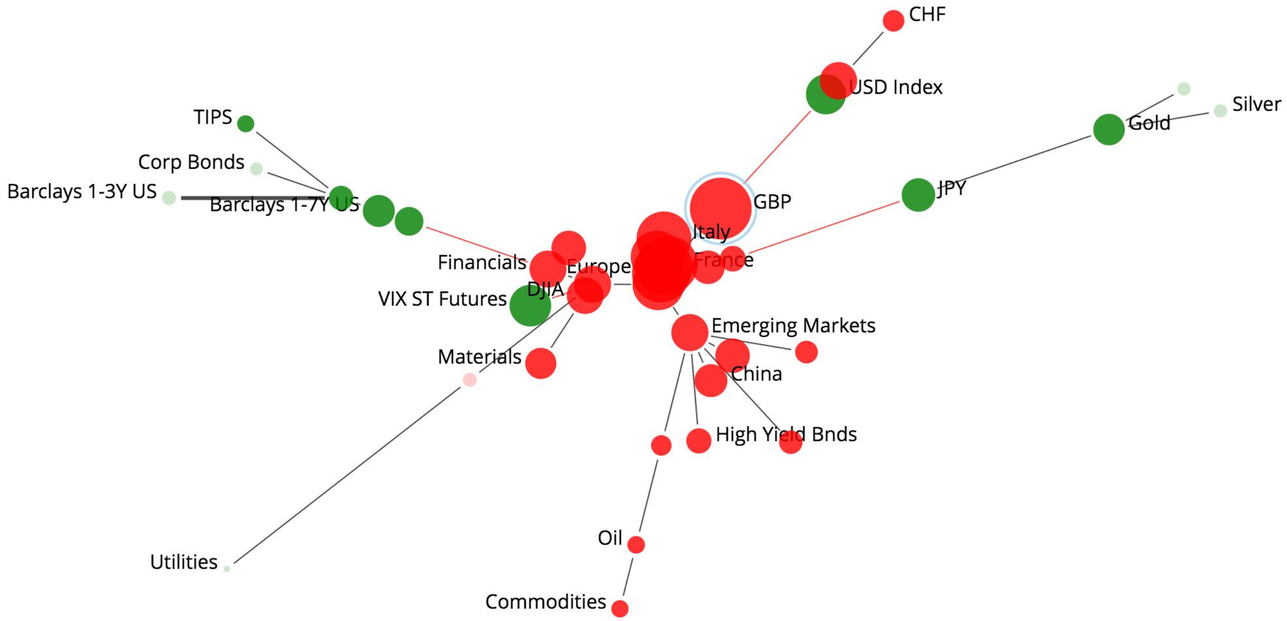
Filtering

www.fna.fi

Filtering

Often networks are large and complex and we want to filter out noise. Filtering techniques give solutions.

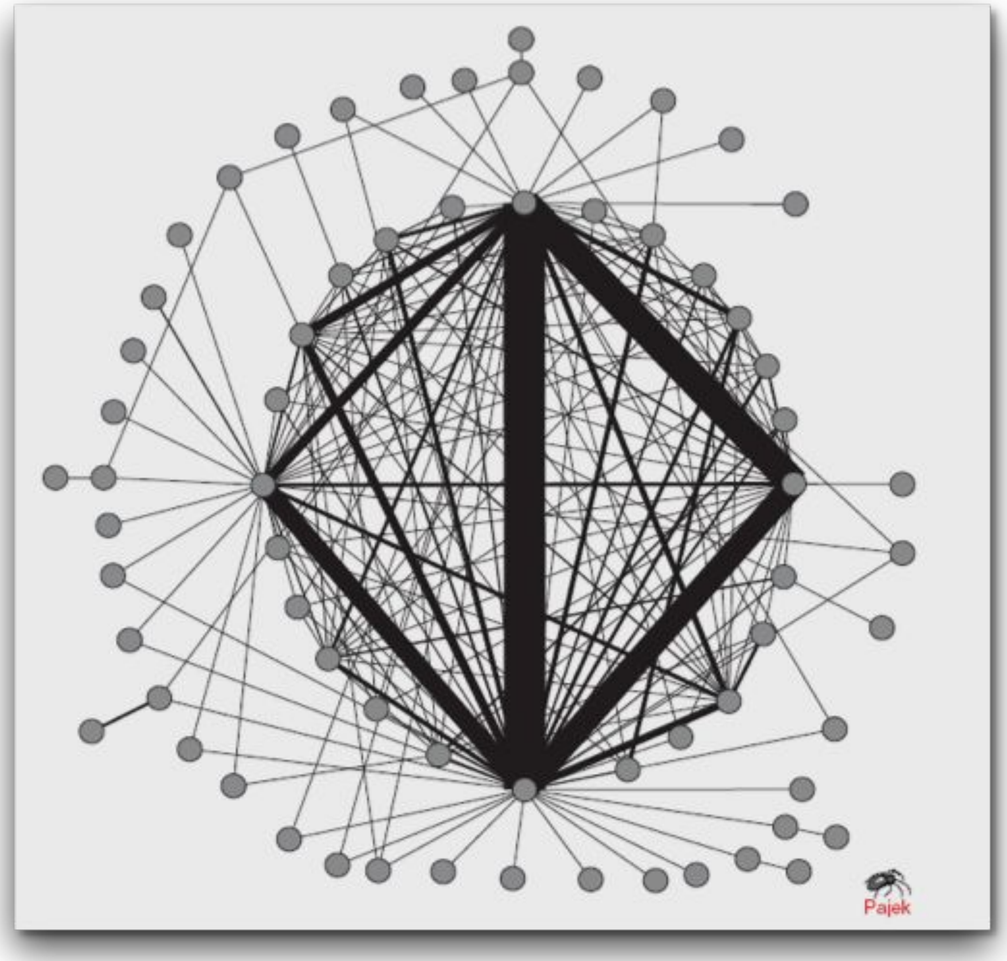




FNA

Top down analysis
Exposure Networks

First Financial Networks



Fedwire Interbank Payment Network (Fall 2001) was one of the first network views into any financial system.

Of a total of around 8000 banks, the 66 banks shown comprise 75% of total value. Of these, 25 banks completely connected

The research was subsequently used e.g. in congressional hearings to showcase the type of information that should be collected by financial institutions after the financial crisis.

Soramaki, K. M Bech, J. Arnold, R.J. Glass and W.E. Beyeler, The Topology of Interbank Payment Flows, Physica A, Vol. 379, pp 317-333, 2007.

Who pays whom?

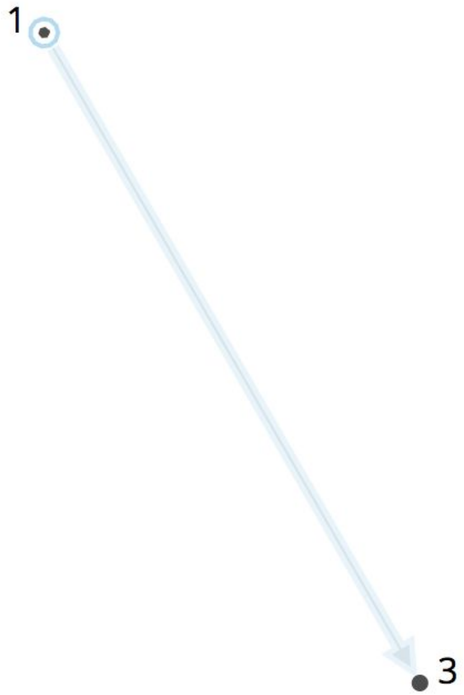
Our data contains:

date	buyer	seller	amount
26 July 2018	1	3	1
26 July 2018	3	5	1
26 July 2018	5	2	1
26 July 2018	3	4	1

...

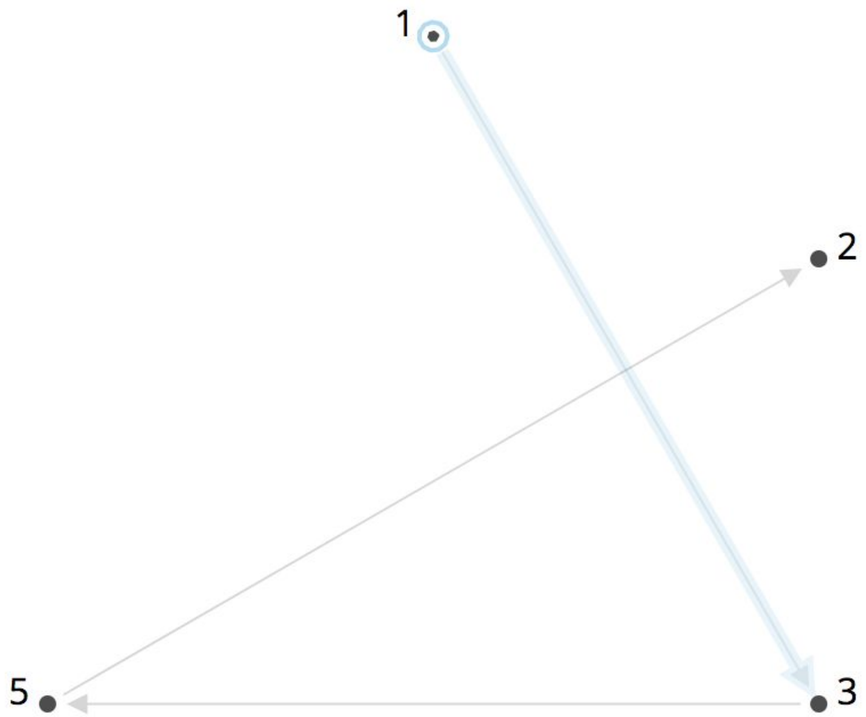
Who pays whom?

Payment from 1 to 3



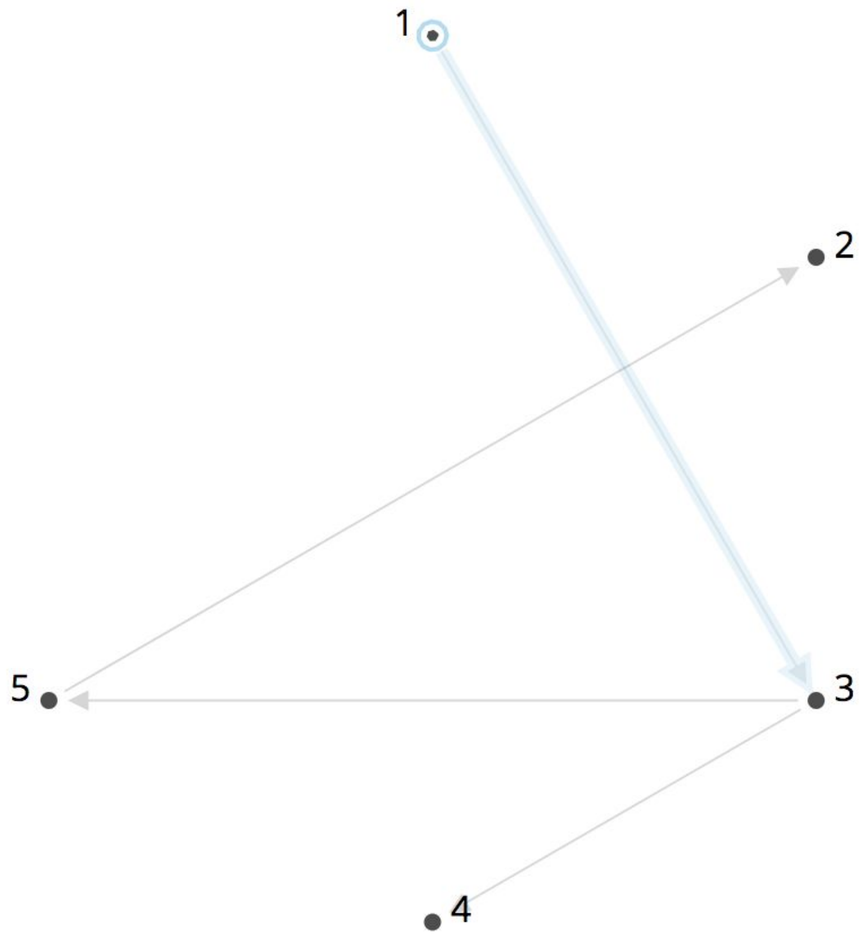
Who pays whom?

Payments from 3 to 4 and 5 to 2



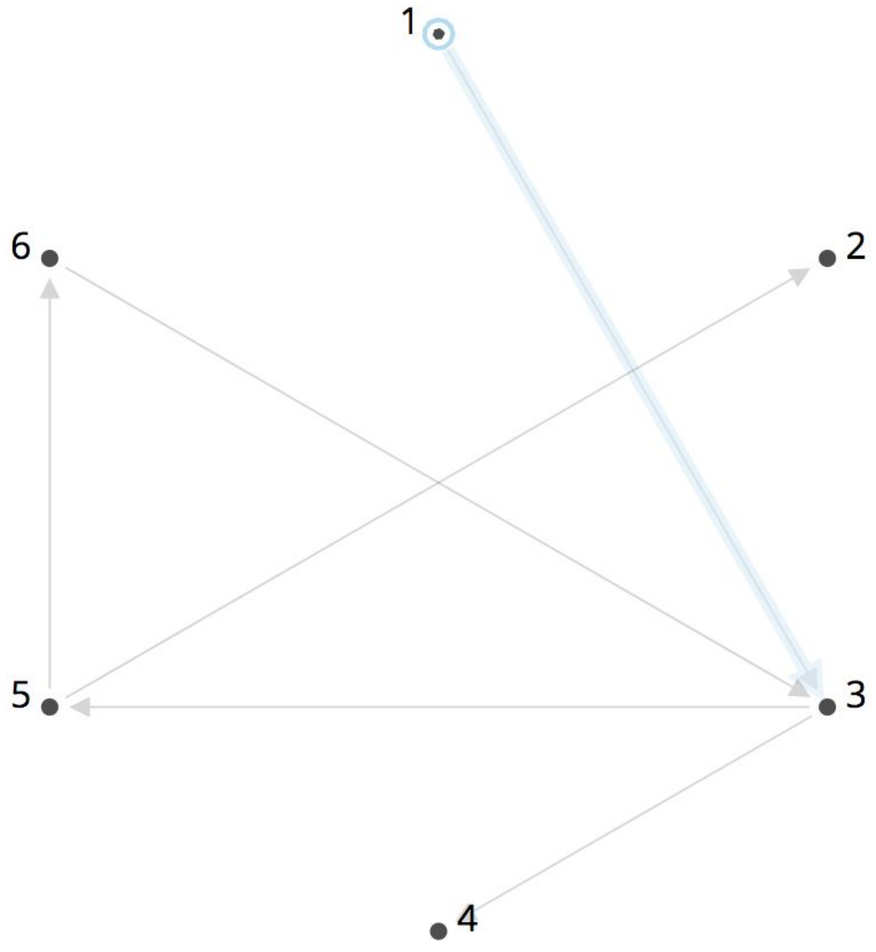
Who pays whom?

Payment from 3 to 4



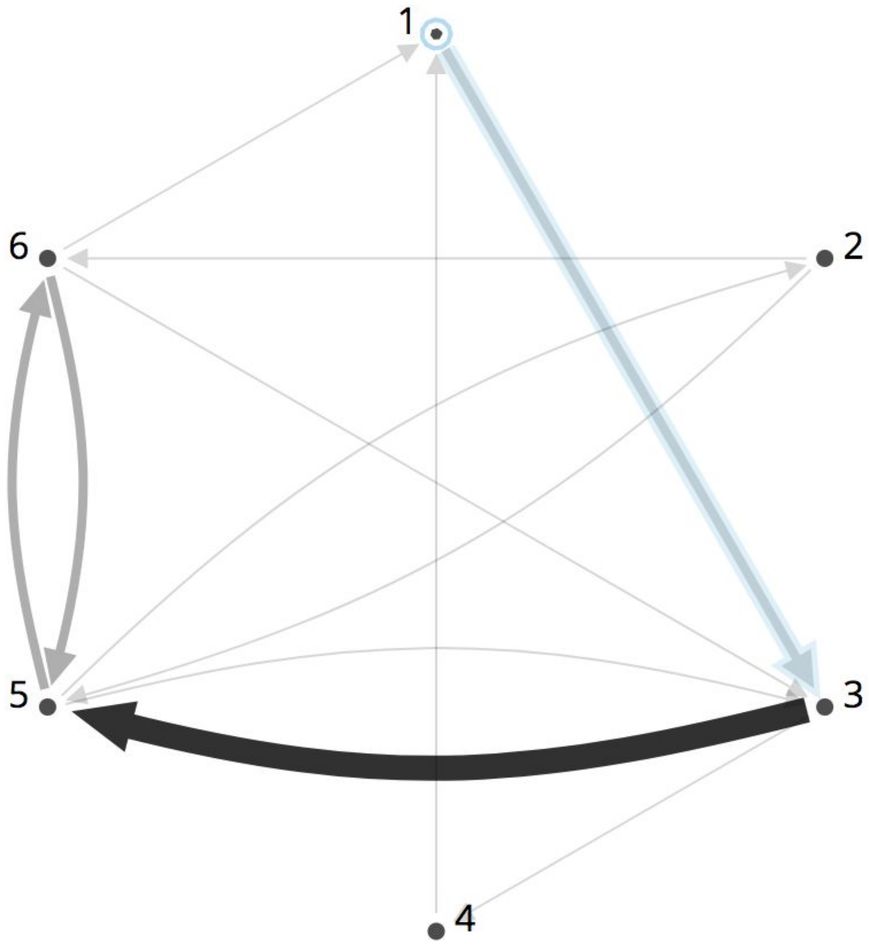
Who pays whom?

Payments from 5 to 6 and 6 to 3



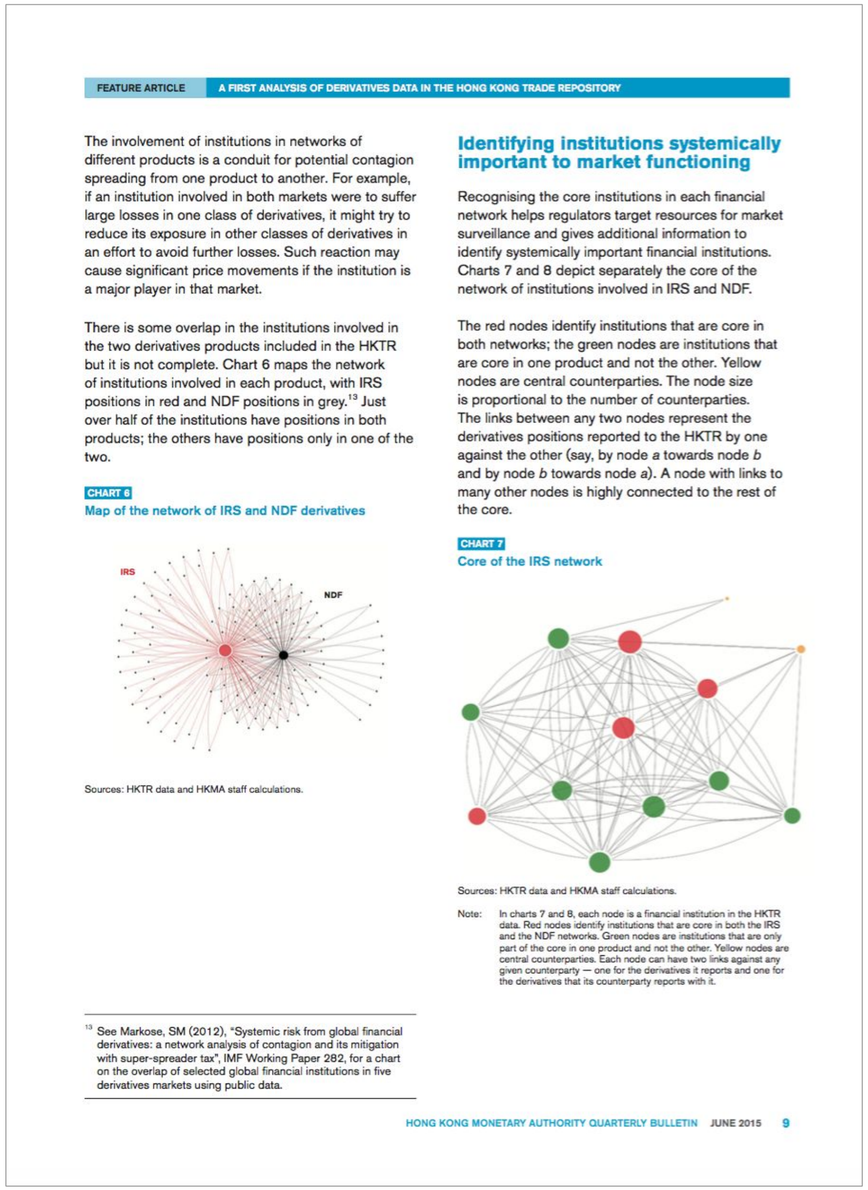
Who pays whom?

More payments ....



Thicker, darker links represent higher link weights, i.e., more payments

Use Case: Understanding Interconnectedness



FEATURE ARTICLE A FIRST ANALYSIS OF DERIVATIVES DATA IN THE HONG KONG TRADE REPOSITORY

The involvement of institutions in networks of different products is a conduit for potential contagion spreading from one product to another. For example, if an institution involved in both markets were to suffer large losses in one class of derivatives, it might try to reduce its exposure in other classes of derivatives in an effort to avoid further losses. Such reaction may cause significant price movements if the institution is a major player in that market.

There is some overlap in the institutions involved in the two derivatives products included in the HKTR but it is not complete. Chart 6 maps the network of institutions involved in each product, with IRS positions in red and NDF positions in grey.[13] Just over half of the institutions have positions in both products; the others have positions only in one of the two.

CHART 6
Map of the network of IRS and NDF derivatives

Sources: HKTR data and HKMA staff calculations.

Identifying institutions systemically important to market functioning

Recognising the core institutions in each financial network helps regulators target resources for market surveillance and gives additional information to identify systemically important financial institutions. Charts 7 and 8 depict separately the core of the network of institutions involved in IRS and NDF.

The red nodes identify institutions that are core in both networks; the green nodes are institutions that are core in one product and not the other. Yellow nodes are central counterparties. The node size is proportional to the number of counterparties. The links between any two nodes represent the derivatives positions reported to the HKTR by one against the other (say, by node *a* towards node *b* and by node *b* towards node *a*). A node with links to many other nodes is highly connected to the rest of the core.

CHART 7
Core of the IRS network

Sources: HKTR data and HKMA staff calculations.

Note: In charts 7 and 8, each node is a financial institution in the HKTR data. Red nodes identify institutions that are core in both the IRS and the NDF networks. Green nodes are institutions that are only part of the core in one product and not the other. Yellow nodes are central counterparties. Each node can have two links against any given counterparty — one for the derivatives it reports and one for the derivatives that its counterparty reports with it.

[13] See Markose, SM (2012), "Systemic risk from global financial derivatives: a network analysis of contagion and its mitigation with super-spreader tax", IMF Working Paper 282, for a chart on the overlap of selected global financial institutions in five derivatives markets using public data.

HONG KONG MONETARY AUTHORITY QUARTERLY BULLETIN JUNE 2015 9

Background

As part of global regulatory reforms, the Hong Kong Monetary Authority (HKMA) started in 2013 to collect derivatives data through the Hong Kong Trade Repository.

Objective

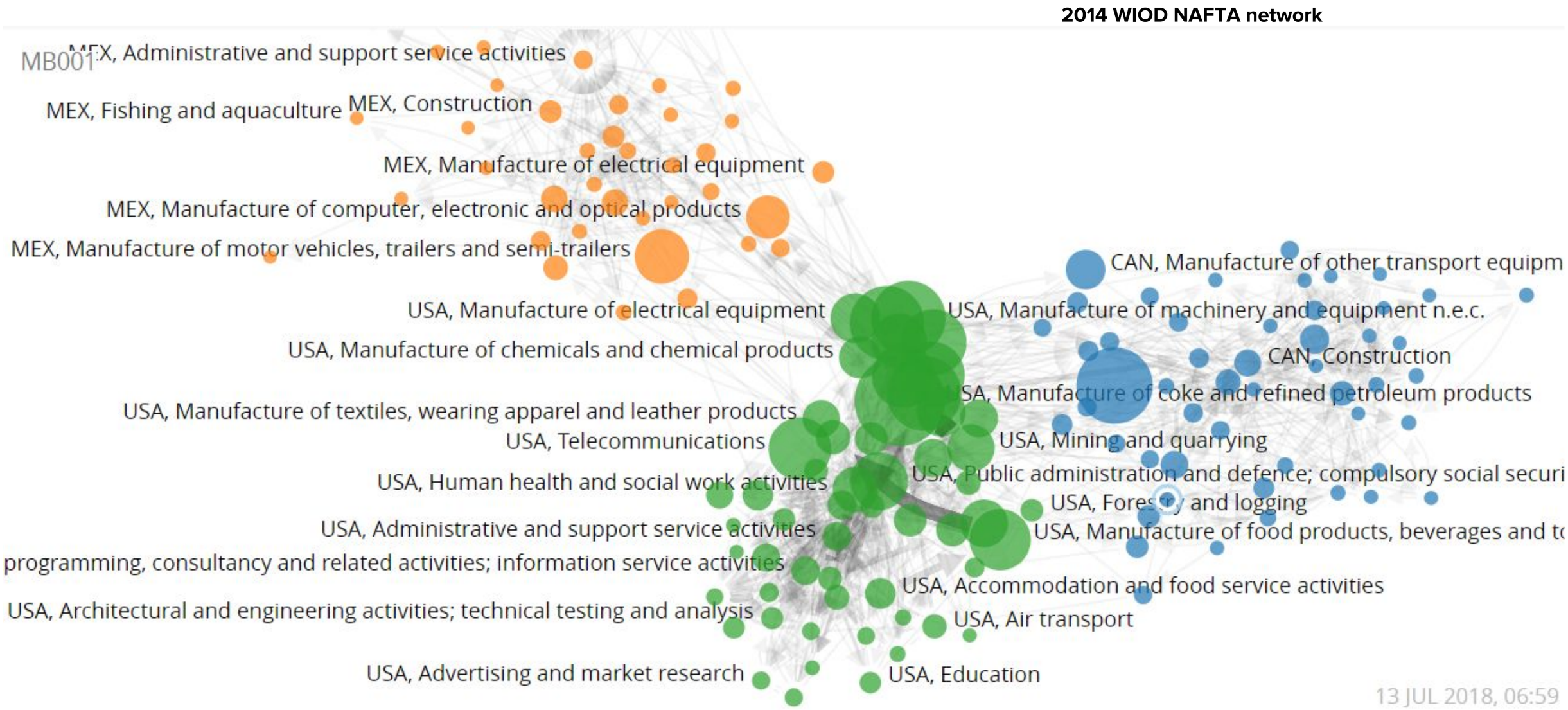
Bring more transparency to derivatives markets using the data collected by trade repositories.

Insights

Initial framework for analysing this new data source to assess the financial stability of the market and potential risks. This includes development of maps for the chain of exposures between institutions.

HKMA: A first analysis of derivatives data in the Hong Kong Trade Repository

World Trade Network



FNA

Interconnectedness in the Global System of CCPs

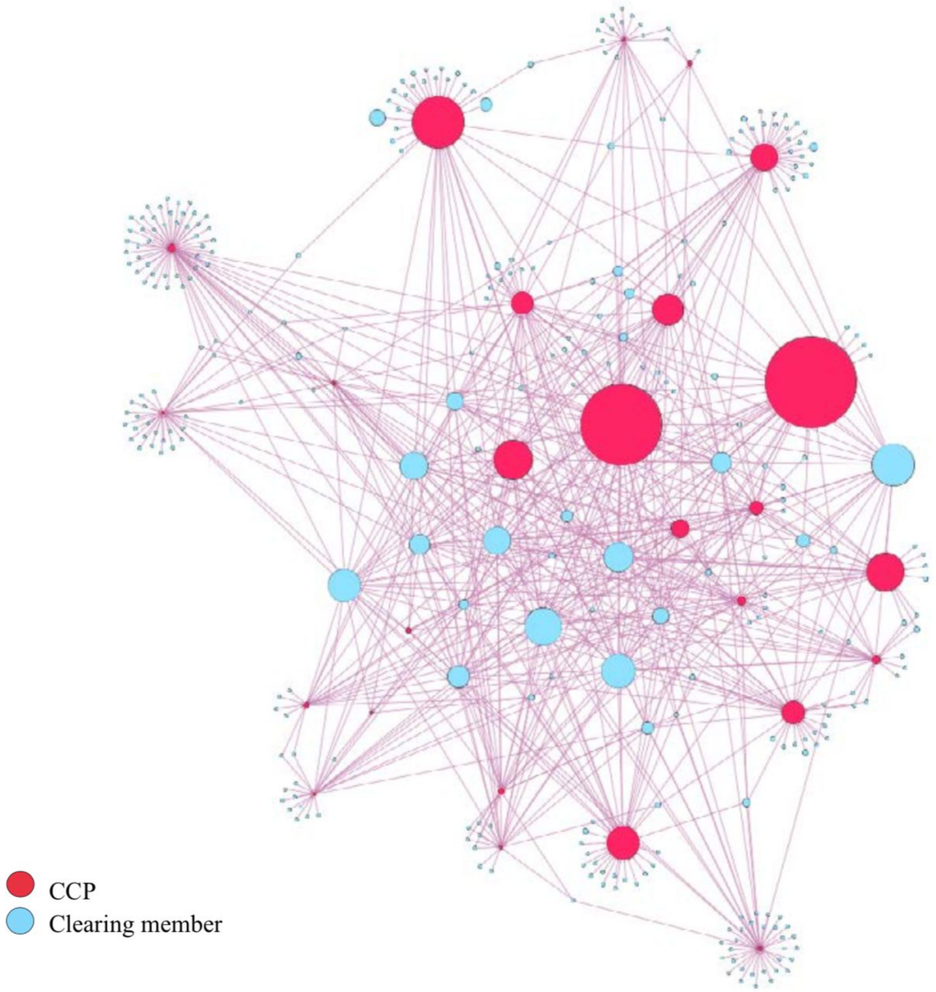
www.fna.fi

Scope of Analysis

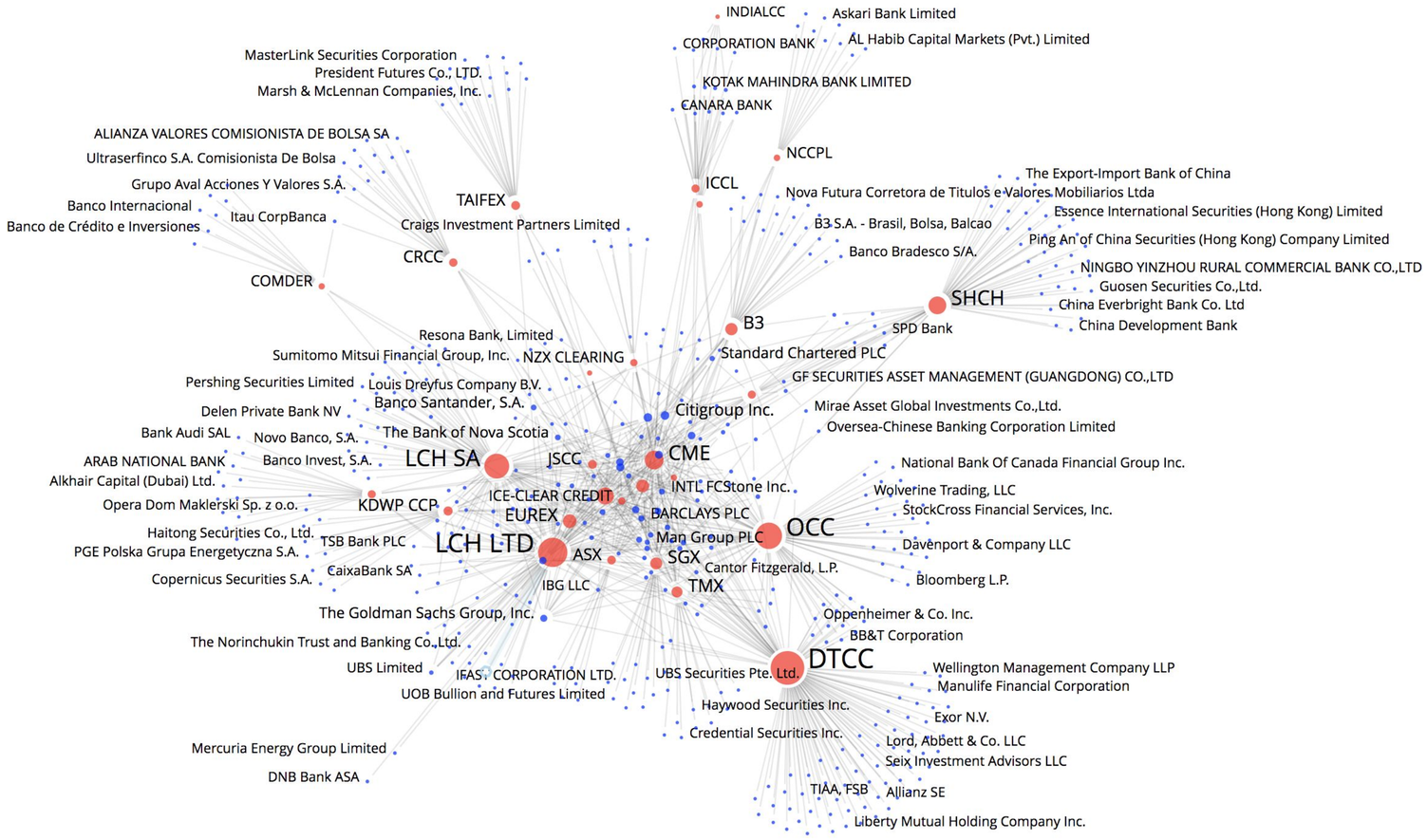
Comparison with BIS "Analysis of Central Clearing Interdependencies" (2017)

	BIS (2017)	**FNA (2018)**
CCPs	26	29
Jurisdictions	20	25
Clearing Members	n/a	811
Parents Organizations	307	563
Roles	5 (member, settlement, LOC, ...)	1 (member)

Private vs Public Data



BIS (2017)



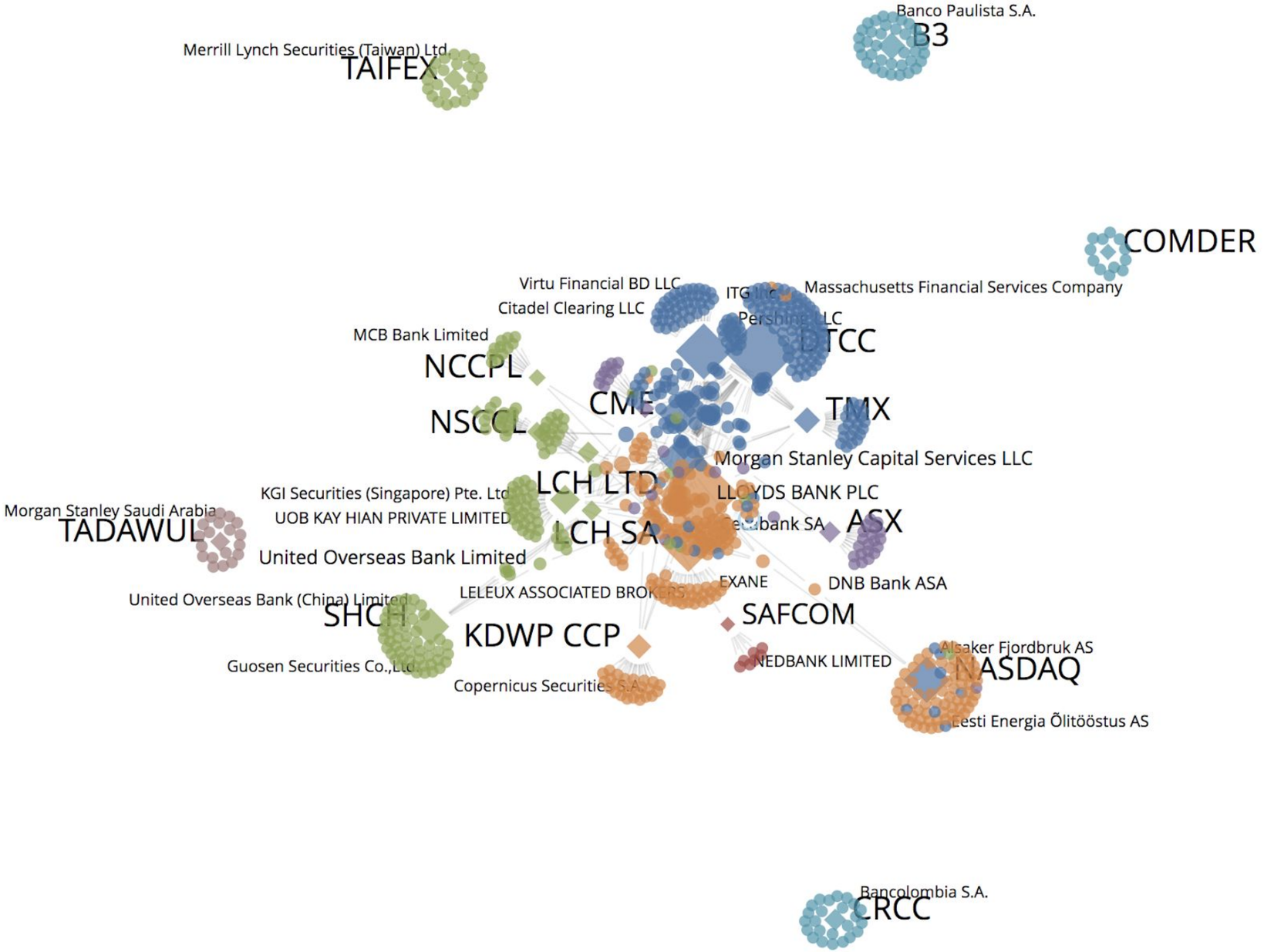
FNA (2018)

CCP Interconnectedness - Subsidiary Level

We see CCPs (diamonds) and their members (circles) from different regions:

- North America (blue)
- Europe (Yellow)
- Asia (green)
- Middle East (brown)
- Latin America (blue)
- Australia & Oceania (purple)

On subsidiary level, we see a tight core with peripheral CCPs and a number of completely disconnected CCPs from Latin America and Middle East.



Banking Groups

210 Banking Groups

Largest (# of entities):

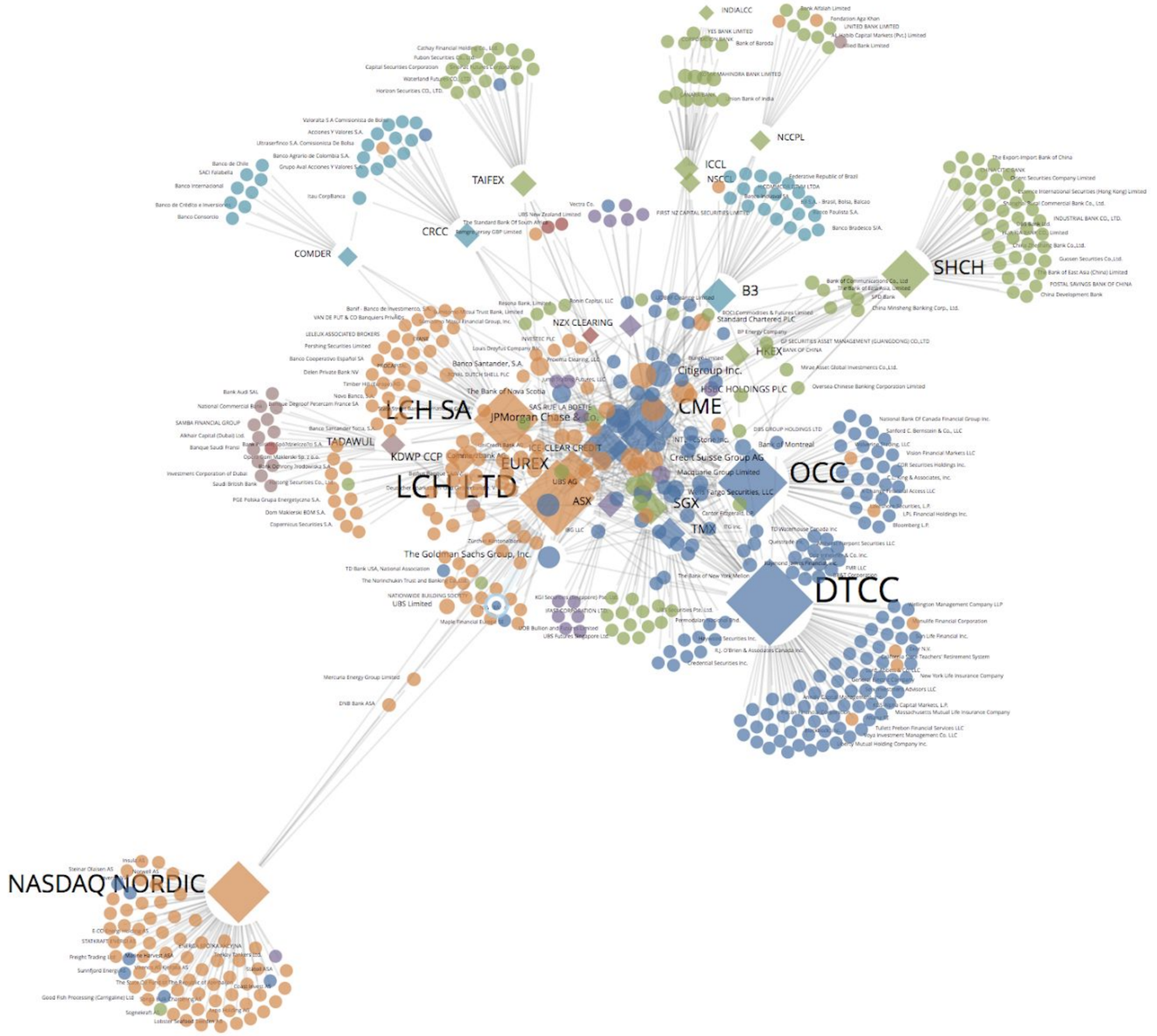
1. Citigroup (19)
2. Morgan Stanley (13)
3. Goldman Sachs (12)
4. JPMorgan Chase (12)
5. Bank of America (12)
6. HSBC (11)
7. Credit Suisse (10)
8. Deutsche Bank (10)
9. Nomura (9)
10. Banco Santander (8)

CCP Interconnectedness on Parent Level

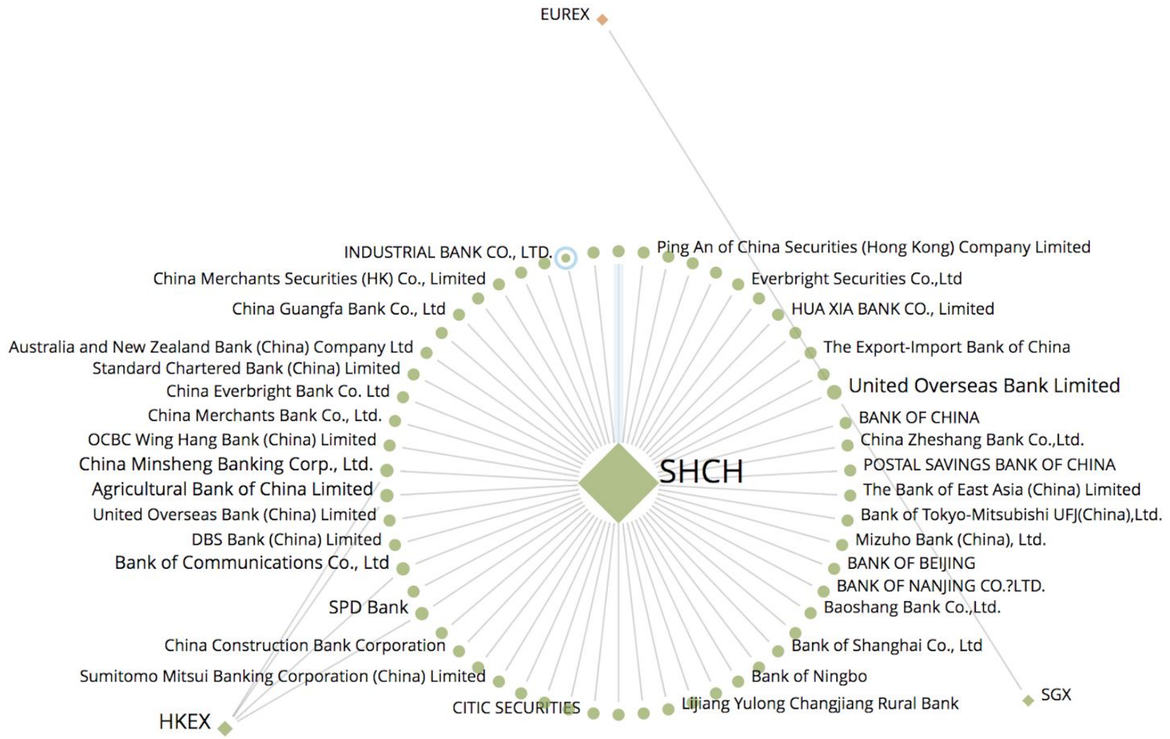
We see CCPs (diamonds) and their members (circles) from different regions:

- North America (blue)
- Europe (Yellow)
- Asia (green)
- Middle East (brown)
- Latin America (blue)
- Australia & Oceania (purple)

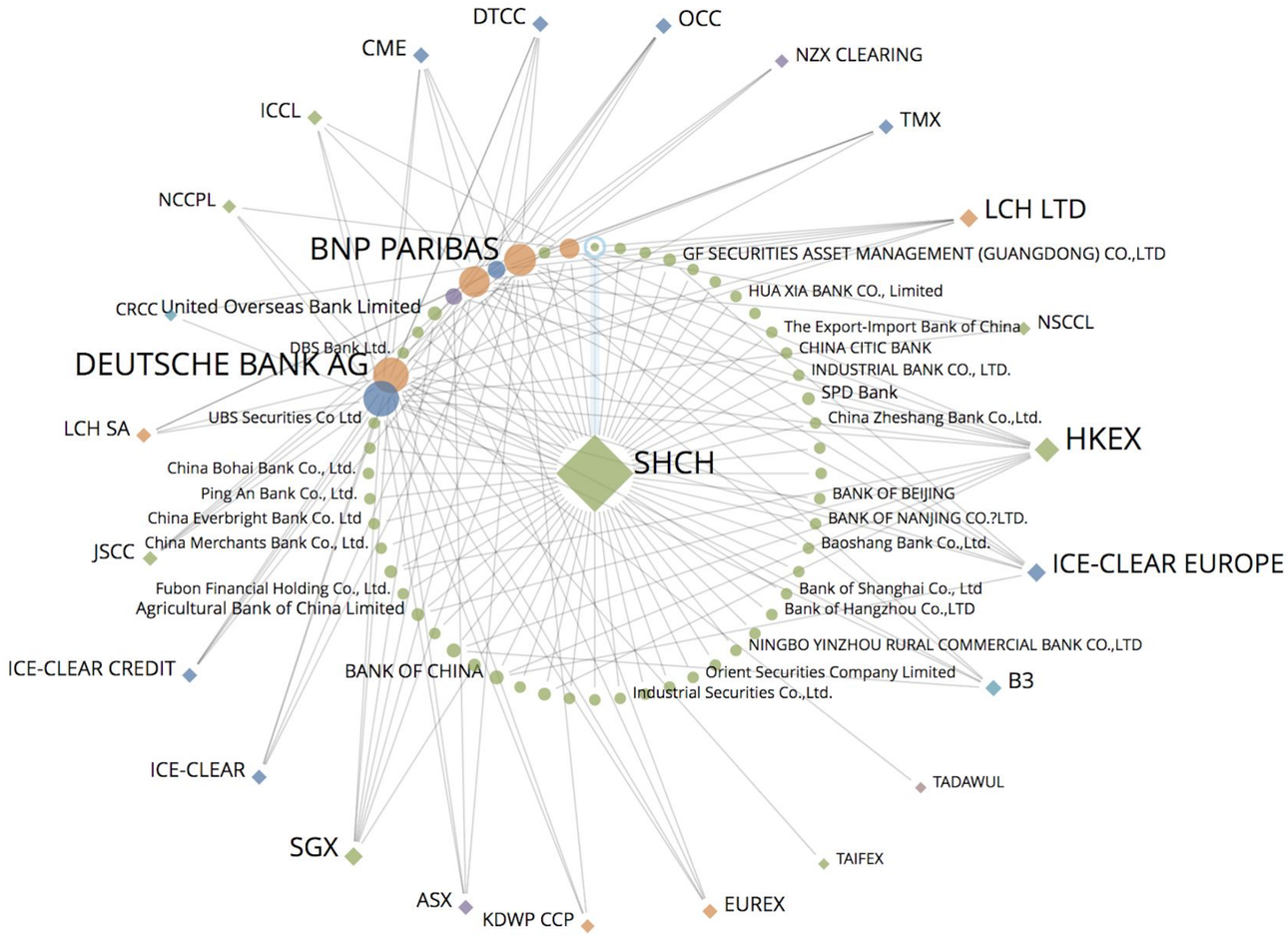
On parent level we see a completely connected network dominated by a core consisting of CCPs from North America and Europe and global banks.



CCP Interconnectedness on Subsidiary vs Parent Level - Example



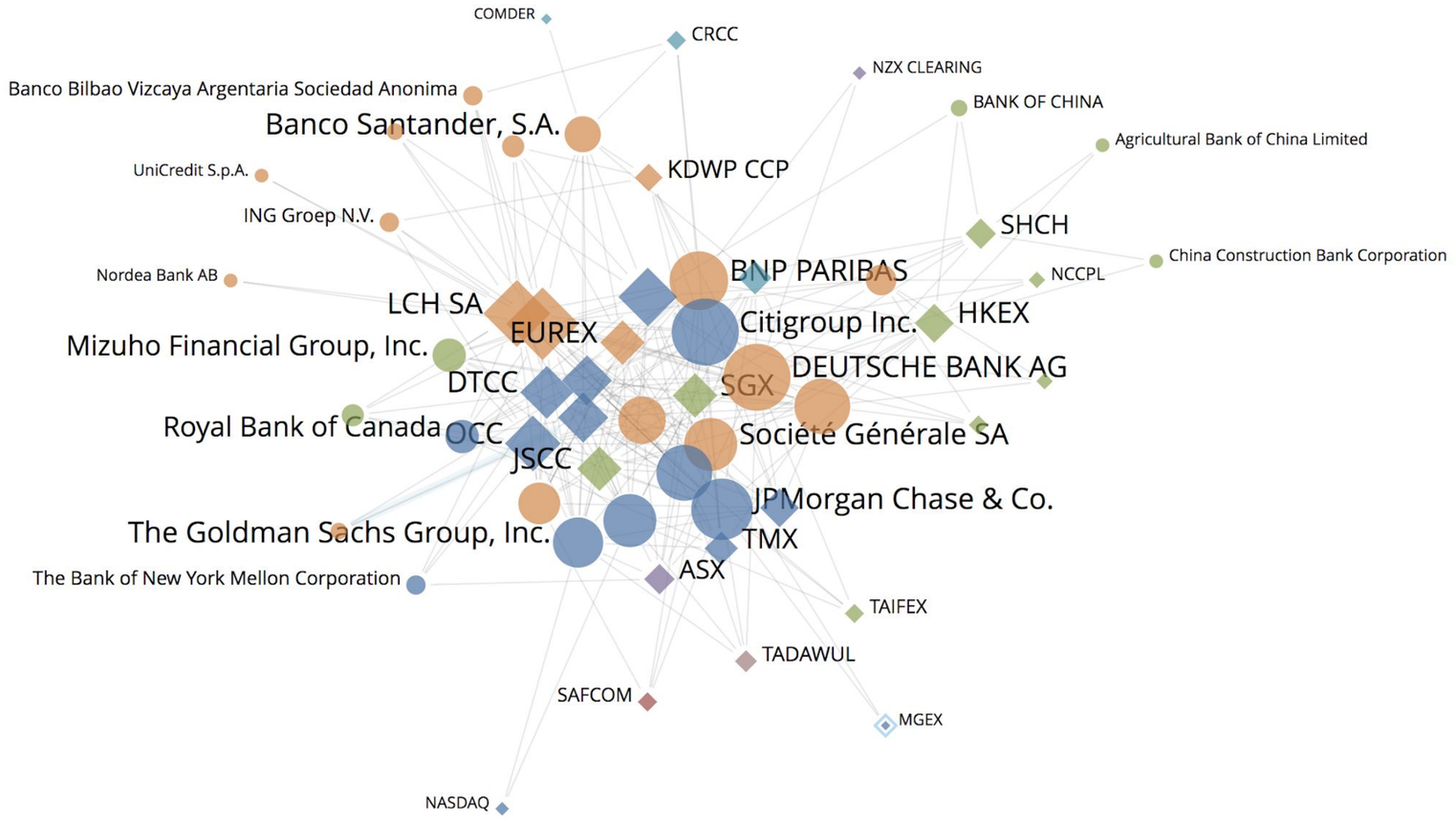
Subsidiary Level
(Connected to 3 CCPs)



Parent Level
(Connected to 23 CCPs)

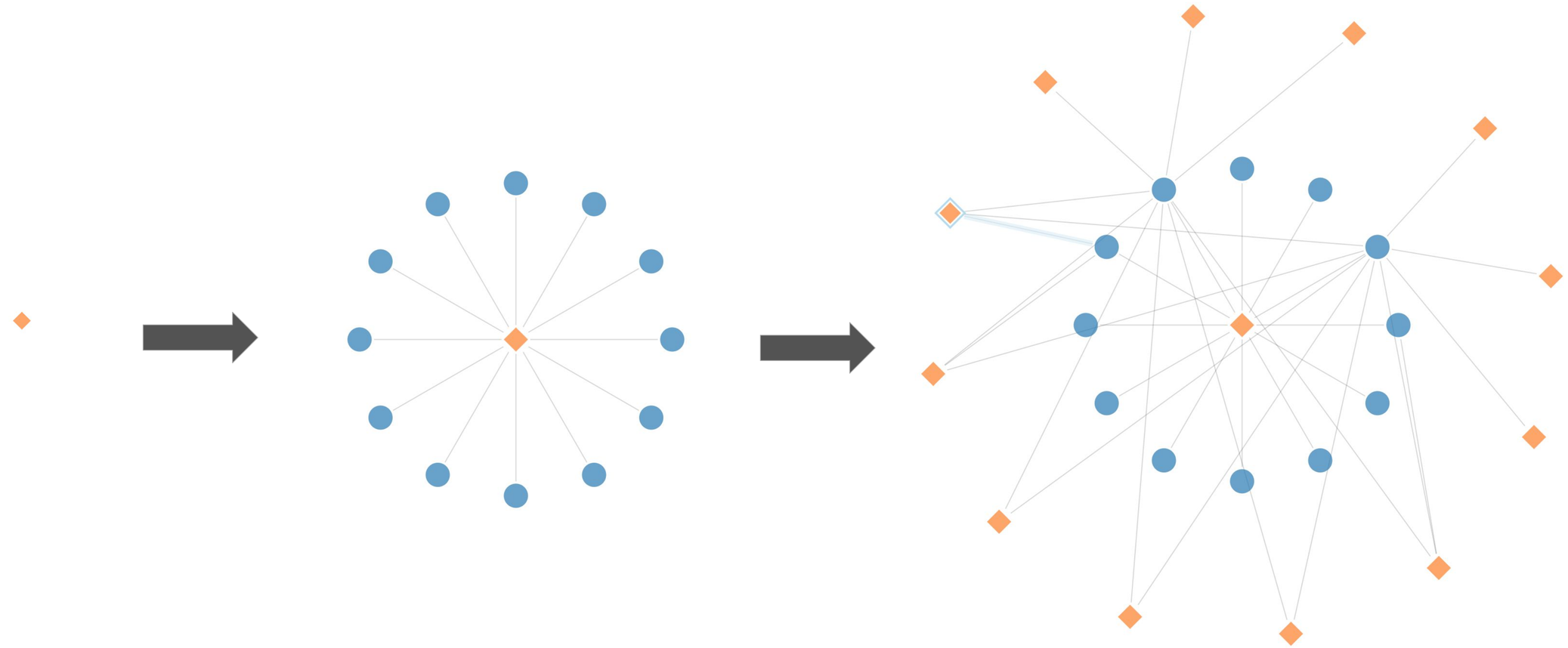
CCP Interconnectedness on GSIB Level

Bank (Parent)	# of FMIs
Citigroup	21
DEUTSCHE BANK	21
JPMorgan Chase & Co.	19
BNP PARIBAS	18
Bank of America	17
HSBC	17
Morgan Stanley	16
Societe Generale	16
The Goldman Sachs	15
Credit Suisse	14



Contagion - CCP Disruption

A disruption in a CCP would affect all of that CCP's clearing members, thereby affecting the other CCP's to which the affected CCP's members belong, possibly creating a cascading cycle as disruption is propagated across members and CCPs



CCP disruption

Propagates to all members

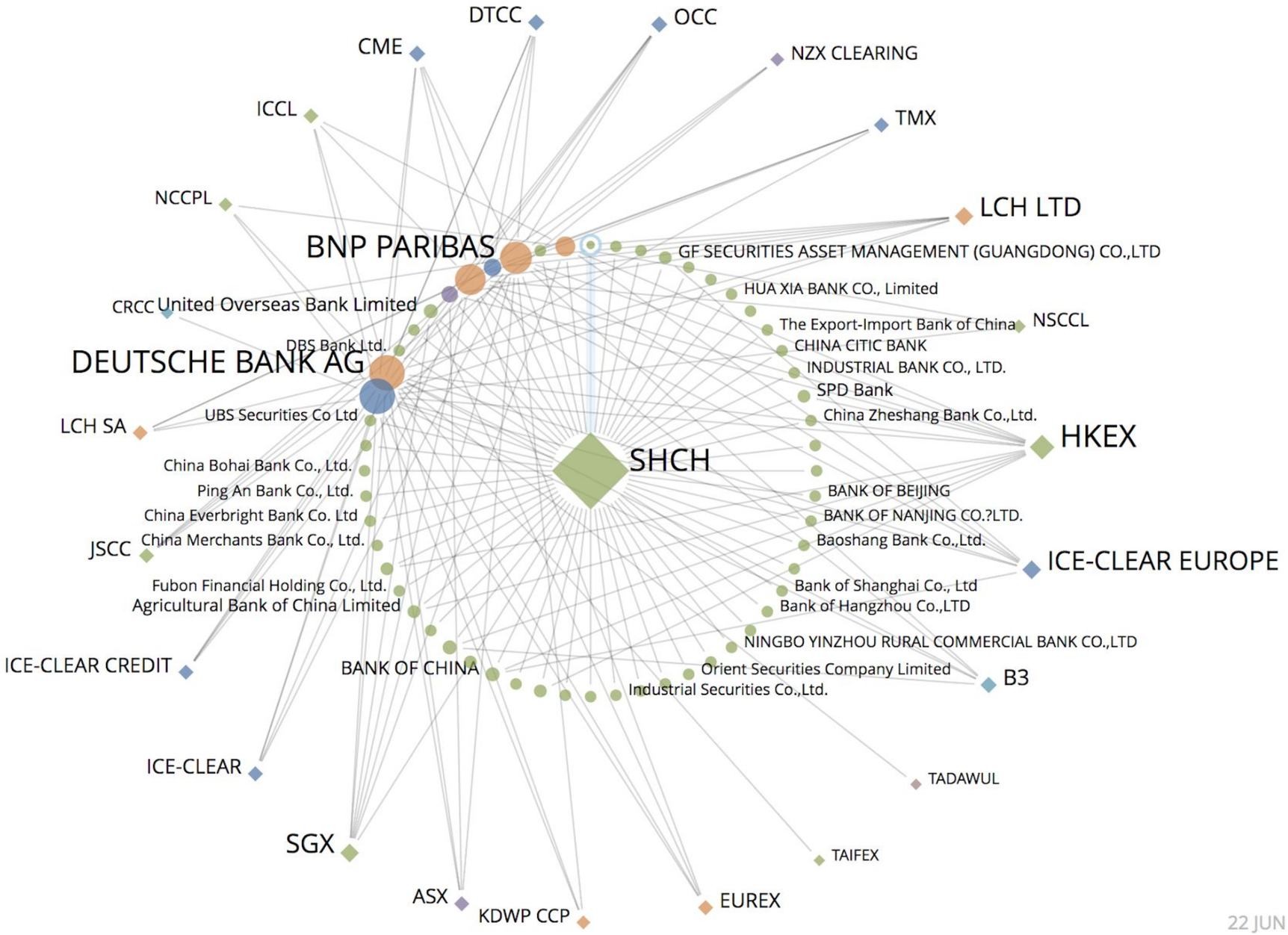
Propagates to other CCPs

Footprint of CCPs - SHCL

SHCL's 56 members are connected to 23 other CCPs

Most members are domestic with a few large global banks based in EU & US.

The most connected CCP is HKEx.

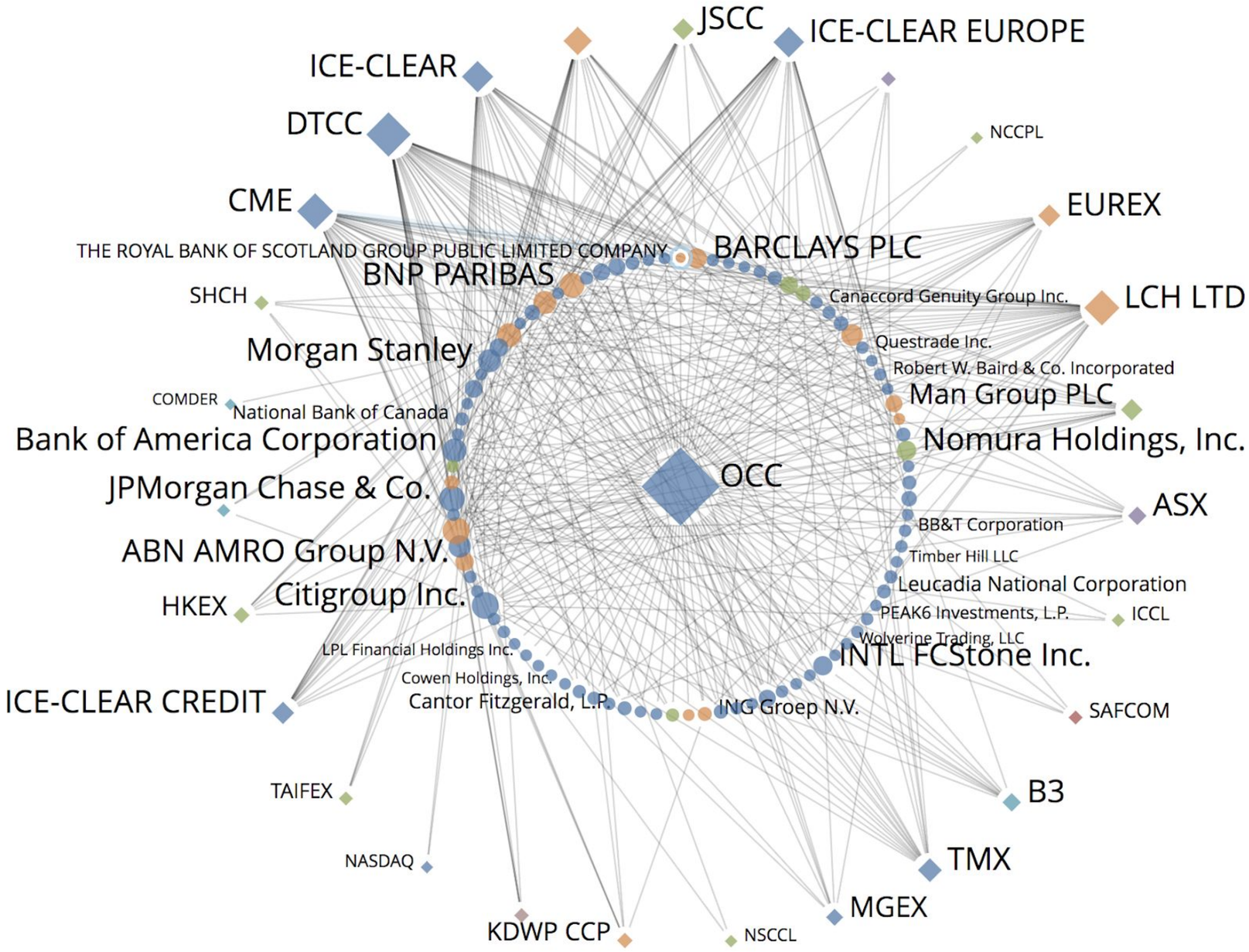


Footprint of CCPs - OCC

OCC's 89 members are connected to 27 other CCPs

The membership is mostly US with a significant EU base.

The most connected CCP's are DTCC and CME.

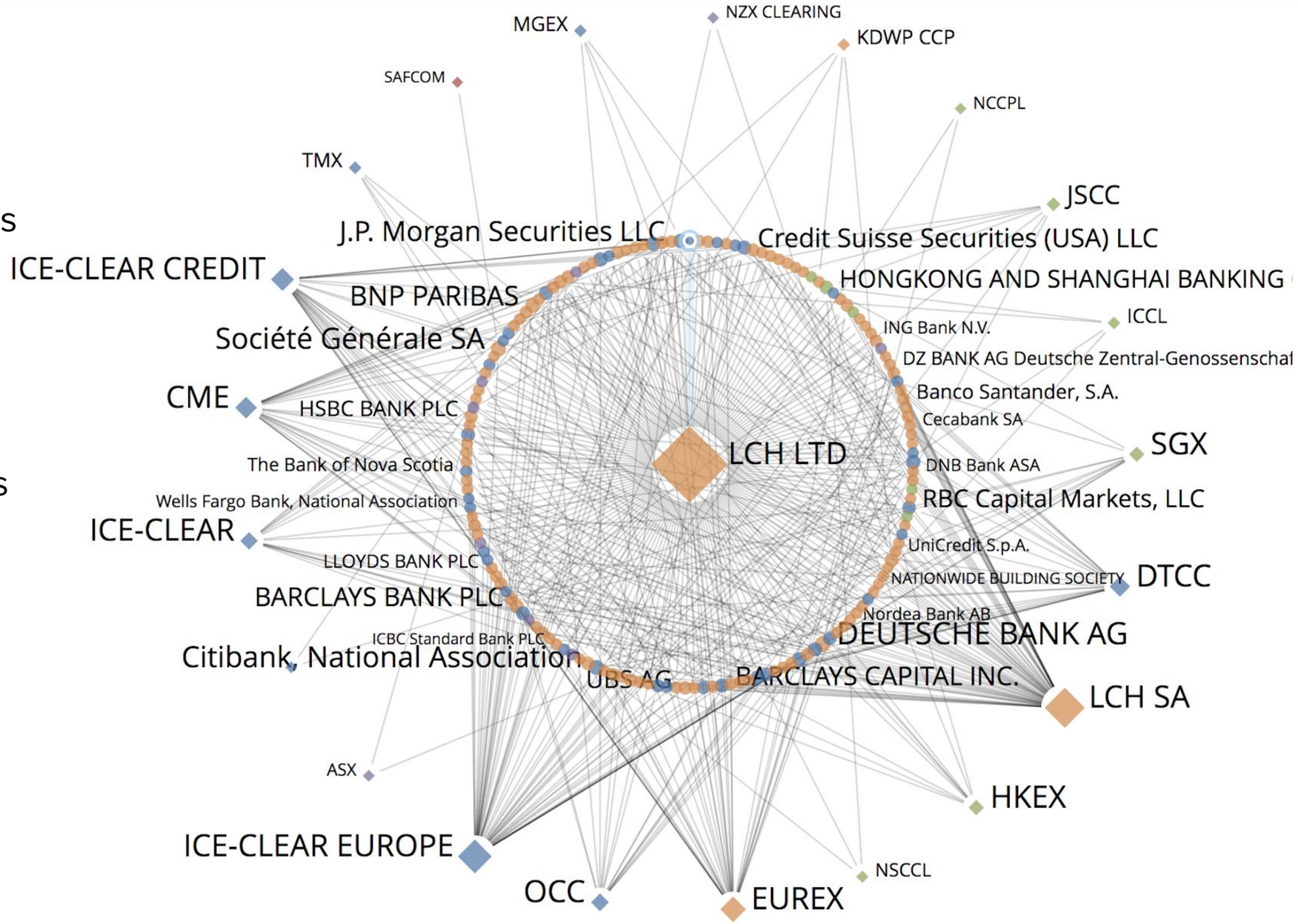


Footprint of CCPs - LCH Ltd

LCH Ltd 100 members are connected to 27 other CCPs

The membership is mostly European with a significant US base.

The most connected CCP is LCH SA and ICE-CLEAR EUROPE.



Contagion – Member Disruption

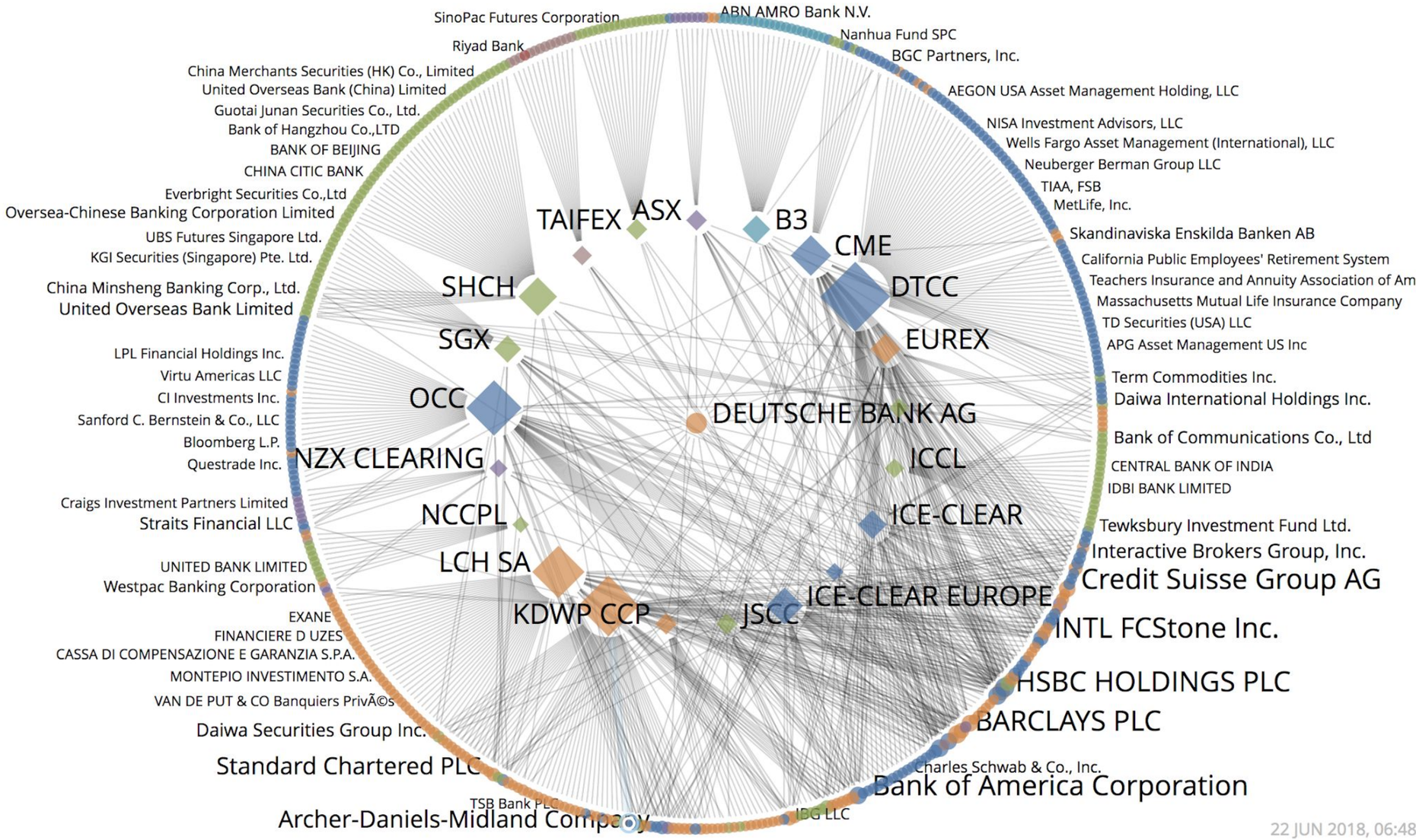
A member disruption can be felt by up to **458** banking groups or banks (of total of 563, or 80%) that are members of the same CCP as the stricken group.

Banking Group	**# banking groups connected via a CCP**
Deutsche Bank	458
Citigroup	446
Morgan Stanley	442
BNP Paribas	423
Goldman Sachs	412
HSBC Holdings	402
JPMorgan Chase	388
Bank of America	382
Credit Suisse	348
Société Générale	340

Contagion – Member Disruption

Deutsche Bank Group participates in 21 CCPs (of 29 mapped).

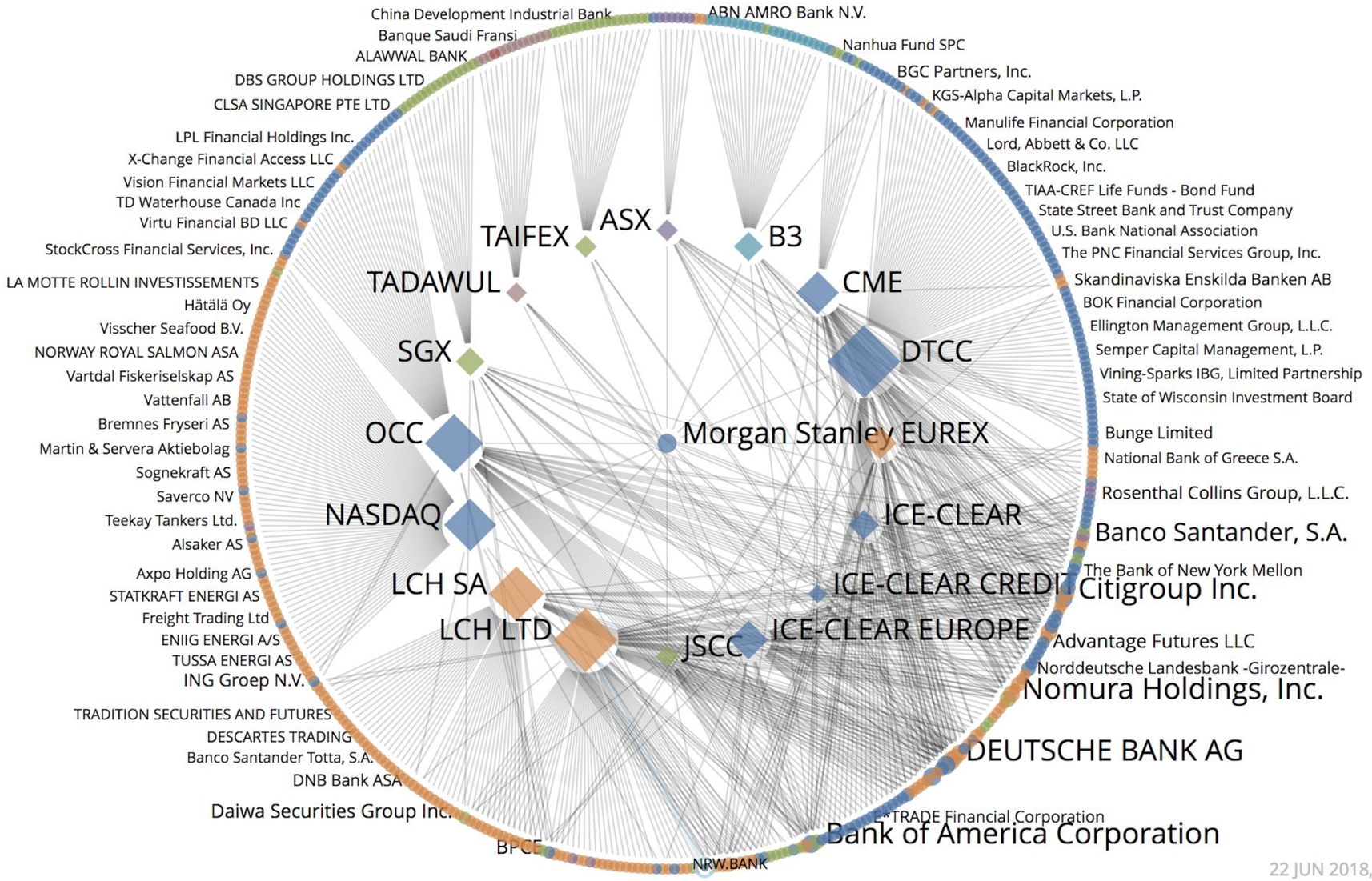
458 other banking groups or banks are members of these CCPs.



Contagion – Member Disruption

Morgan Stanley participates in 16 CCPs (of 29 mapped).

442 other banking groups or banks are members of these CCPs.



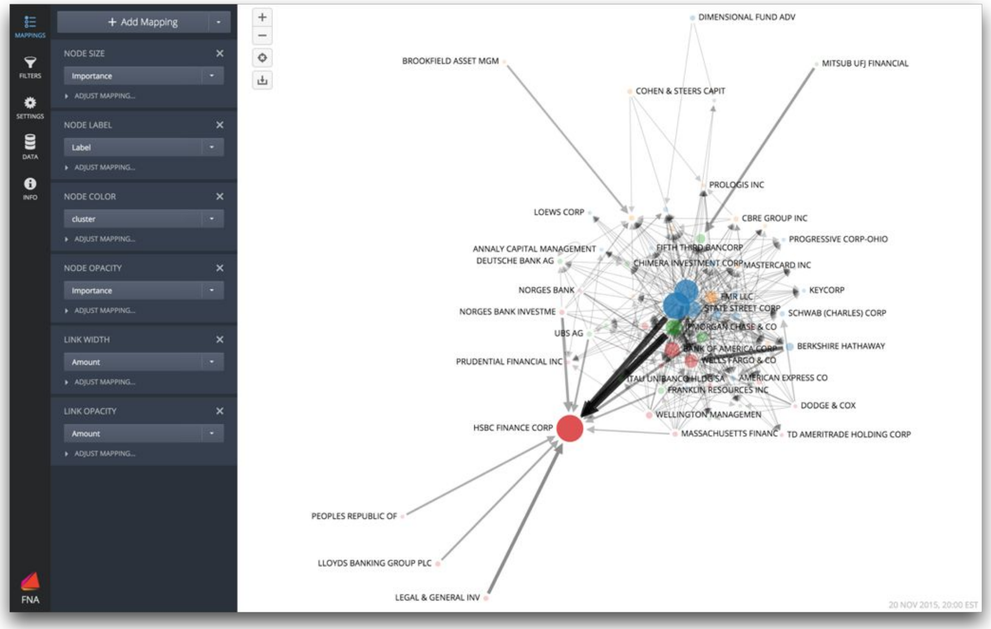
FNA

Top Down Analysis
Correlation Networks

www.fna.fi

Transactions & Similarity Based Networks

Transaction: payment, trade, exposure, supply, flow, ...



Similarity: correlation, partial correlation, granger causality, transfer entropy, ...

Stavroglou et al (2016)
[Causality Networks of Financial Assets]

Typical view of cross asset correlations

Correlation Matrix Over the Last 15 Years (2001-2015)

		Equity					Fixed Income			Alternative Strategies					Alternative Assets		
		Large Cap	Mid Cap	Small Cap	Int'l	Emerging Mkts	Corp.	High Yield	Treas.	Long/ Short	Mkt Neutral	Event Driven	FI Arbitrage	Mgd Futures	Real Estate	Currency	Comm-odities
Equity	Large Cap	1.00															
	Mid Cap	0.93	1.00														
	Small Cap	0.88	0.96	1.00													
	Int'l	0.88	0.85	0.79	1.00												
	Emer. Mkts	0.78	0.80	0.75	0.87	1.00											
Fixed Income	Corp.	0.17	0.20	0.13	0.30	0.31	1.00										
	High Yield	0.66	0.71	0.66	0.69	0.71	0.52	1.00									
	Treas.	-0.36	-0.35	-0.36	-0.27	-0.25	0.63	-0.19	1.00								
Alt. Strategies	Long/Short	0.75	0.79	0.72	0.84	0.80	0.27	0.61	-0.28	1.00							
	Mkt Neutral	0.27	0.29	0.29	0.26	0.24	-0.09	0.37	-0.29	0.25	1.00						
	Event Driven	0.64	0.70	0.64	0.71	0.70	0.23	0.67	-0.34	0.83	0.32	1.00					
	FI Arbitrage	0.42	0.46	0.37	0.48	0.48	0.42	0.63	-0.10	0.51	0.36	0.55	1.00				
	Mgd Futures	-0.13	-0.10	-0.12	0.00	0.01	0.19	-0.12	0.29	0.19	-0.01	0.09	0.00	1.00			
Alt. Assets	Real Estate	0.08	0.12	0.14	0.08	0.05	-0.03	0.06	-0.08	0.05	0.14	0.04	0.03	0.00	1.00		
	Currency	-0.03	-0.02	0.01	0.02	0.01	0.11	-0.07	0.13	0.13	0.02	0.03	0.00	0.61	0.07	1.00	
	Commodities	0.32	0.38	0.33	0.45	0.47	0.11	0.35	-0.17	0.49	0.28	0.49	0.45	0.18	0.10	-0.08	1.00

High (0.9-1.0) | Moderate High (0.7-0.9) | Moderate (0.3-0.7) | Low (0.0-0.3) | Negative (<0.0)

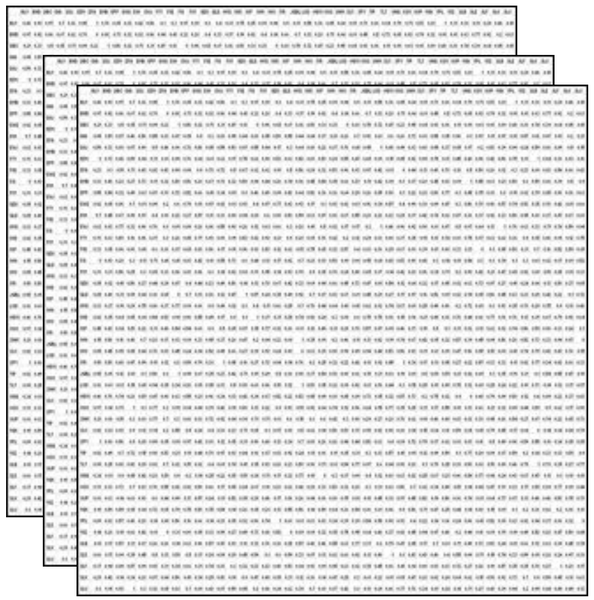
Low Diversification ← → High Diversification

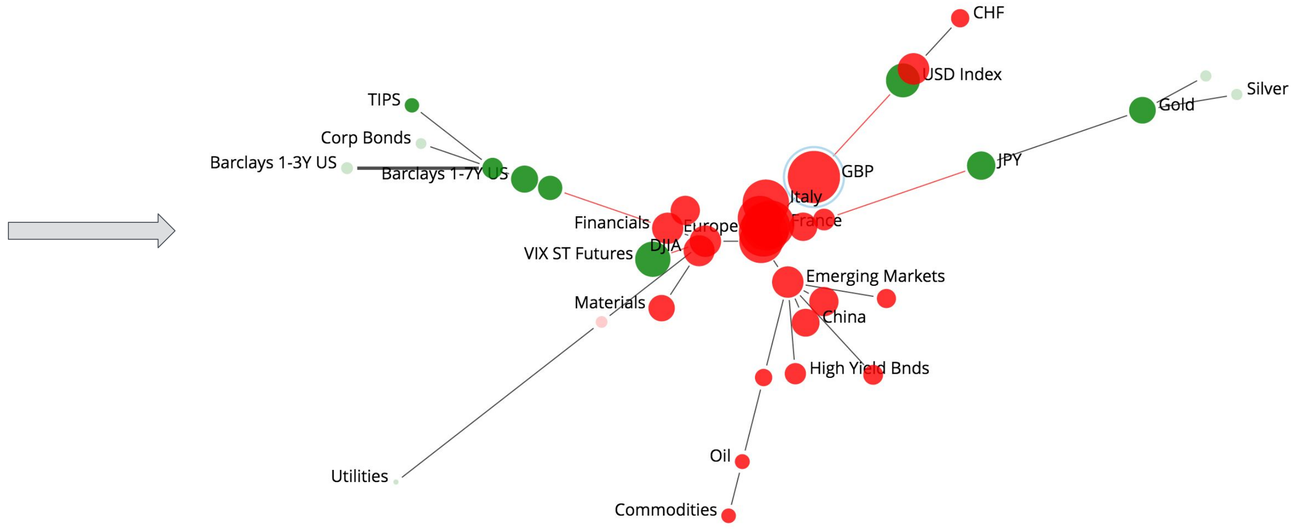
Correlation Networks

Interconnectivity of markets has increased

We need to be able to understand correlations structures of much larger scale.

Networks help develop intuition, and understand stress tests.





Universe of Global Assets (ETFs)

BND	Total Bond Index
DBC	Commodities
DIA	DJIA
DXJ	Japan Stocks (in JPY)
EEM	Emerging Markets
EFA	EAFE
EMB	EMBI
EPP	Asia ex Japan
EWG	Germany
EWI	Italy
EWJ	Japan
EWQ	France
EWU	UK
FXB	GBP
FXC	CAD
FXE	EUR
FXI	China
FXY	JPY
GDX	Gold Miners
GLD	Gold
IEF	Barclays 1-7Y US
IYR	Real Estate
JNK	High Yield Bonds
LQD	Corp Bonds
SLV	Silver
SPY	S&P 500
TIP	TIPS
TLT	20Y+ Gov't
USO	Oil
UUP	USD Index
VGK	Europe
VPL	Asia
VXX	VIX ST Futures
XIU	TSX 60
XLB	Materials
XLE	Energy
XLF	Financials
XLK	Tech
XLU	Utilities
CSJ	Barclays 1-3Y US
FXF	CHF

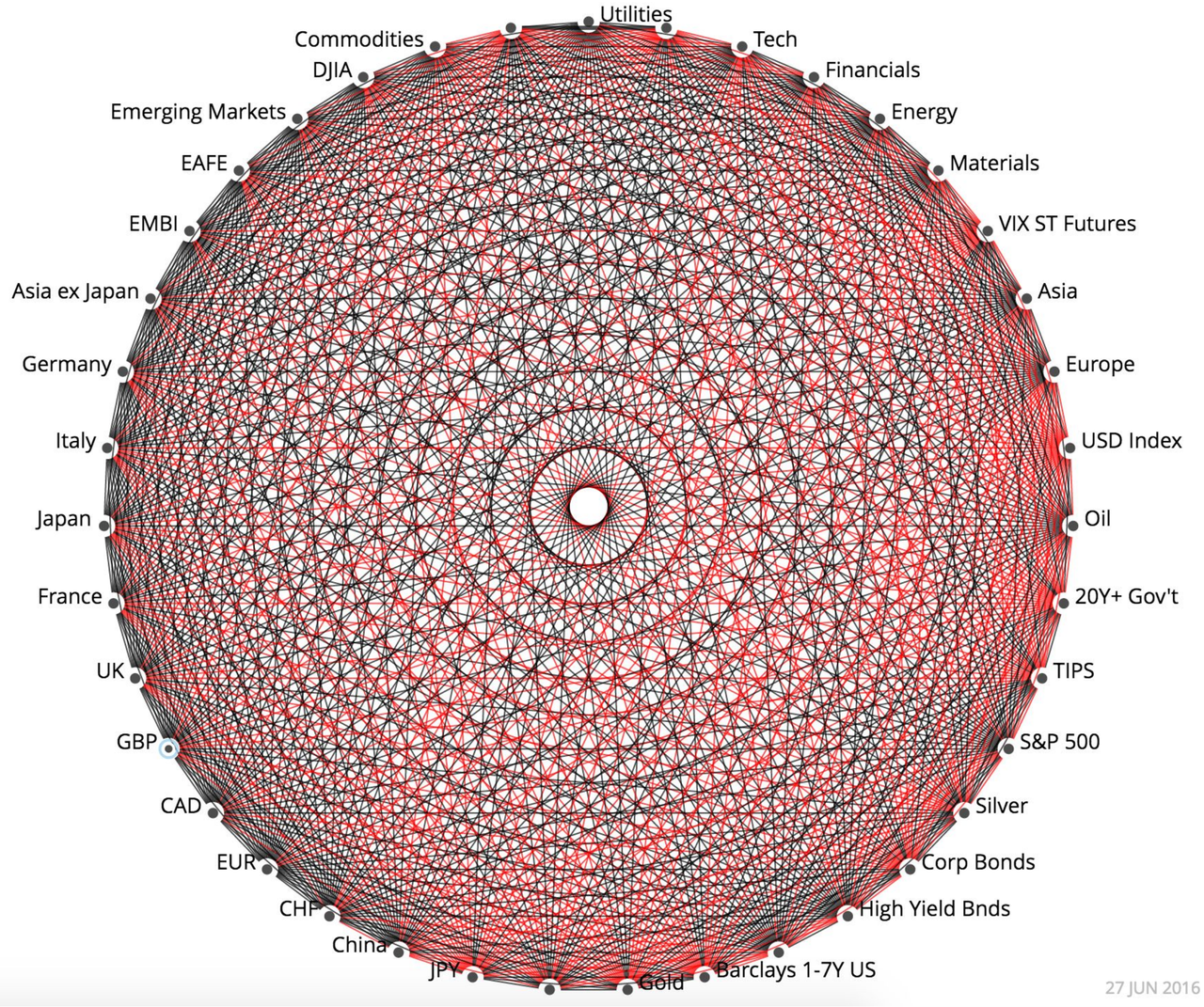
Correlation Network of the Assets

We can view any matrix as a network.

We encode correlations as links between the correlated nodes/assets.

Red link = negative correlation
Black link = positive correlation

However, this simple encoding does not give us much.

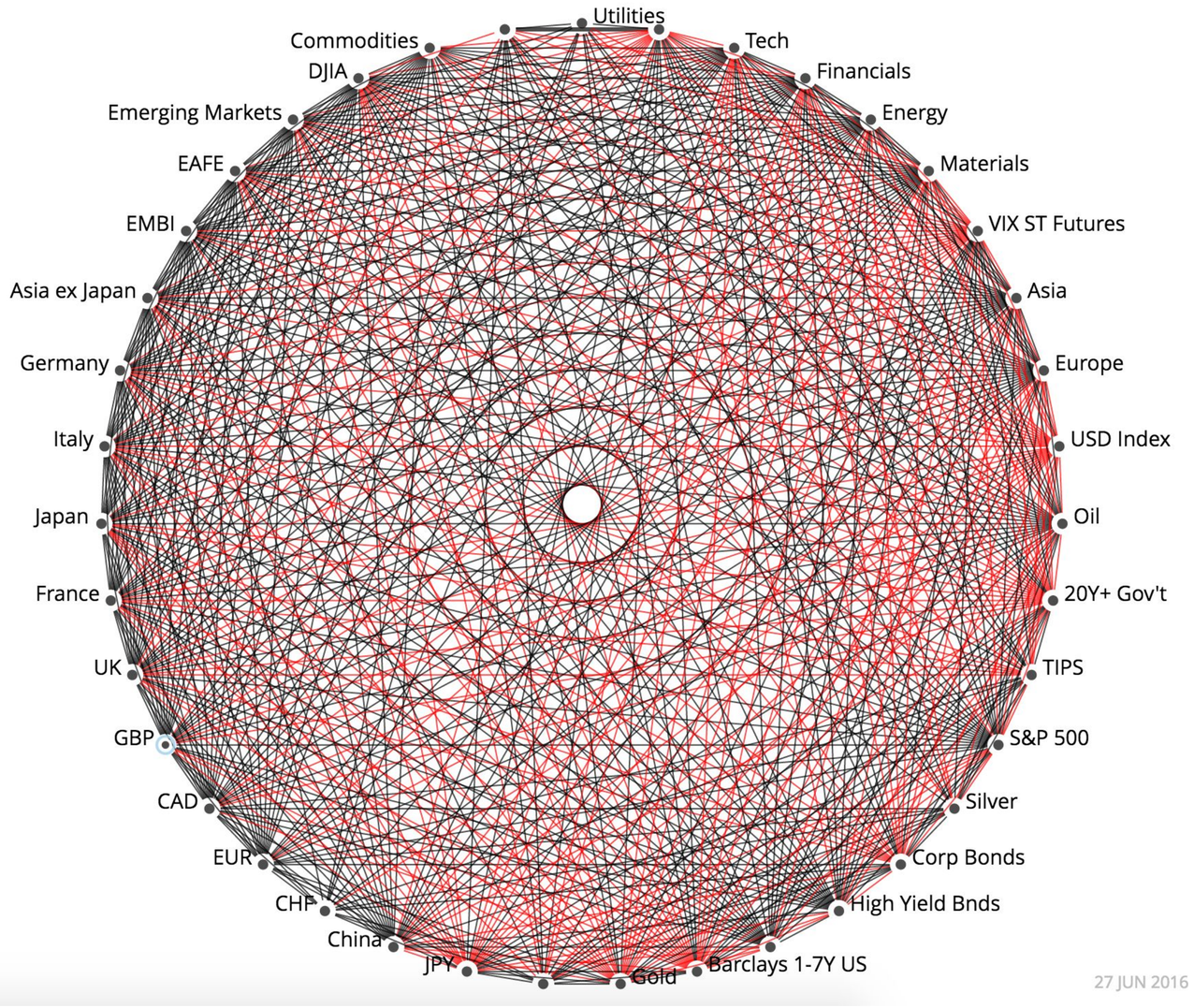


Transactions & Similarity Based Networks

Not all correlations are statistically significantly different from 0.

Absence of link marks that asset is not significantly correlated (here at 95% level).

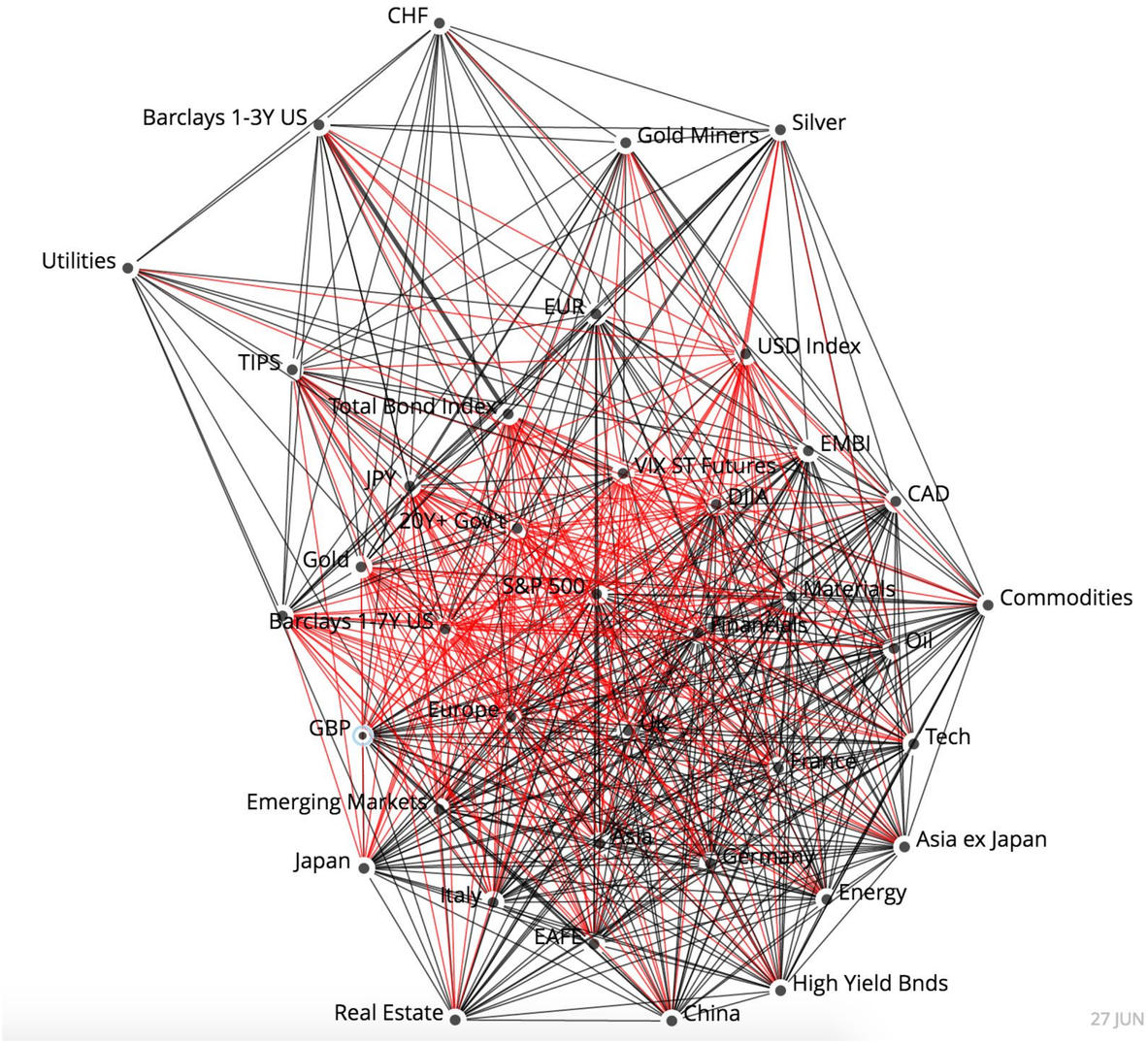
Due to the large number of estimates, we also need for multiple comparisons correction. Eg. Bonferroni or FDR.



Network Layout

We can use network layouts to better detect patterns from noise.

E.g. we can try a Force-Directed network layout to identify clusters.

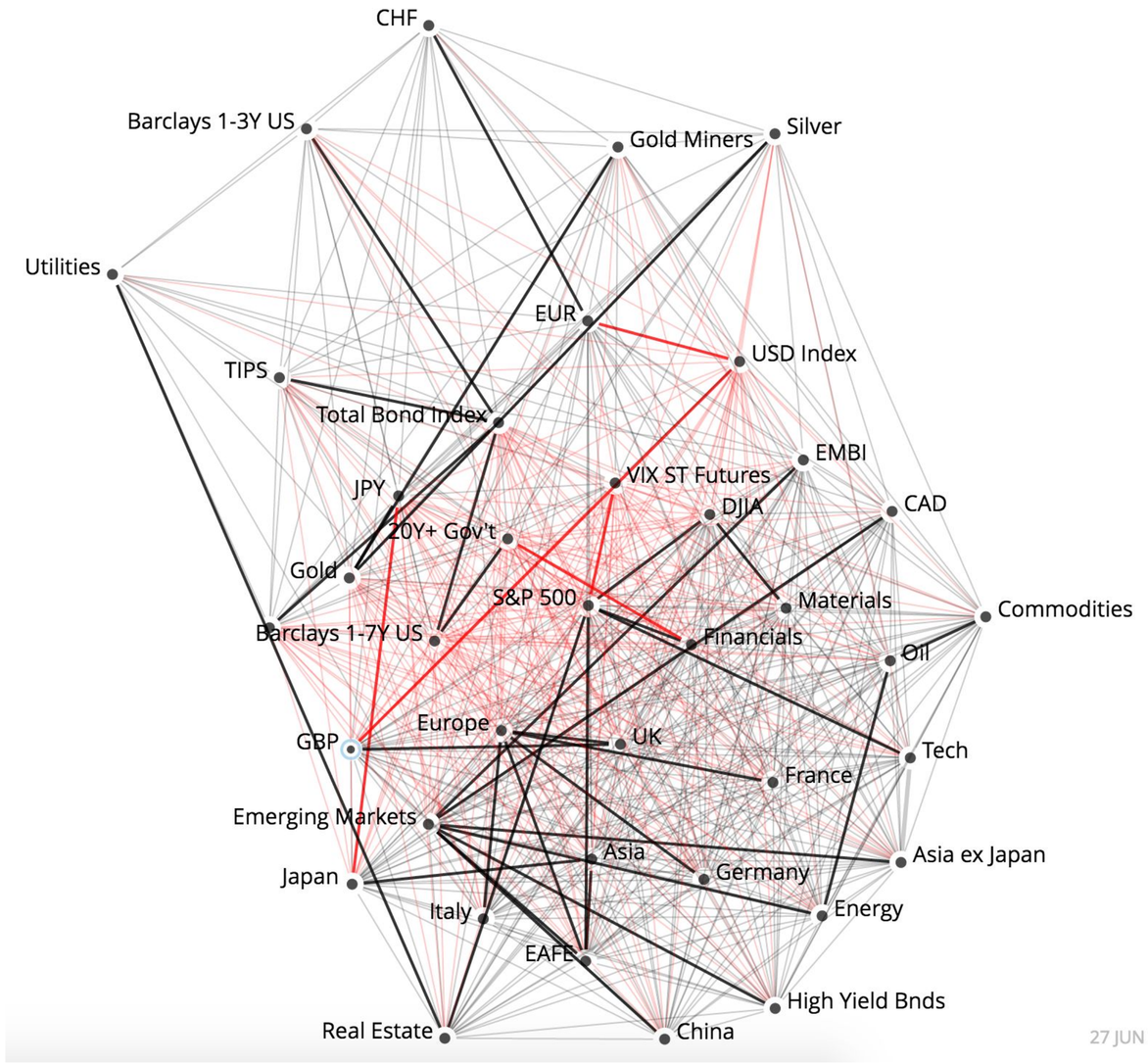


Filtering

Next, we identify the Minimum Spanning Tree and filter out other correlations (Mantegna, ‘99).

We need a distance function, here we look at maximum spanning tree with distance function: abs(cor)

This shows us the backbone correlation structure.

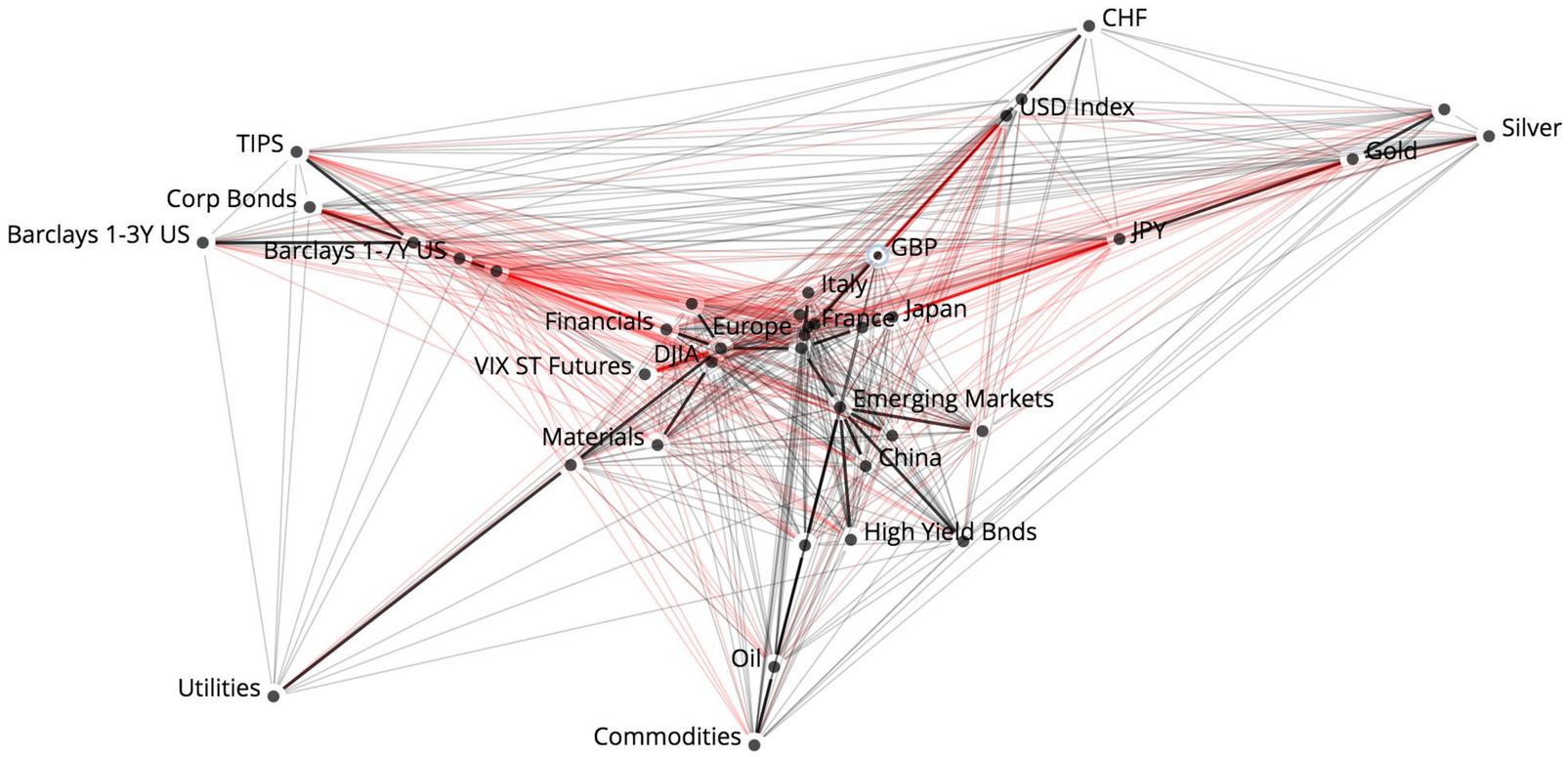


Radial Tree Layout

We use a radial tree layout algorithm (Bachmeier et al. ‘05) that places the assets so that:

- Shorter links in the tree indicate higher correlations
- Longer links indicate lower correlations

As a result, we also see how the assets cluster by asset class.



Radial Tree Layout

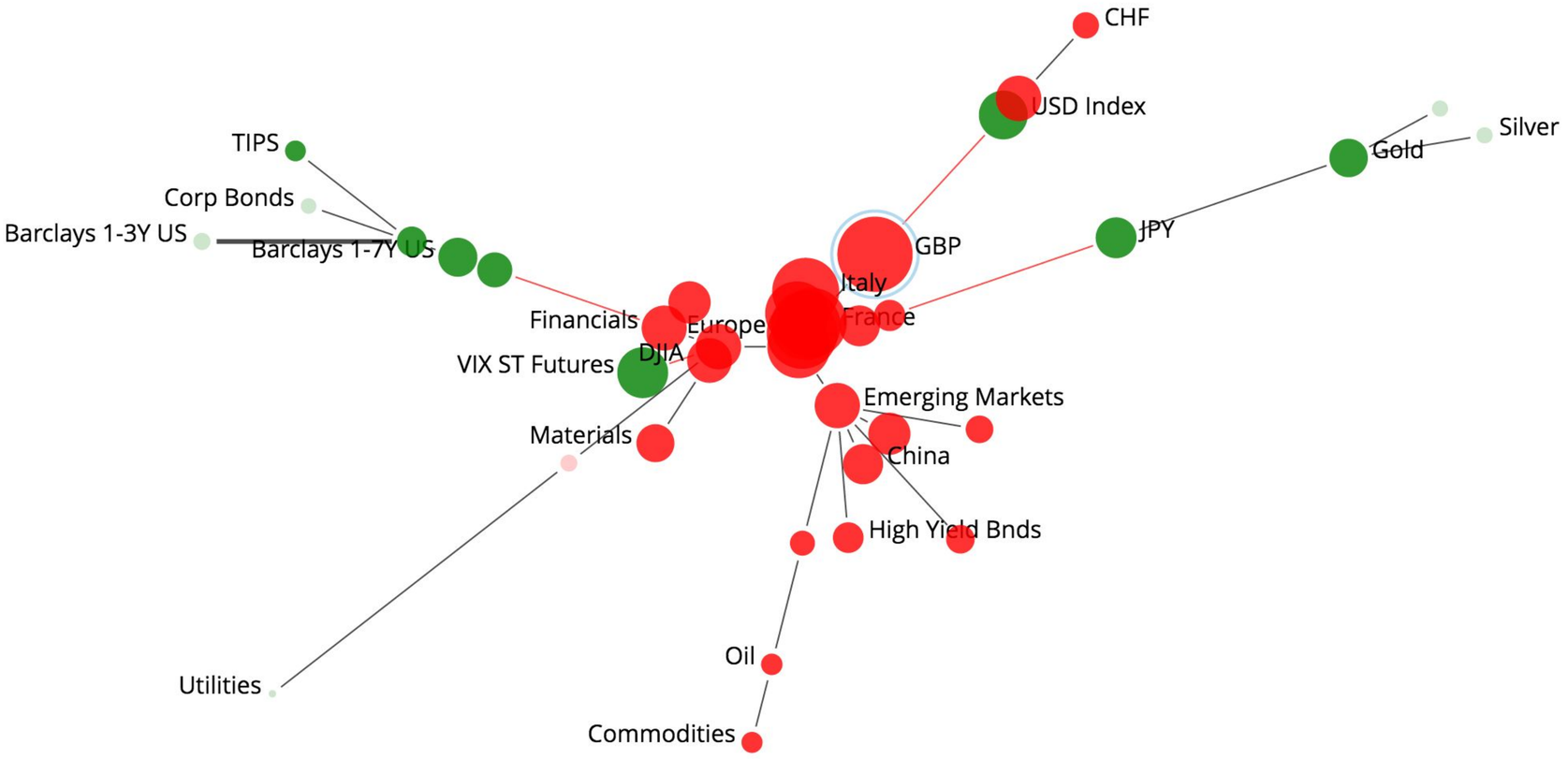
Focus on the links in the Spanning Tree to highlight clustering structure.

Node color indicates last daily return

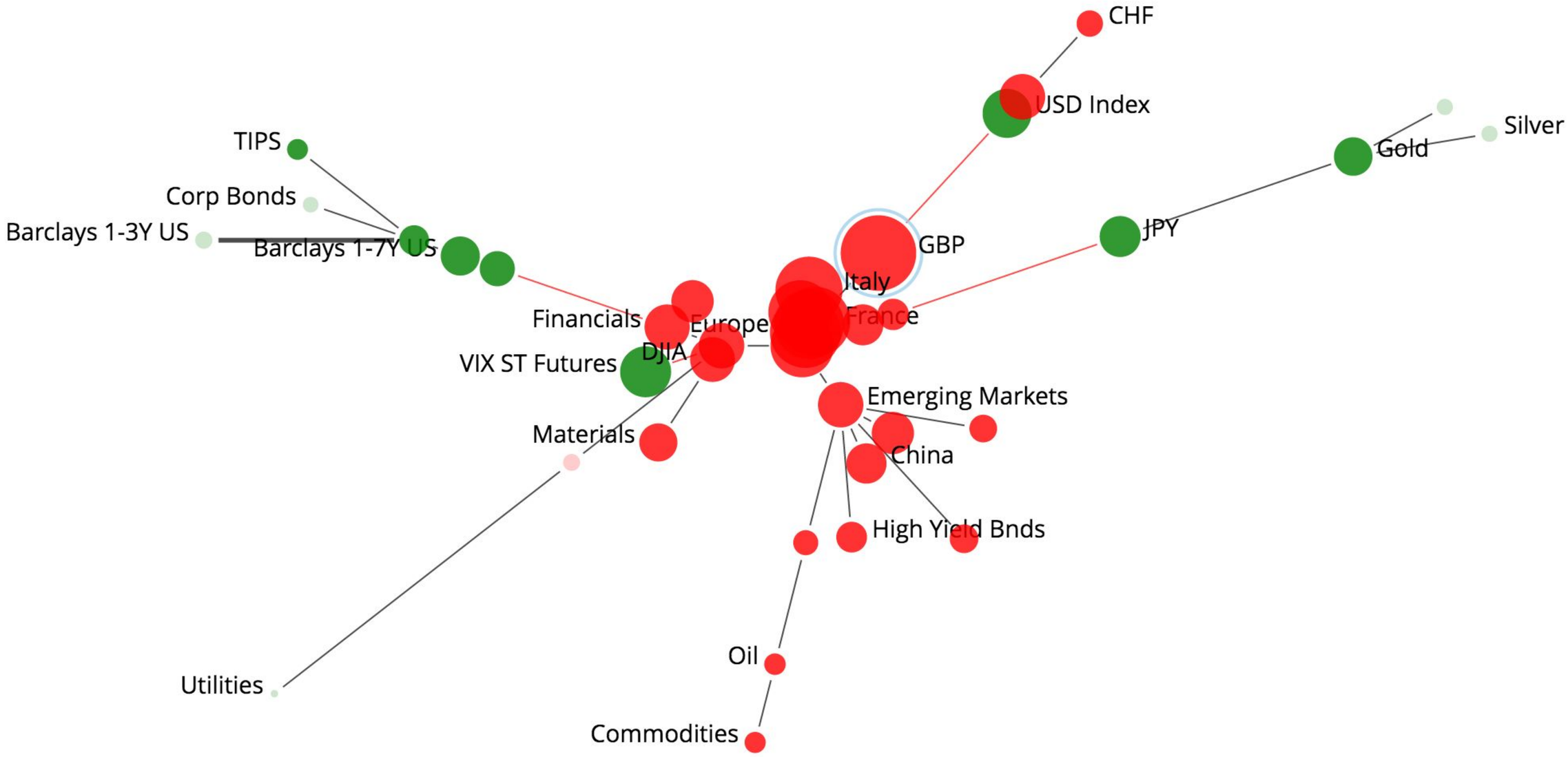
- Green = positive
- Red = negative

Node size indicates magnitude of return

Bright colors are VaR exceptions



Brexit, Friday 24 June 2016



Financial Cartography

Coordinate system

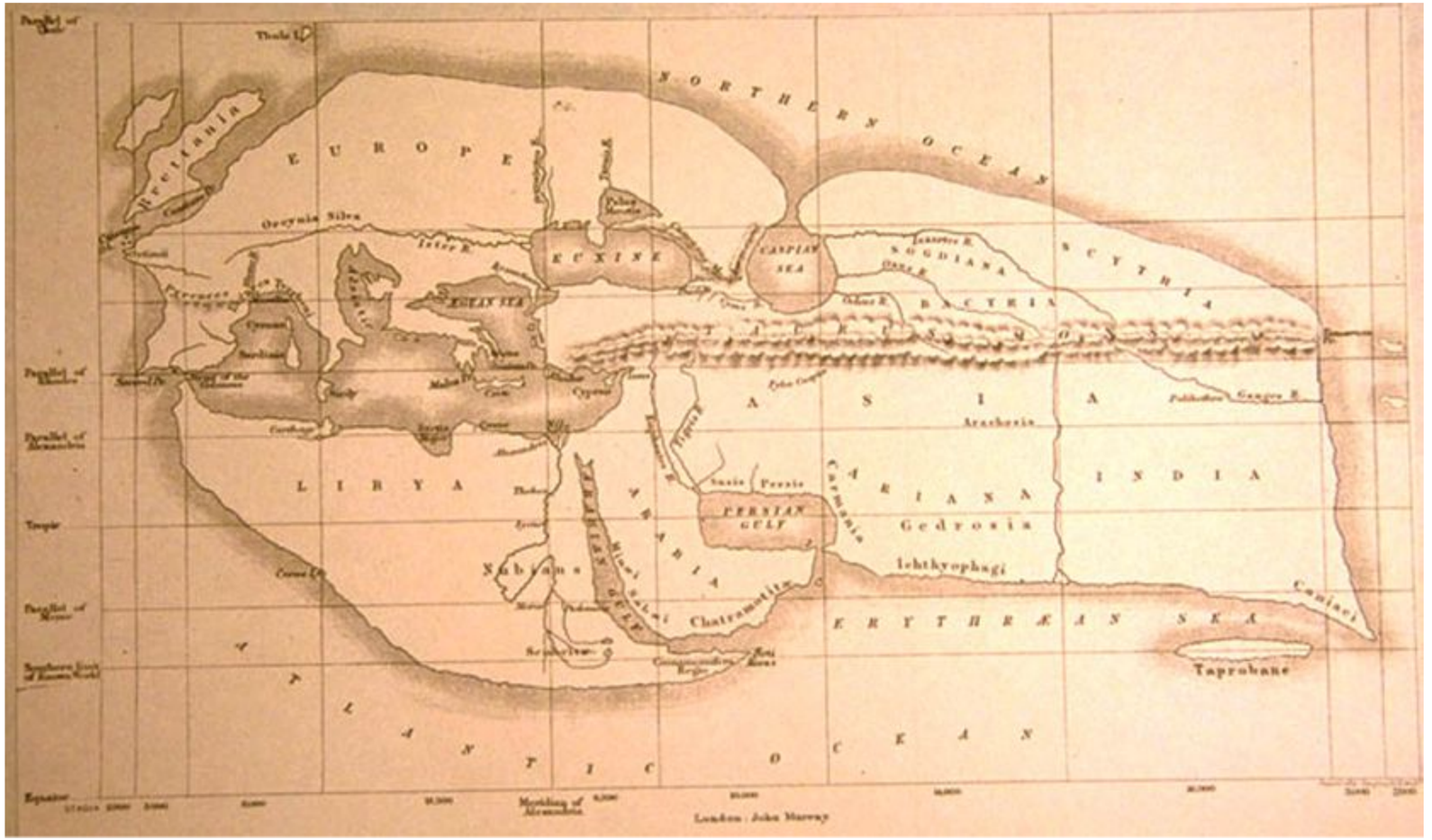
-> layout algorithm

System for visual encoding of map data

-> node sizes & colors

Dimensionality reduction & filtering

-> minimum spanning tree



Use Case: Monitoring Housing Markets

THE WALL STREET JOURNAL.

Subscribe | Sign In

BUSINESS | JOURNAL REPORTS: LEADERSHIP

New Tools Give Better Picture, Literally, of Financial-System Risk

Researchers are using network analytics and advanced data modeling to identify weak spots in the system that otherwise might go unnoticed

This sprawling tree shows housing prices in U.S. markets moving with little correlation in 2000. The tree has gotten shorter and shorter since, indicating higher correlation between markets.

PHOTO: FINANCIAL NETWORK ANALYTICS

Soramaki et al (2016). 'A Network-Based Method for Visual Identification of Systemic Risks': Journal of Network Theory in Finance

online | pdf

ASSET
NETWORK
STRESS
LIBRARY
ALERTS
PORTFOLIO
INFO
TOUR
FNA

House Price Index Correlations

Correlations in changes of House Price Index by state. Data source: Quandl / Federal Reserve Economic Data.

The data displays a clear cascade of negative house price shocks starting in Q1 2007. Also notable is the increase in the overall level of correlations (and connectedness) from 2000 to 2014.

How to read it

Nodes represent US States. Node color reflects change in price: green is positive, and red is negative. Node size scales with the magnitude of change, bigger nodes have larger price change. Outlier movements are marked with bright colors.

Links show strongest correlations. Among these correlations, shorter link means a stronger correlation.

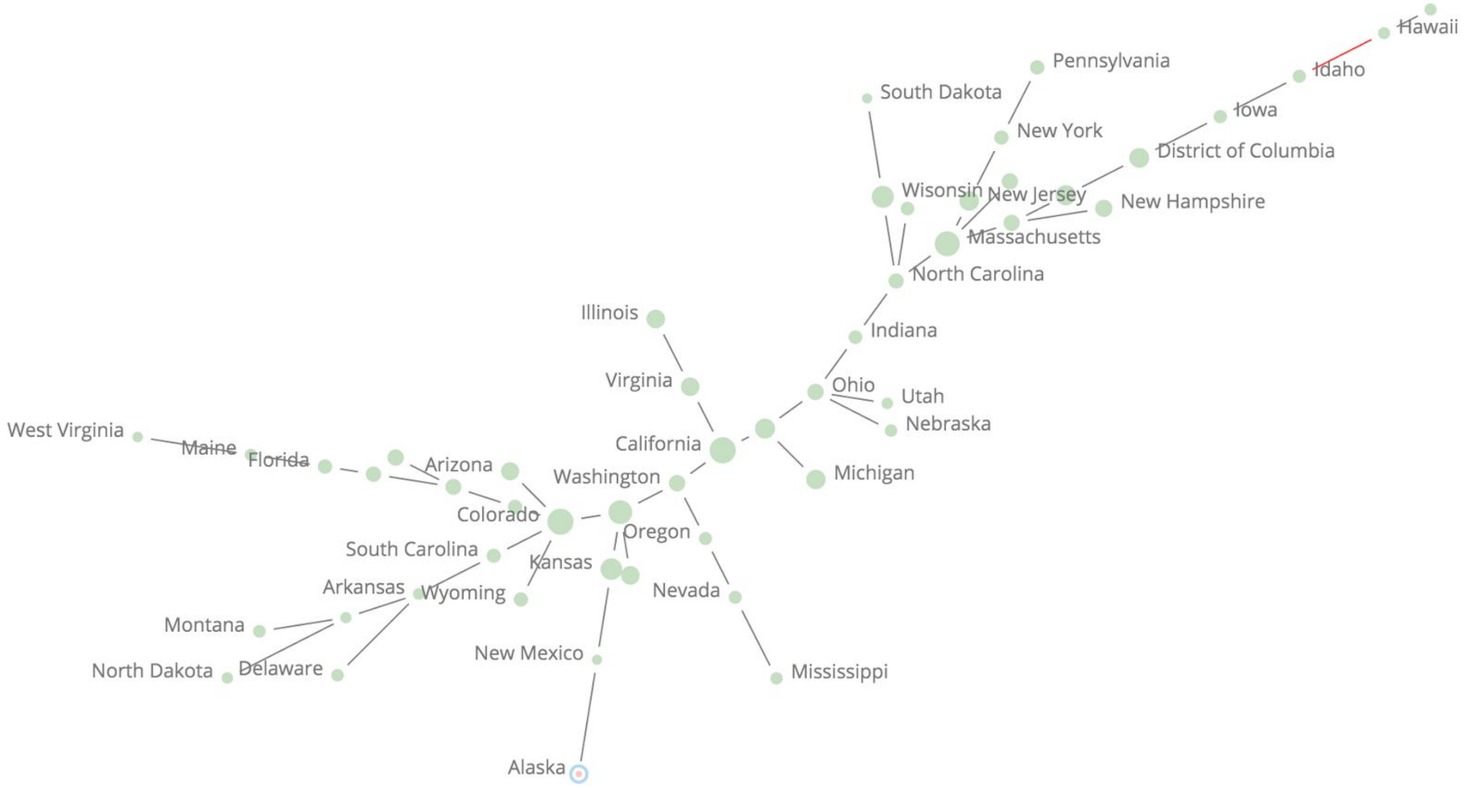
In this example we look at US house prices across states. We see the US states as nodes and strong correlation between house prices as link. In 2000 the tree is very spread out and prices are going slightly up. This is a time when ABS are developed with the assumption that real-estate risk can be diversified across US states.

Sat, 1 Jan 2000 PLAY 1 Jul 1990 - 1 Jul 2015

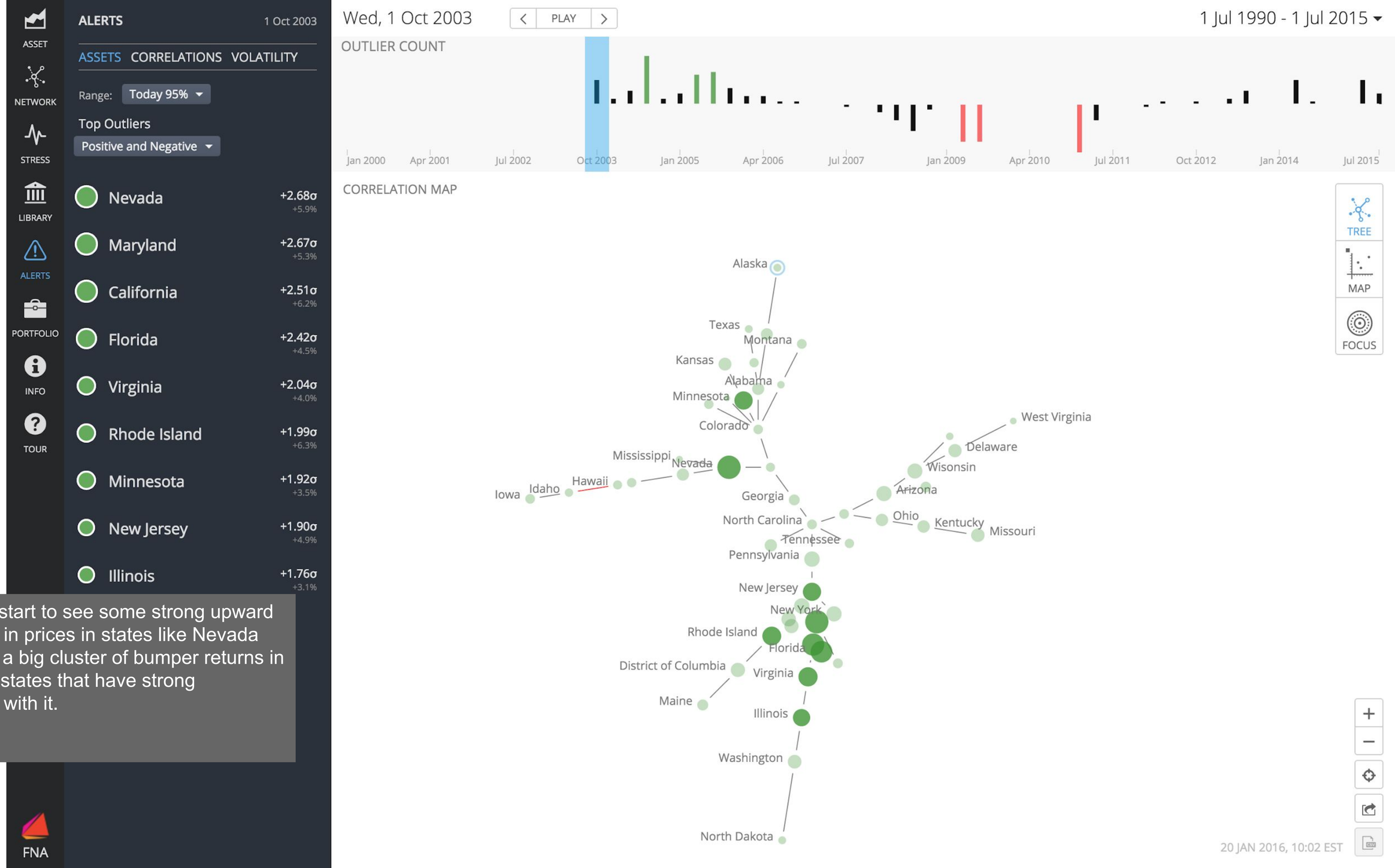
OUTLIER COUNT

CORRELATION MAP

TREE
MAP
FOCUS

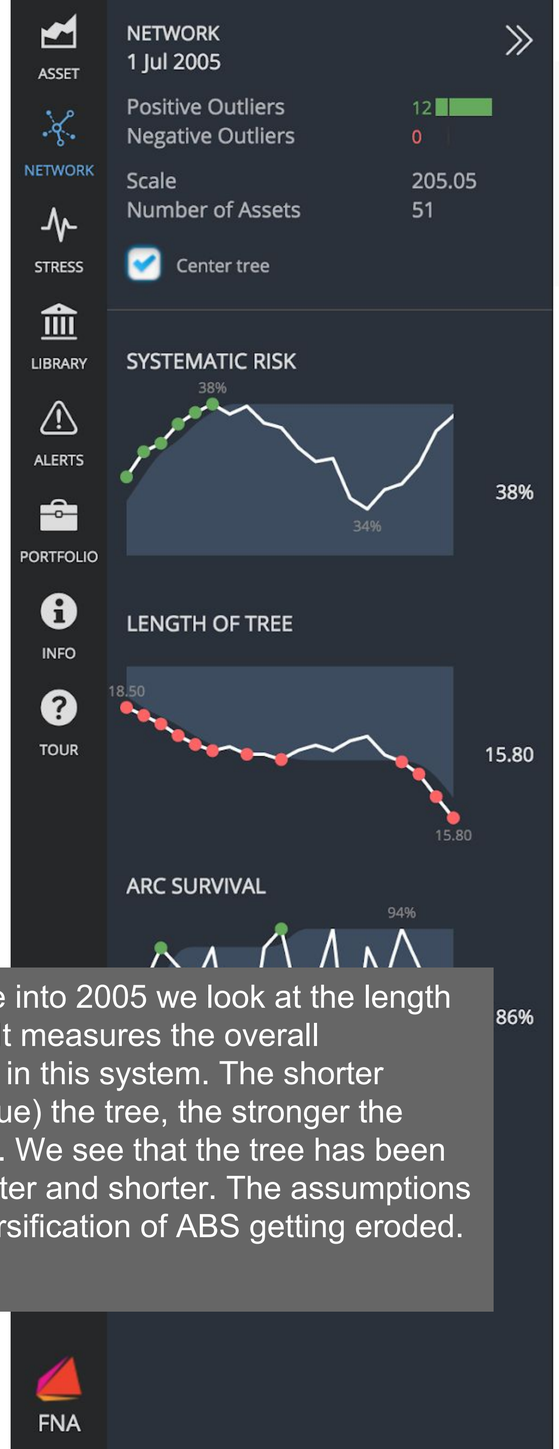


20 JAN 2016, 10:02 EST

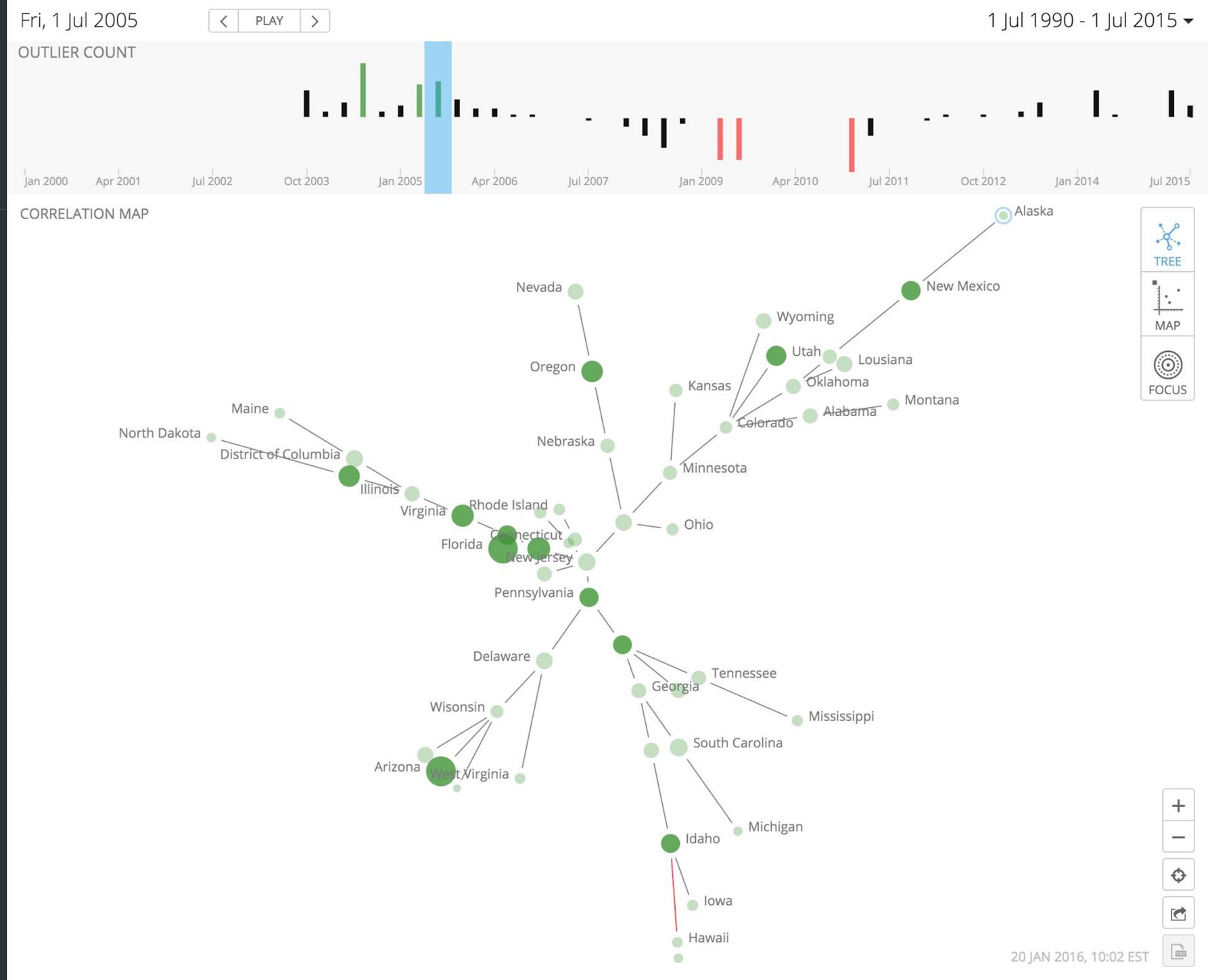


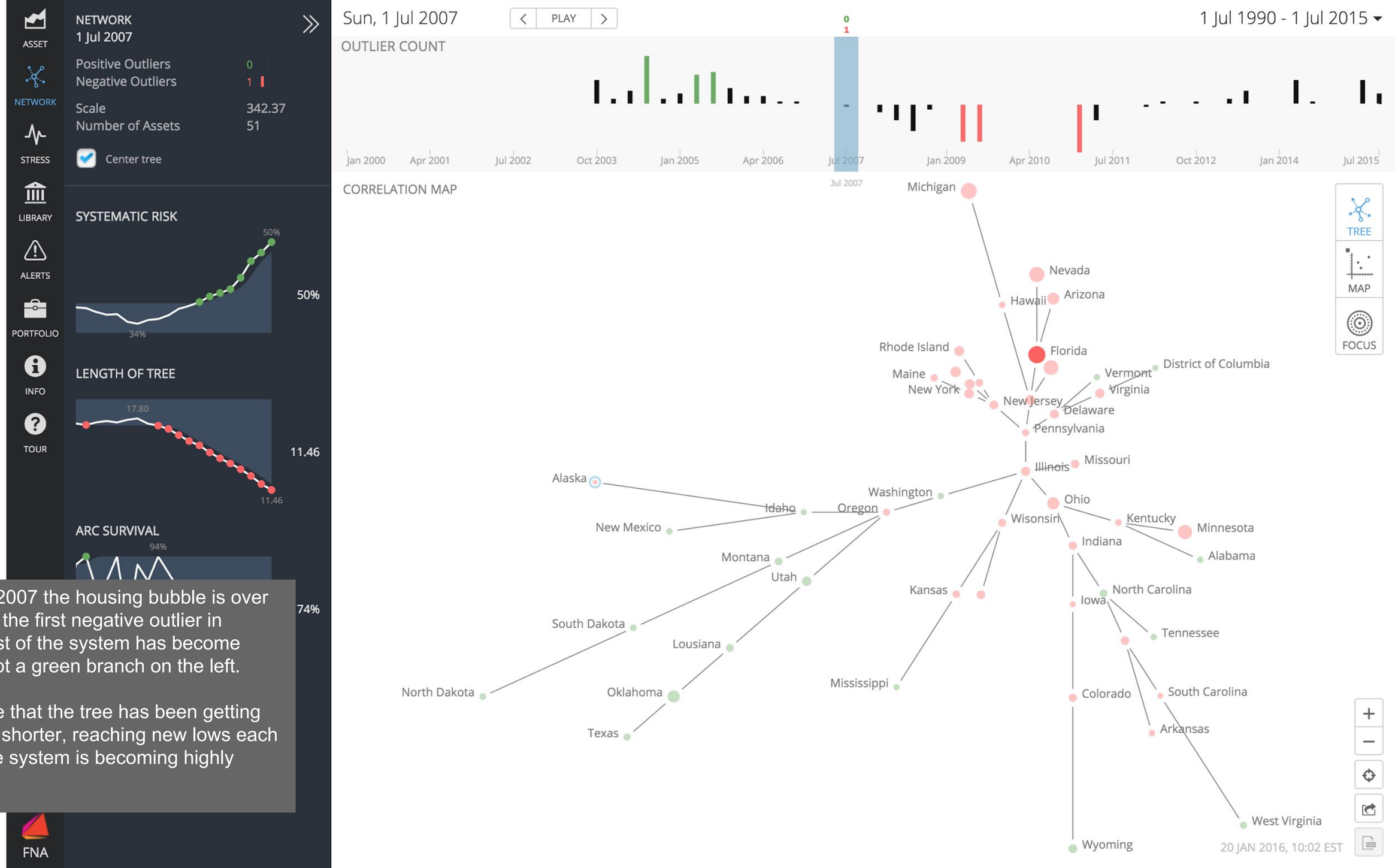
In 2003 we start to see some strong upward movements in prices in states like Nevada and we see a big cluster of bumper returns in Florida and states that have strong correlations with it.

As we move into 2004, into the peak of the housing bubble we see that most states now have outlier price changes and Nevada for example has an almost 12% rise in house prices in one quarter.



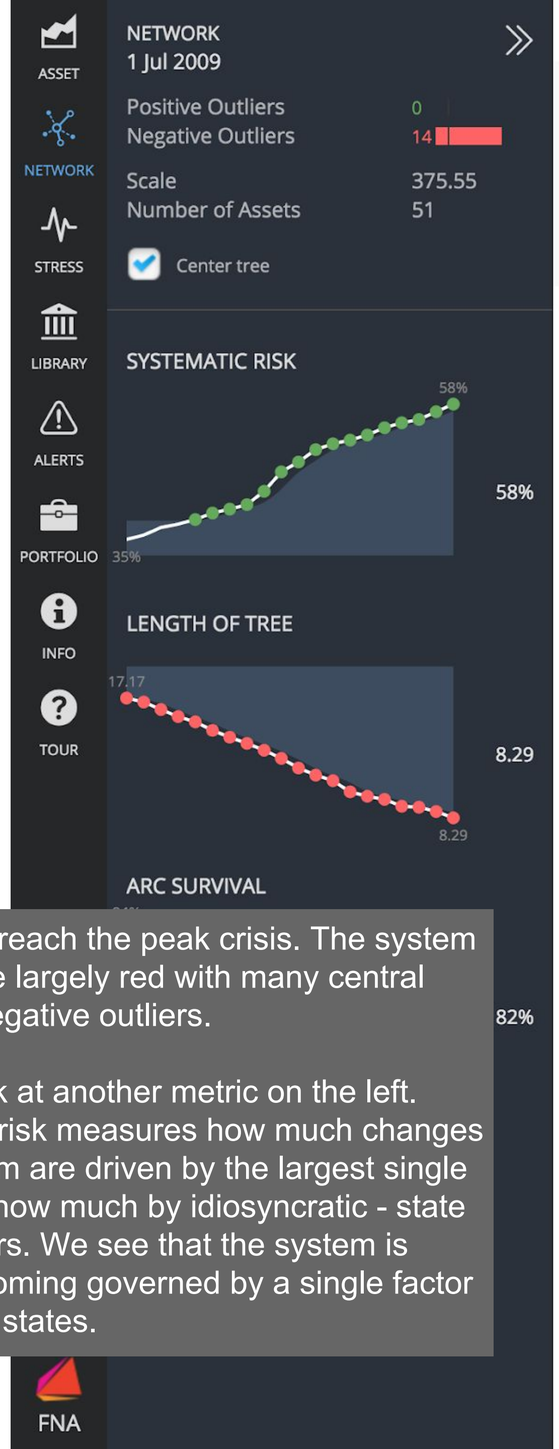
As we move into 2005 we look at the length of the tree. It measures the overall correlations in this system. The shorter (smaller value) the tree, the stronger the correlations. We see that the tree has been getting shorter and shorter. The assumptions behind diversification of ABS getting eroded.





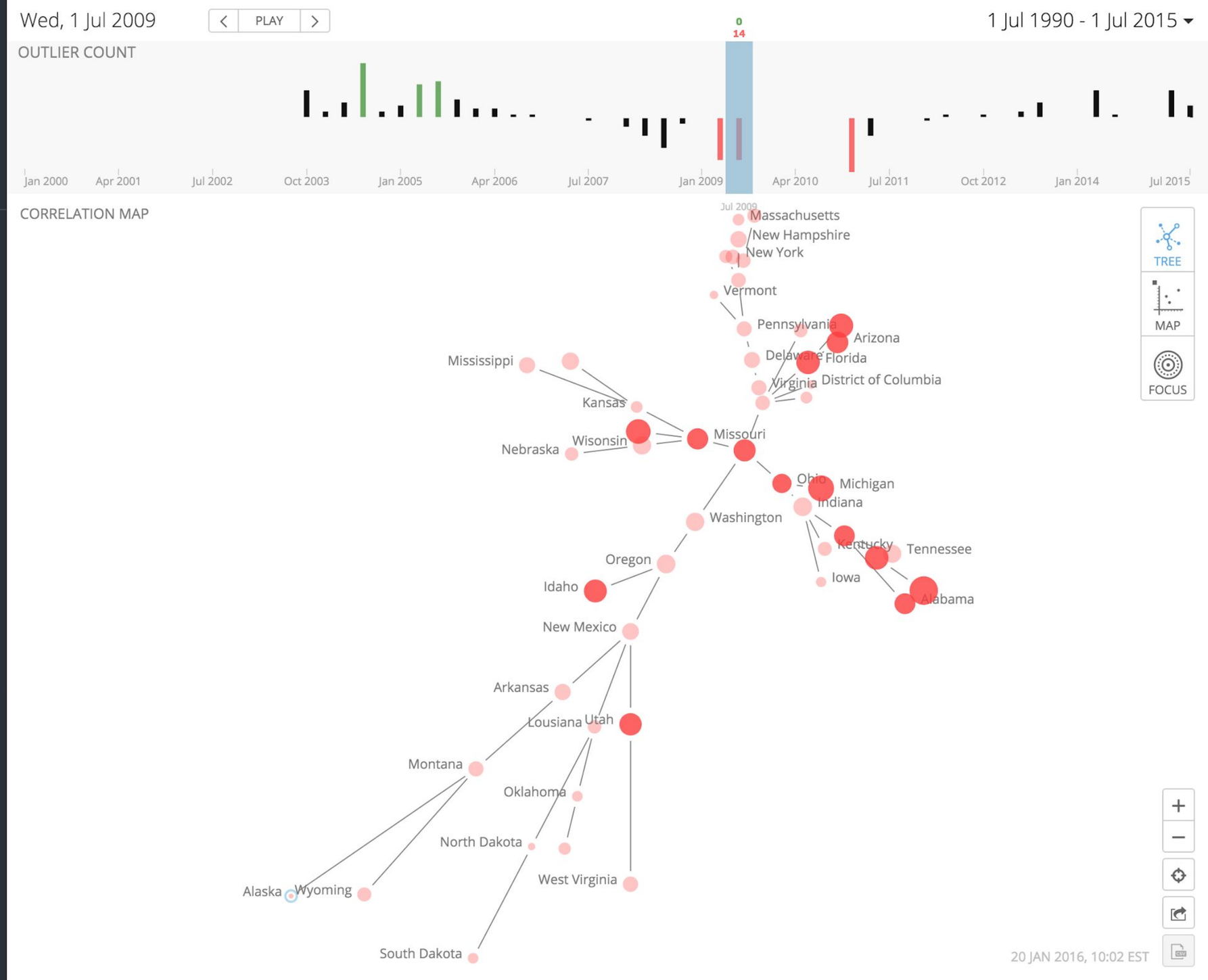
In summer 2007 the housing bubble is over and we see the first negative outlier in Florida. Most of the system has become redm, except a green branch on the left.

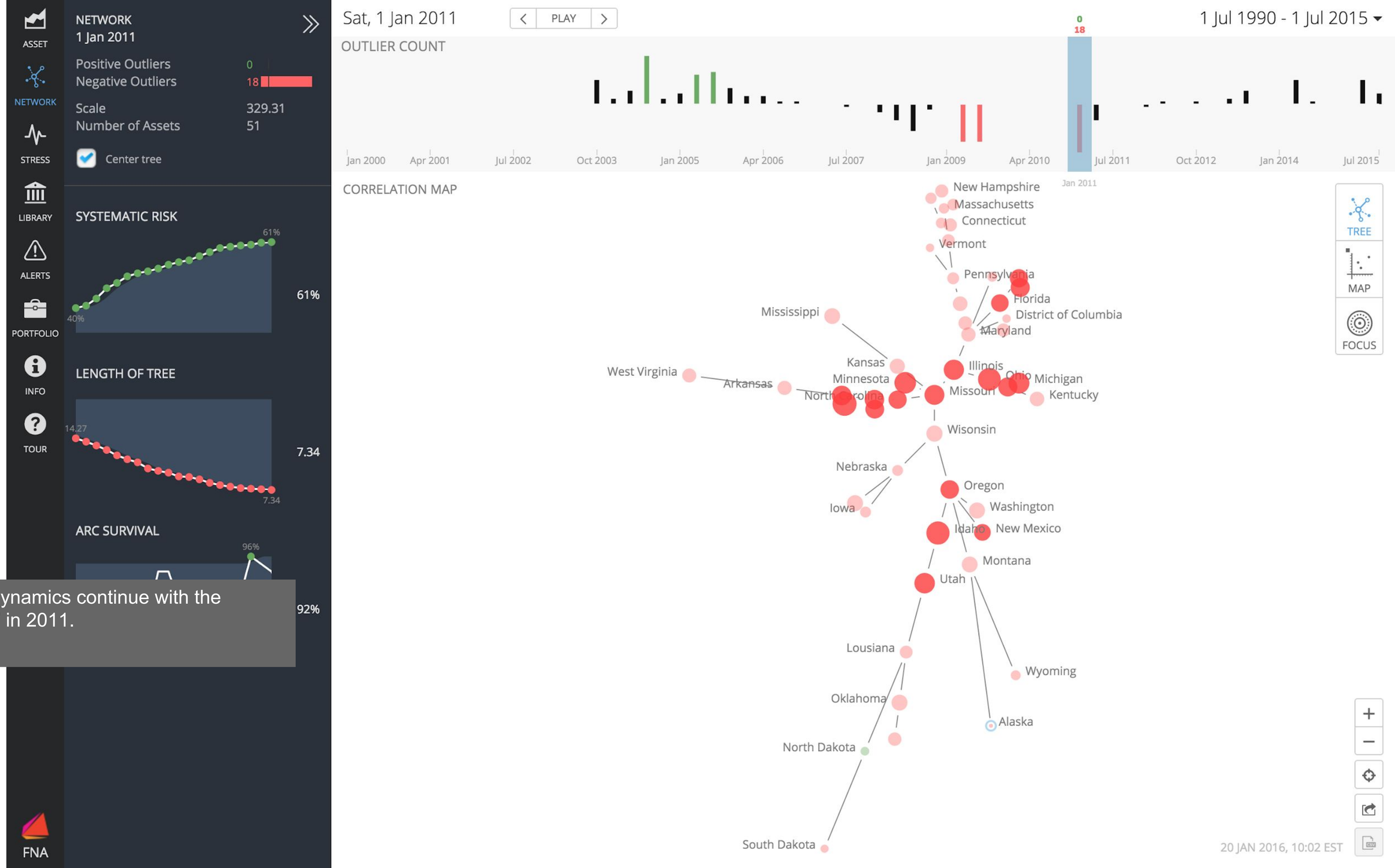
We also see that the tree has been getting shorter and shorter, reaching new lows each quarter. The system is becoming highly coupled.



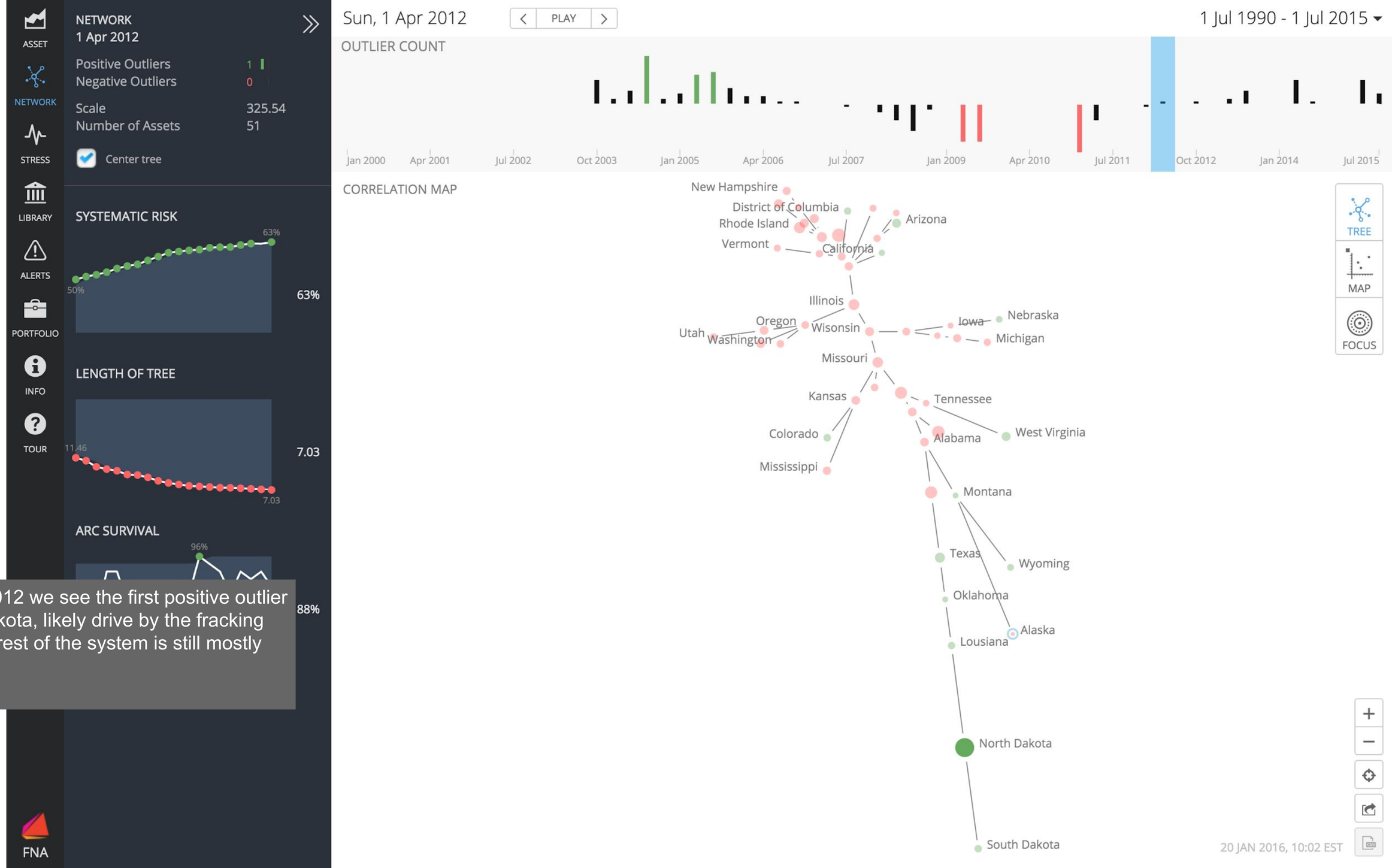
In 2009 we reach the peak crisis. The system has become largely red with many central states as negative outliers.

We can look at another metric on the left. Systematic risk measures how much changes in the system are driven by the largest single factor, and how much by idiosyncratic - state level - factors. We see that the system is quickly becoming governed by a single factor affecting all states.





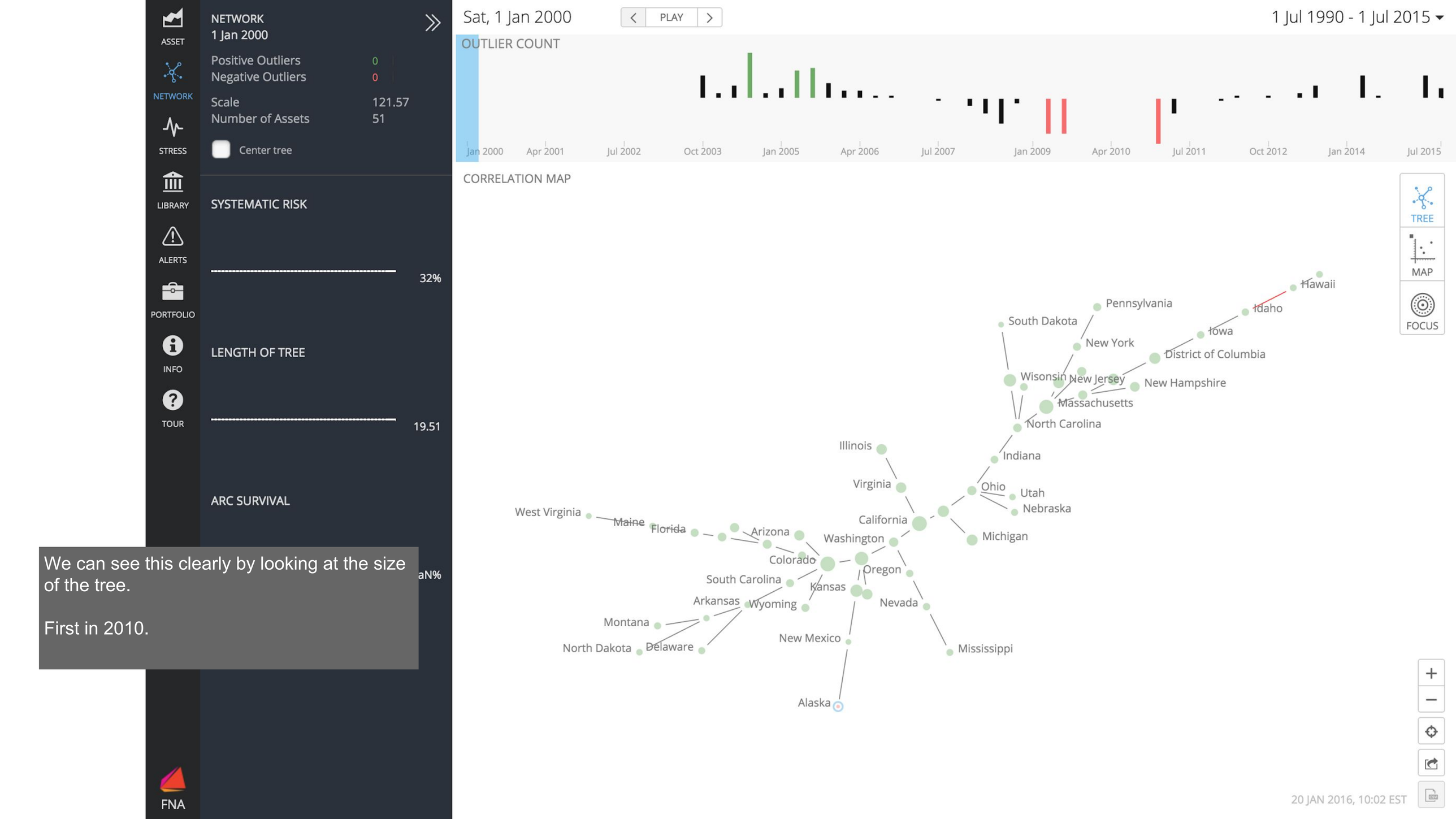
The same dynamics continue with the “double dip” in 2011.



In Spring 2012 we see the first positive outlier in North Dakota, likely drive by the fracking boom. The rest of the system is still mostly negative.

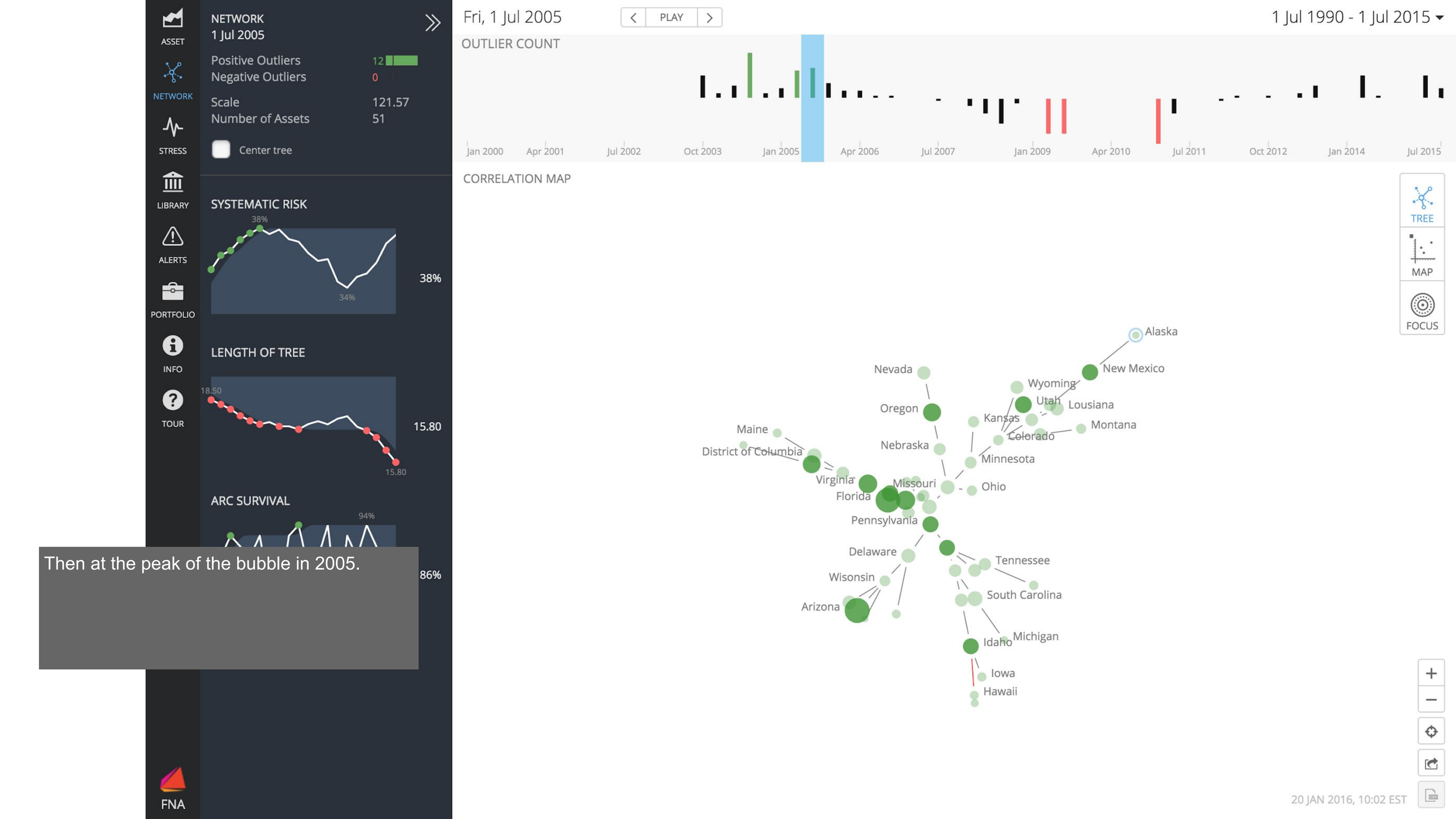
If we fast forward to the latest observation, July 2015, we see a period of positive changes in prices with outliers scattered across the network.

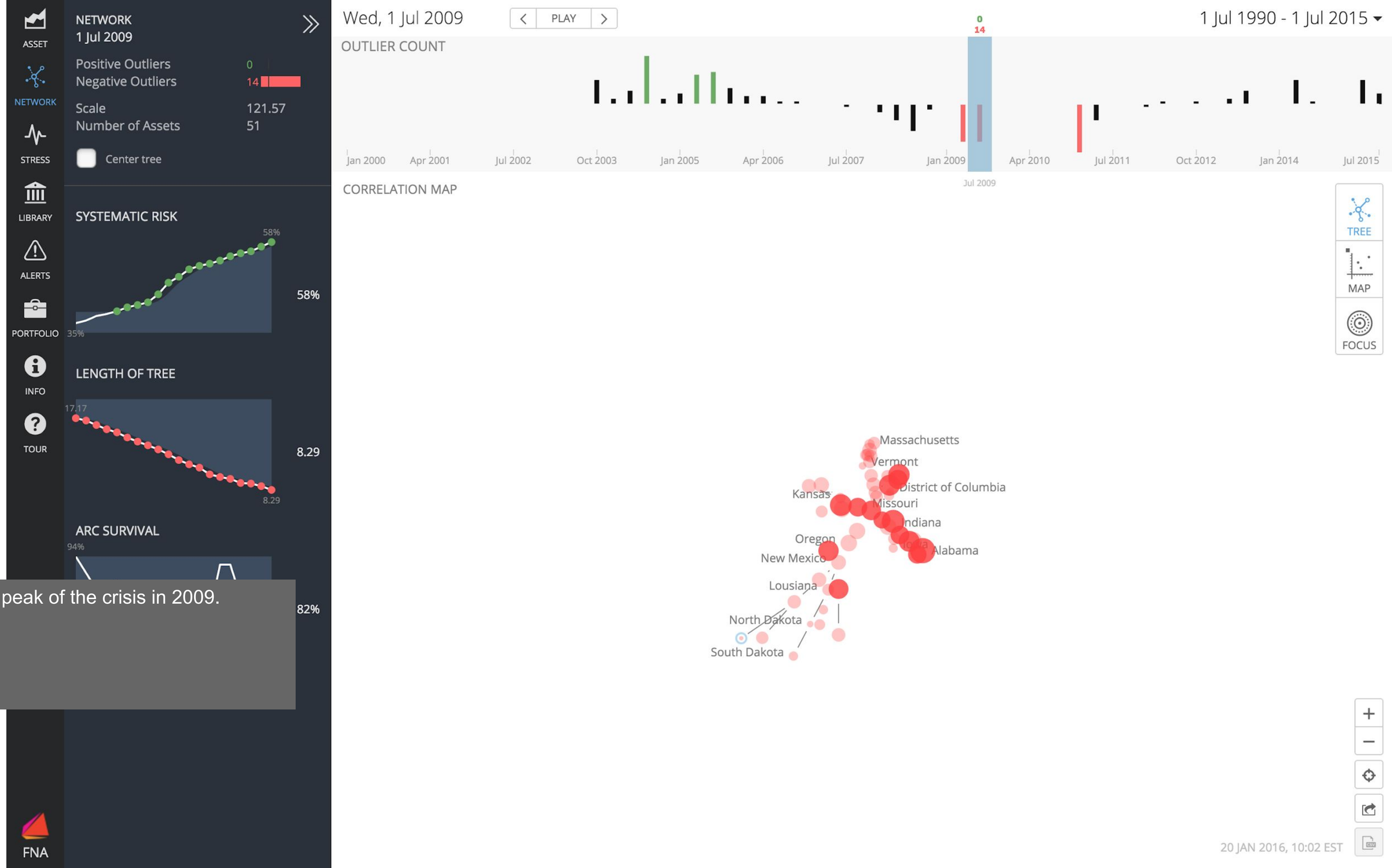
We also see both systematic risk and correlations at their peak. We have not returned to the pre-bubble system state but are in a very risky territory still..



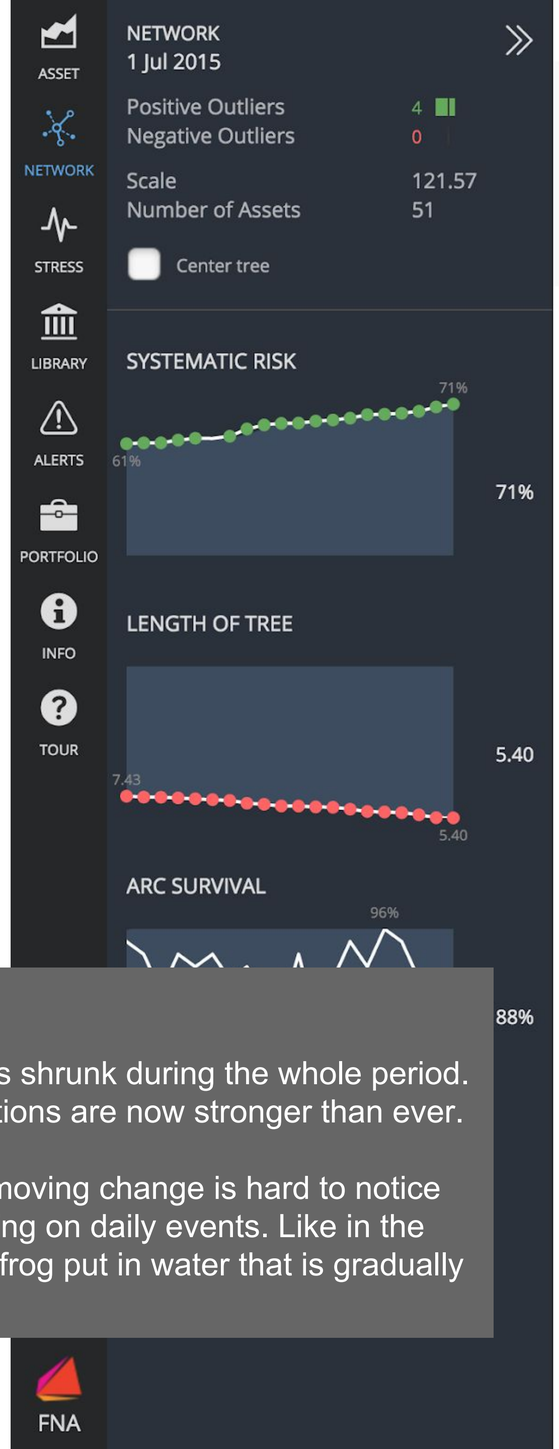
We can see this clearly by looking at the size of the tree.

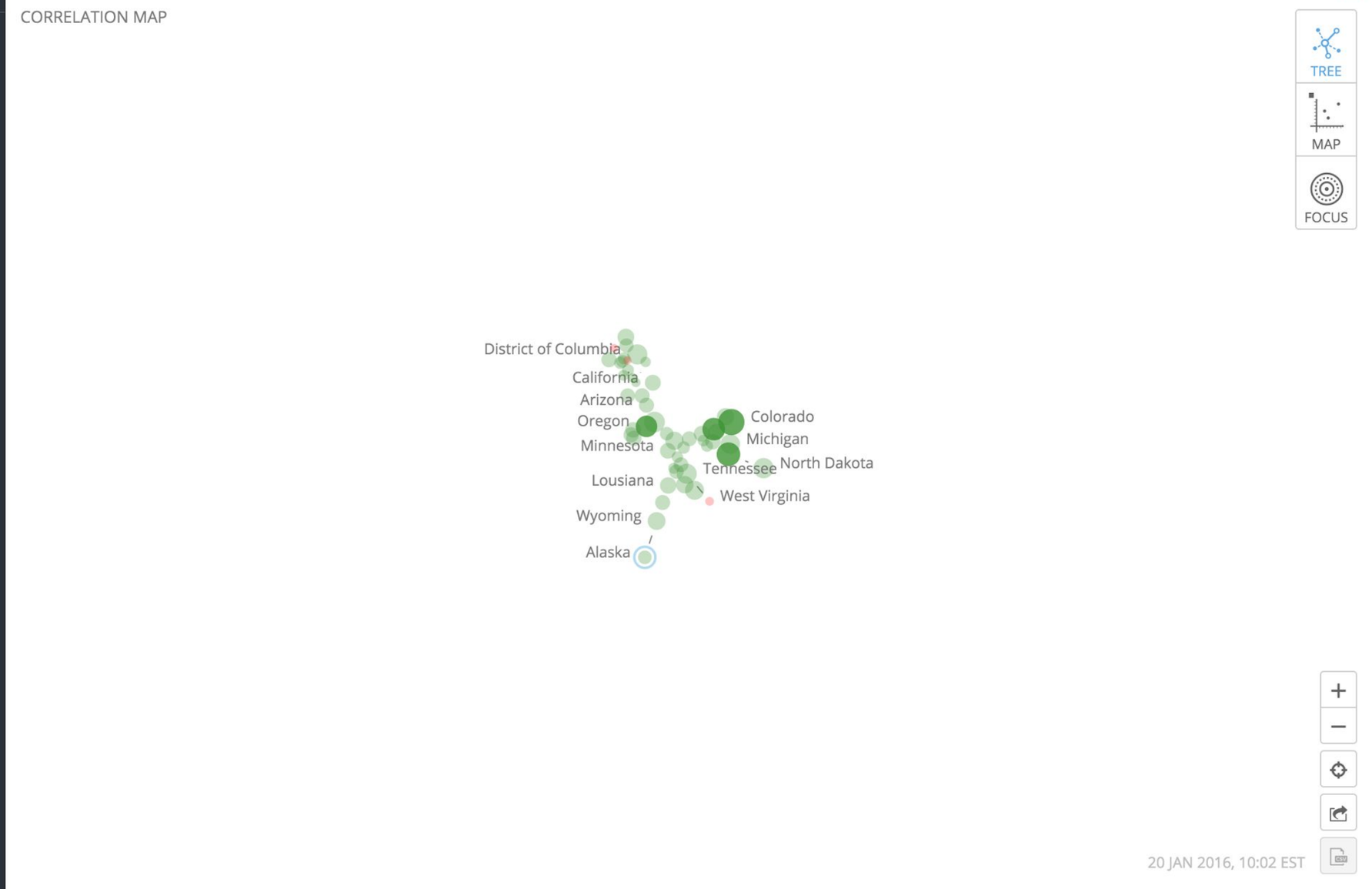
First in 2010.


ASSET
NETWORK
STRESS
LIBRARY
ALERTS
PORTFOLIO
INFO
TOUR
FNA
NETWORK
1 Jul 2005
Positive Outliers 12
Negative Outliers 0
Scale 121.57
Number of Assets 51
Center tree
SYSTEMATIC RISK
38%
34%
38%
LENGTH OF TREE
18.50
15.80
15.80
ARC SURVIVAL
94%
86%
Then at the peak of the bubble in 2005.
Fri, 1 Jul 2005
PLAY
1 Jul 1990 - 1 Jul 2015
OUTLIER COUNT
Jan 2000
Apr 2001
Jul 2002
Oct 2003
Jan 2005
Apr 2006
Jul 2007
Jan 2009
Apr 2010
Jul 2011
Oct 2012
Jan 2014
Jul 2015
CORRELATION MAP
TREE
MAP
FOCUS
Alaska
New Mexico
Nevada
Wyoming
Utah
Lousiana
Oregon
Kansas
Montana
Colorado
Maine
Nebraska
District of Columbia
Minnesota
Virginia
Missouri
Ohio
Florida
Pennsylvania
Delaware
Tennessee
Wisonsin
South Carolina
Arizona
Michigan
Idaho
Iowa
Hawaii
20 JAN 2016, 10:02 EST



Then at the peak of the crisis in 2009.

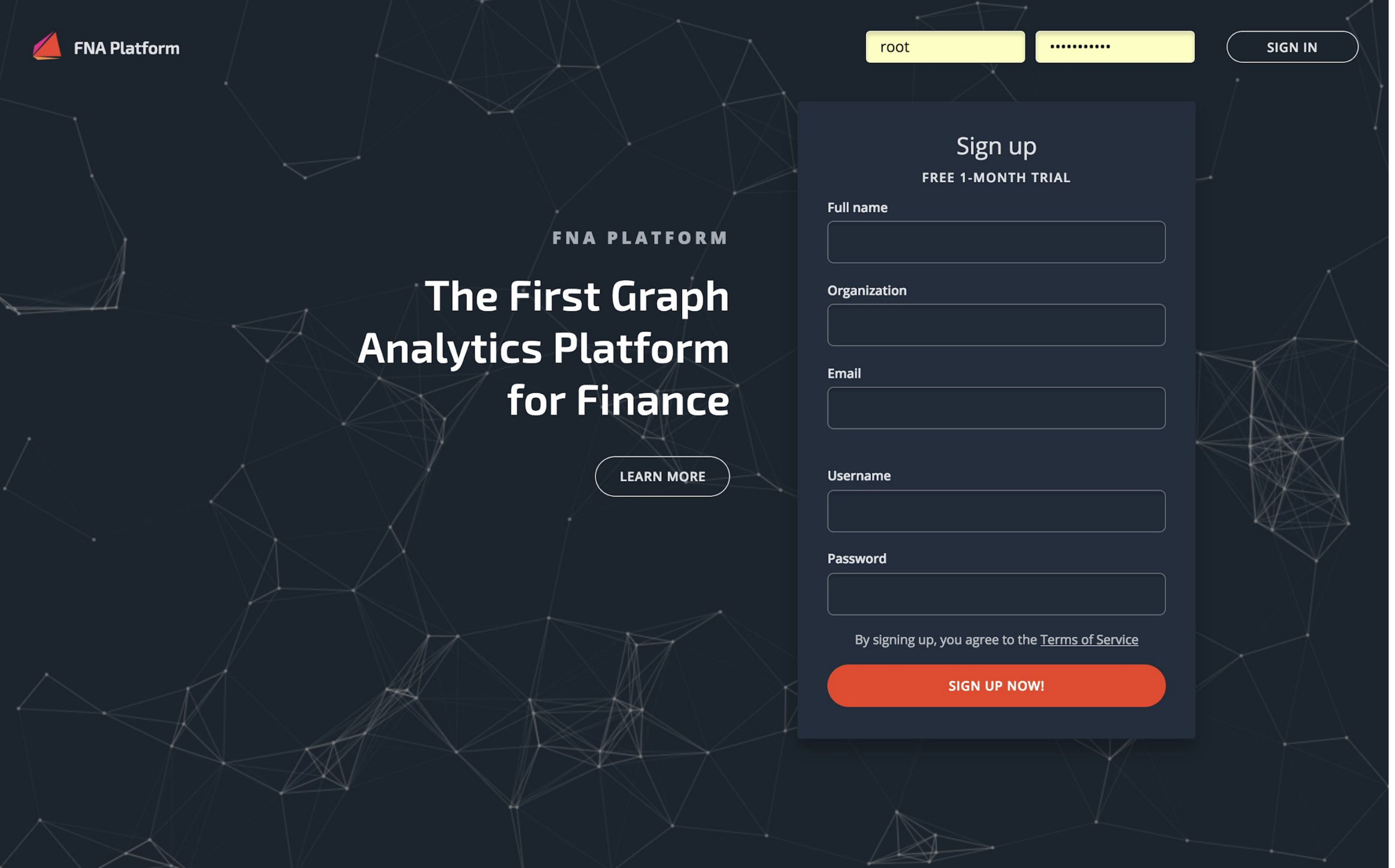




And now.

The tree has shrunk during the whole period. The correlations are now stronger than ever.

Such slow moving change is hard to notice when focusing on daily events. Like in the story of the frog put in water that is gradually heated.


FNA Platform
root
SIGN IN
FNA PLATFORM
The First Graph Analytics Platform for Finance
LEARN MORE
Sign up
FREE 1-MONTH TRIAL
Full name
Organization
Email
Username
Password
By signing up, you agree to the Terms of Service
SIGN UP NOW!

FNA

Dr. Kimmo Soramäki
Founder & CEO
FNA - Financial Network Analysis Ltd.

kimmo@fna.fi
tel. +44 20 3286 1111

Address
4-8 Crown Place
London EC2A 4BT
United Kingdom

FNA

Introduction to Network Science & Visualization II

Dr. Kimmo Soramäki
Founder & CEO, FNA

Agenda

Financial Crime & Cyber Risks

- Fraud, AML & KYCC
- DDoS Attacks
- Related Parties Analysis

Financial Market Infrastructures

- Monitoring Members
- Designing liquidity efficient FMIs
- Predicting Liquidity
- Detecting Anomalous transactions

FNA

Fraud

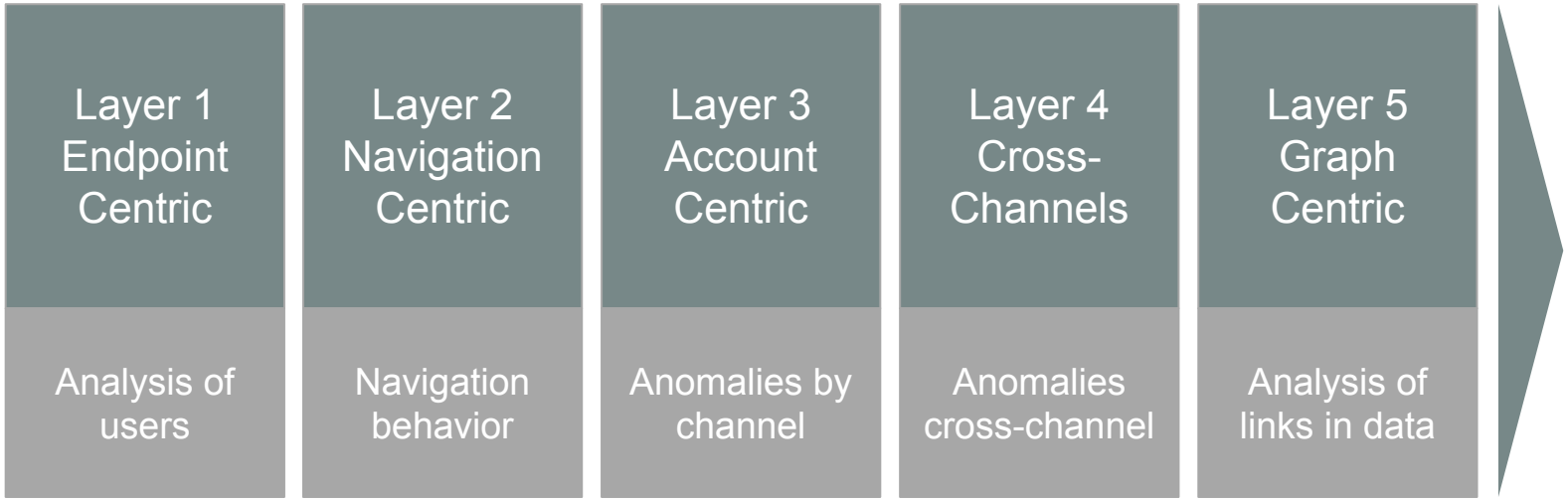
The Challenge

Intensified regulatory pressures has increased the number of false positives generated by existing software solutions

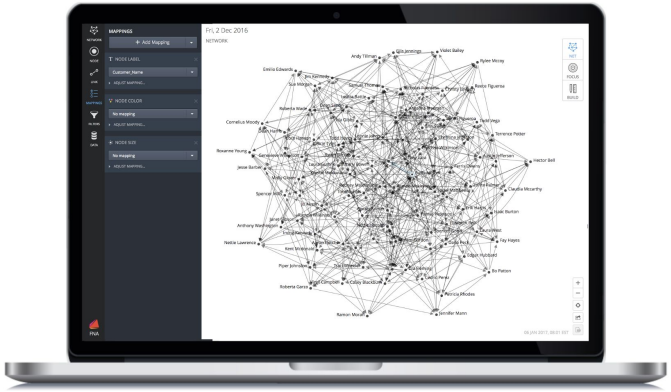
It is increasingly difficult for banks and financial institutions to quickly identify fraudsters.

The cost is $50 billion in fraudulent transactions happening each year.

Gartner's Layered Model of Fraud Prevention



Use Case: Improving Fraud Detection



Background
In 2017 global banks were fined £5B for failures to detect and address money laundering. Current methods are insufficient in identifying money laundering, and costly in terms of large amounts of manual labour needed.

Method
Payments form networks which can be automatically analyzed by network science algorithms. Existing research on large datasets proved that particular graph properties are good predictors of fraudulent transactions.

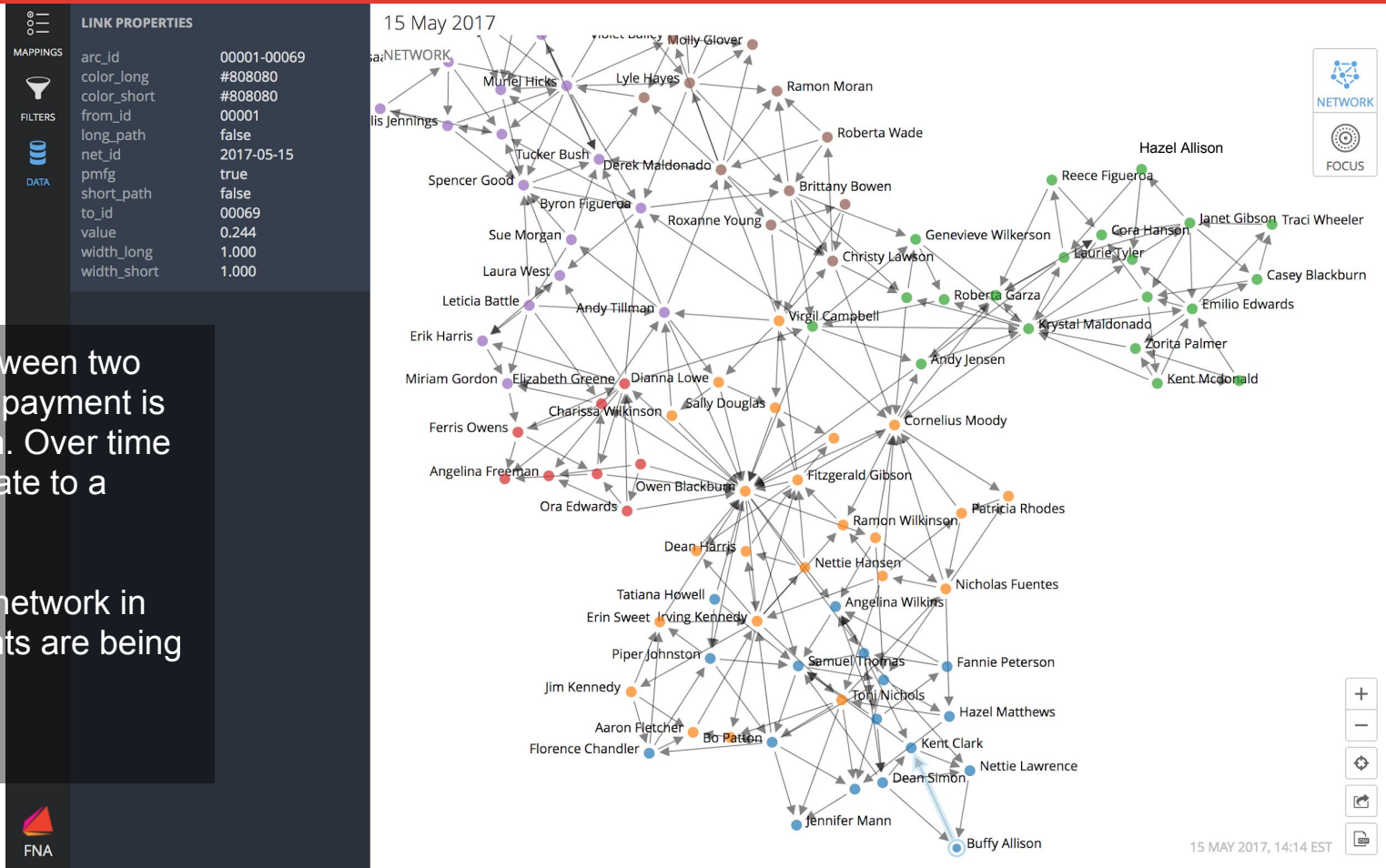
Benefits
Graphs improve fraud detection by eliminating false positives and identifying true positives more accurately - saving time and money.

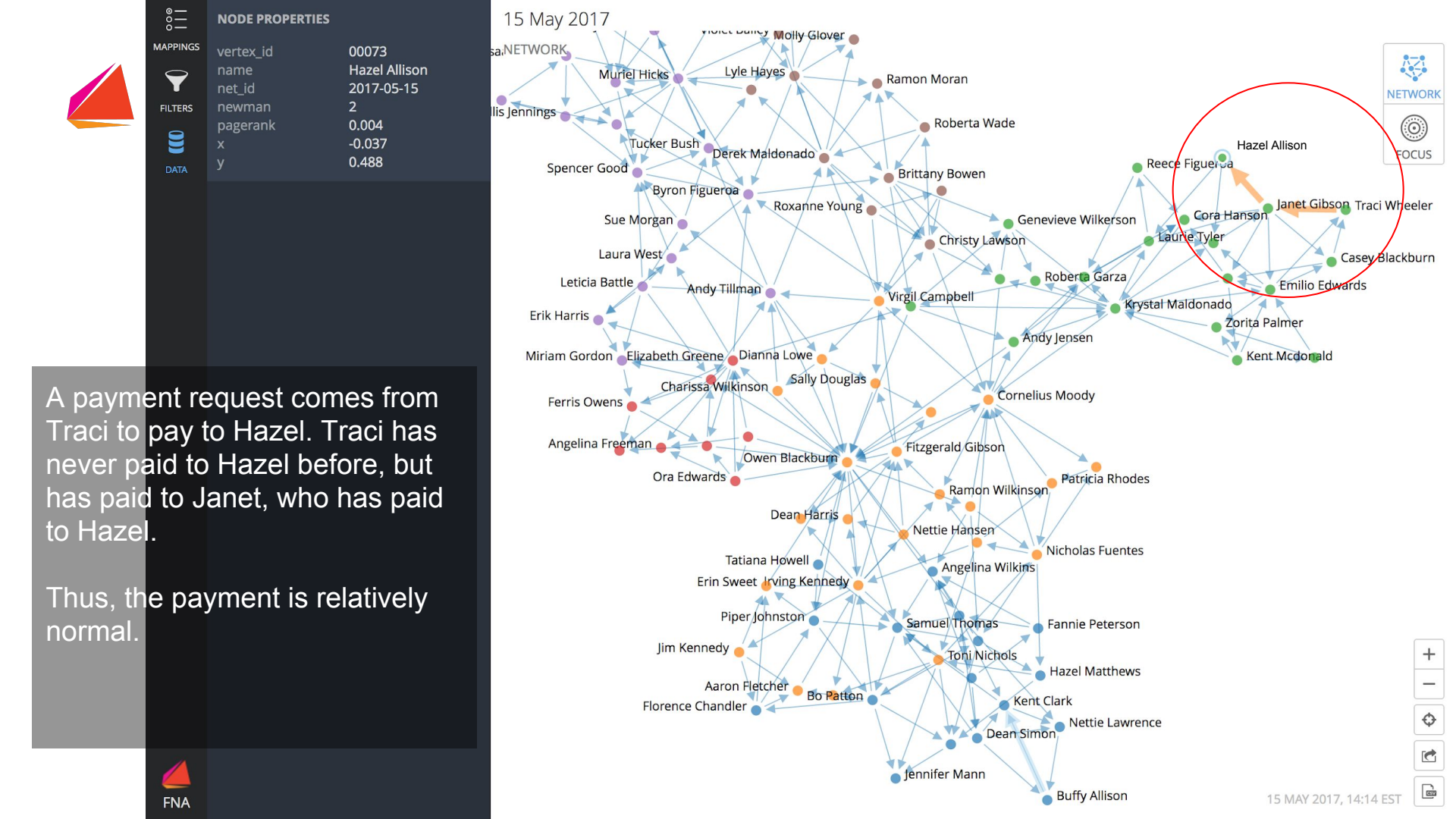
Molloy et al (2017), Graph Analytics for Real-time Scoring of Cross-channel Transactional Fraud

Basic Idea: Using Graphs in Automated Fraud Detection

We create a link between two account holders if a payment is made between them. Over time these links accumulate to a network.

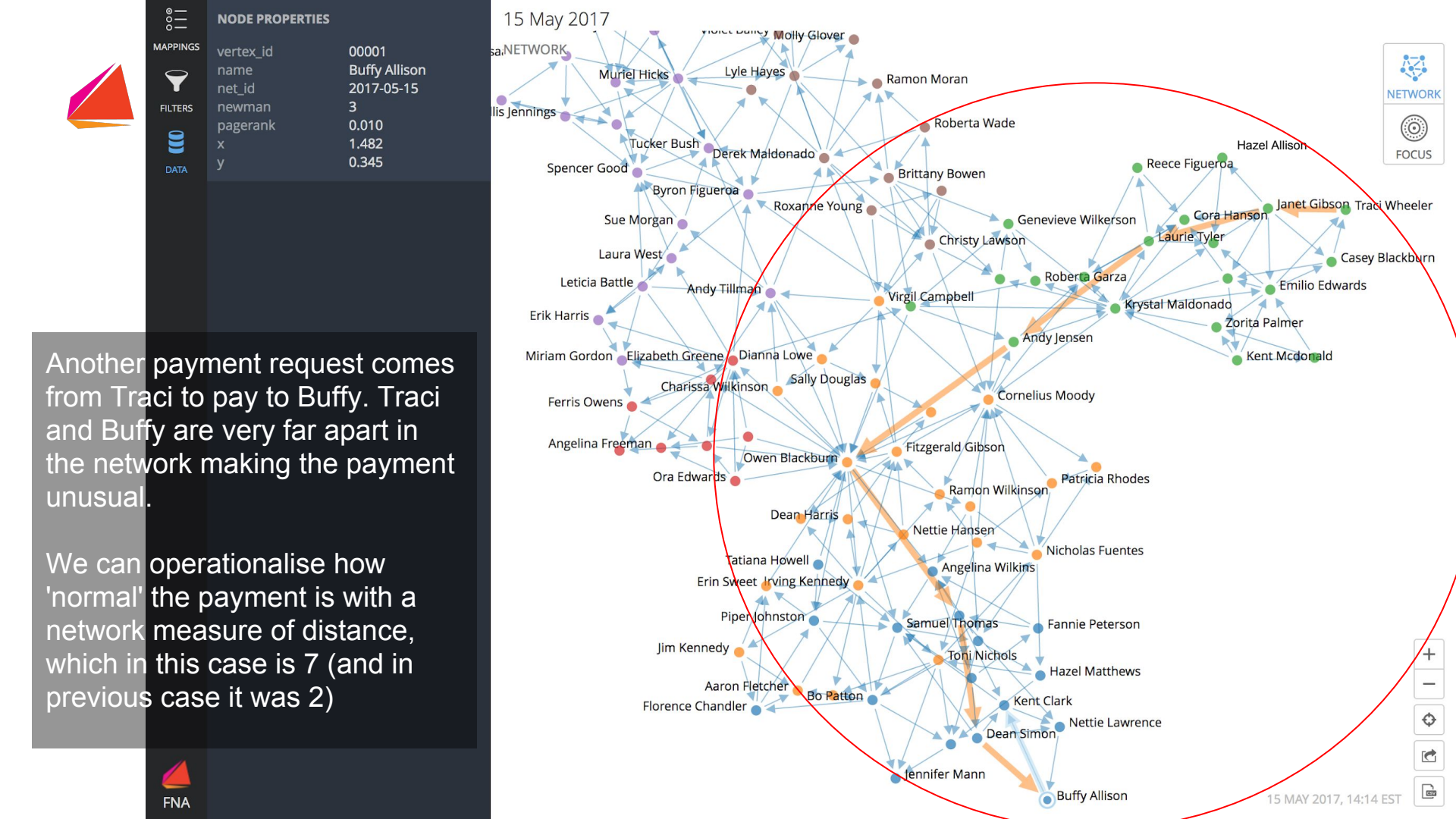
We can update the network in real-time as payments are being processed.

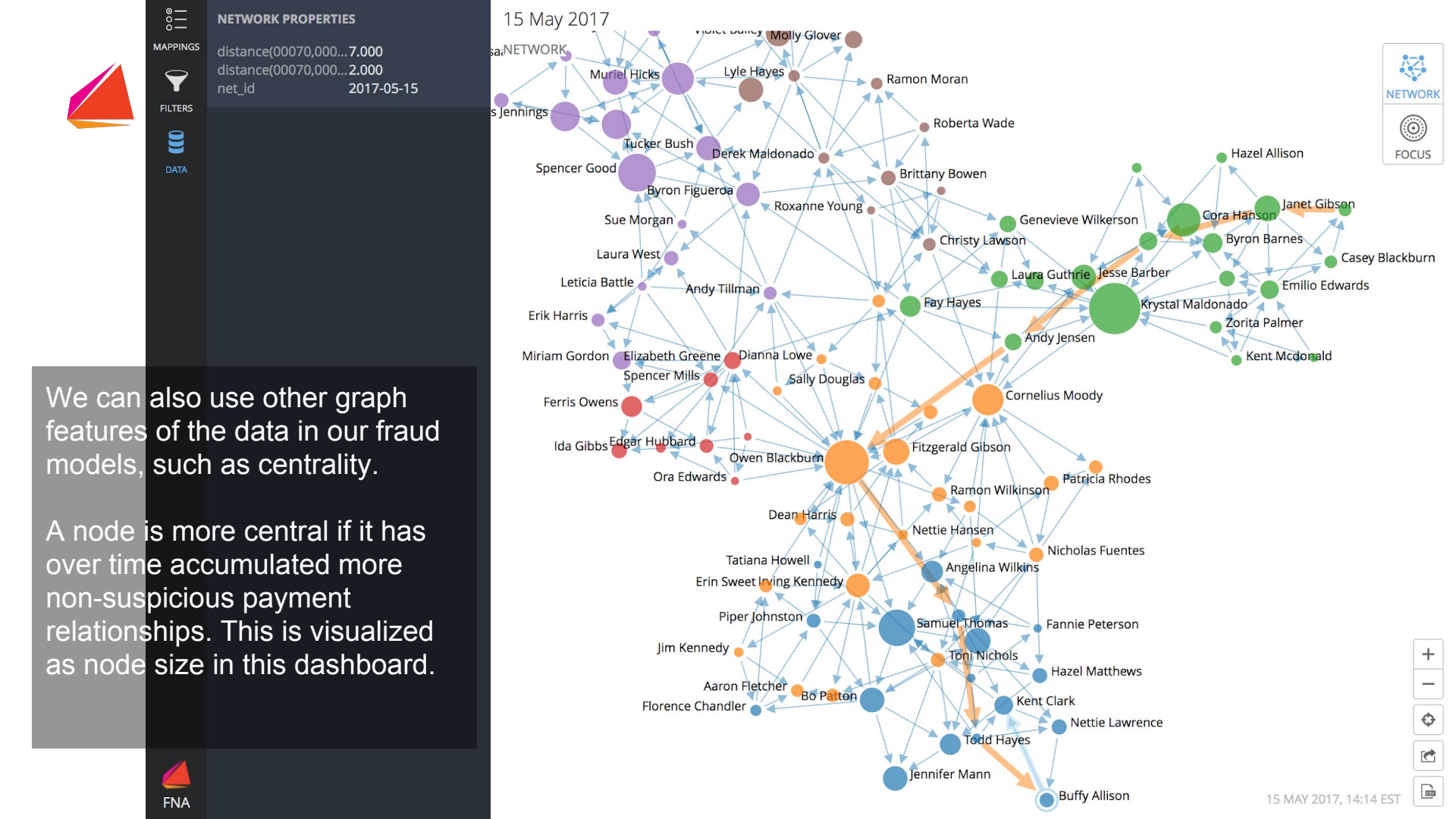




A payment request comes from Traci to pay to Hazel. Traci has never paid to Hazel before, but has paid to Janet, who has paid to Hazel.

Thus, the payment is relatively normal.


MAPPINGS
FILTERS
DATA
NODE PROPERTIES
vertex_id 00001
name Buffy Allison
net_id 2017-05-15
newman 3
pagerank 0.010
x 1.482
y 0.345
15 May 2017
NETWORK
FOCUS
Another payment request comes from Traci to pay to Buffy. Traci and Buffy are very far apart in the network making the payment unusual.
We can operationalise how 'normal' the payment is with a network measure of distance, which in this case is 7 (and in previous case it was 2)
FNA
15 MAY 2017, 14:14 EST



We can also use other graph features of the data in our fraud models, such as centrality.

A node is more central if it has over time accumulated more non-suspicious payment relationships. This is visualized as node size in this dashboard.

FNA

AML and Suspicious Activity

Company Interconnectedness

Challenge

Understand corporate interconnections for due diligence, fraud/criminal investigation, KYC, KYCC.

Current Situation

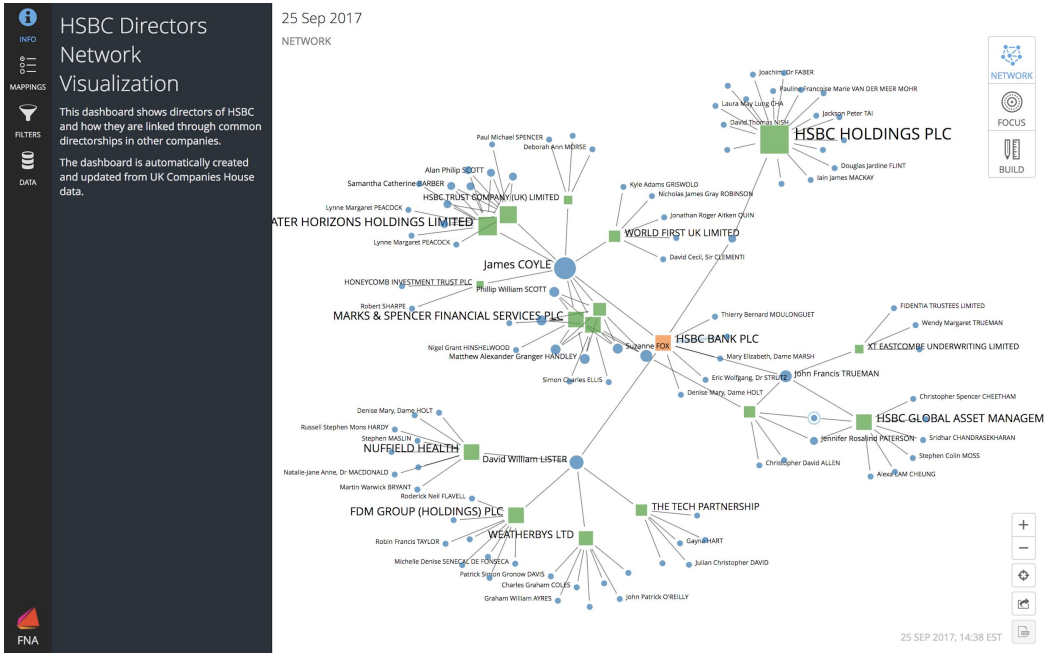
Manual investigation.

Solution

FNA has built a connection to Companies House register to automatically build graphs for any UK company.

Benefits

Save time and achieve a systemic view of company interconnectedness.



FNA DASHBOARDS EDITOR CONTACT US HELP DEMOS

MAPPINGS FILTERS DATA

7% NODES 1% NODE VALUE 3% LINKS 100% LINK VALUE

+ Add Filter

Hide disconnected nodes

NODE FILTER

ego_d_CA_UK

Data range

0 1 2 2 3

6 May 2018

NETWORK

NETWORK FOCUS BUILD

CAMBRIDGE ANALYTICA(UK) LIMITED

Alexander James Ashburner NIX

06 MAY 2018, 12:56 EST

FNA DASHBOARDS EDITOR CONTACT US HELP DEMOS

MAPPINGS FILTERS DATA

41% NODES 28% NODE VALUE 33% LINKS 100% LINK VALUE

+ Add Filter

Hide disconnected nodes

NODE FILTER

ego_d_CA_UK

Data range

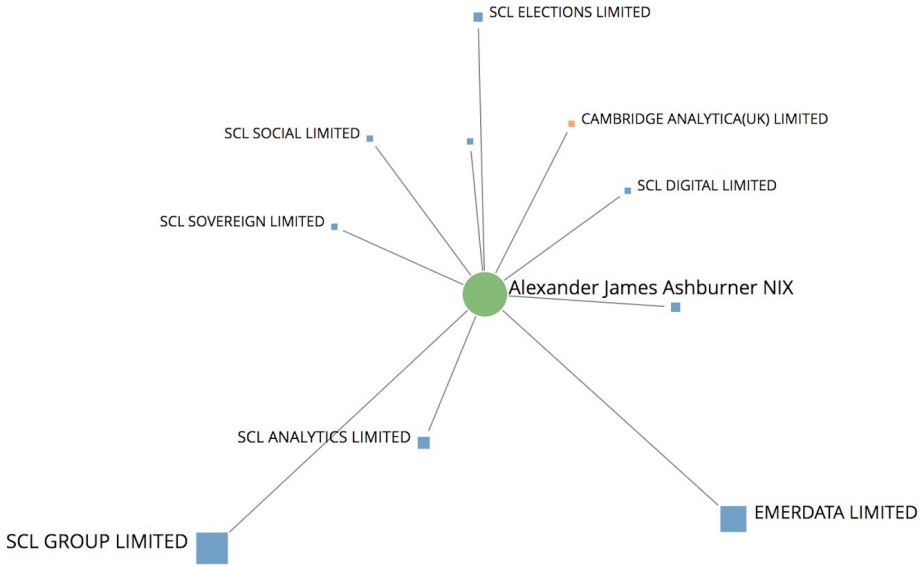
0 2

0 1 2 2 3

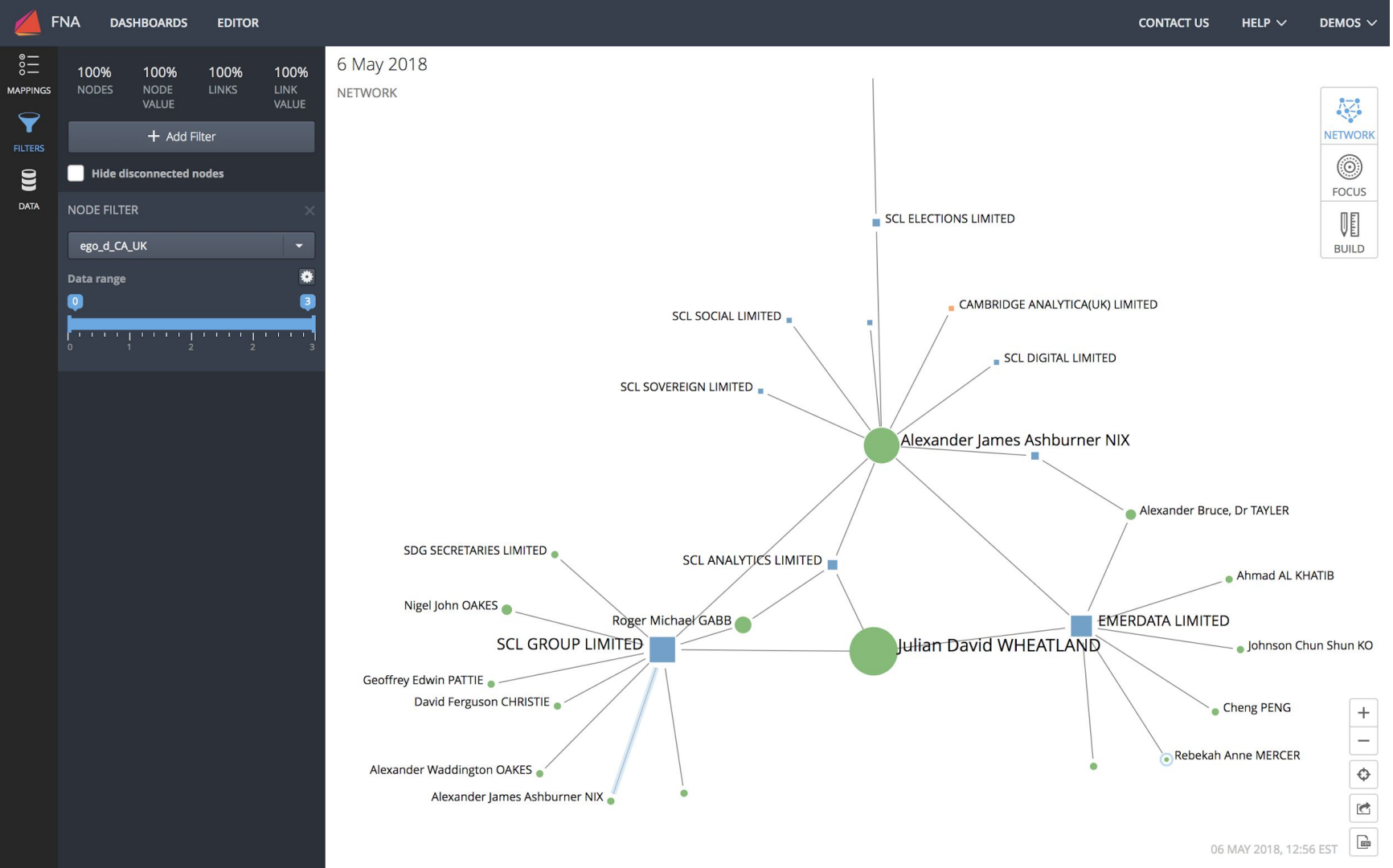
6 May 2018

NETWORK

NETWORK FOCUS BUILD



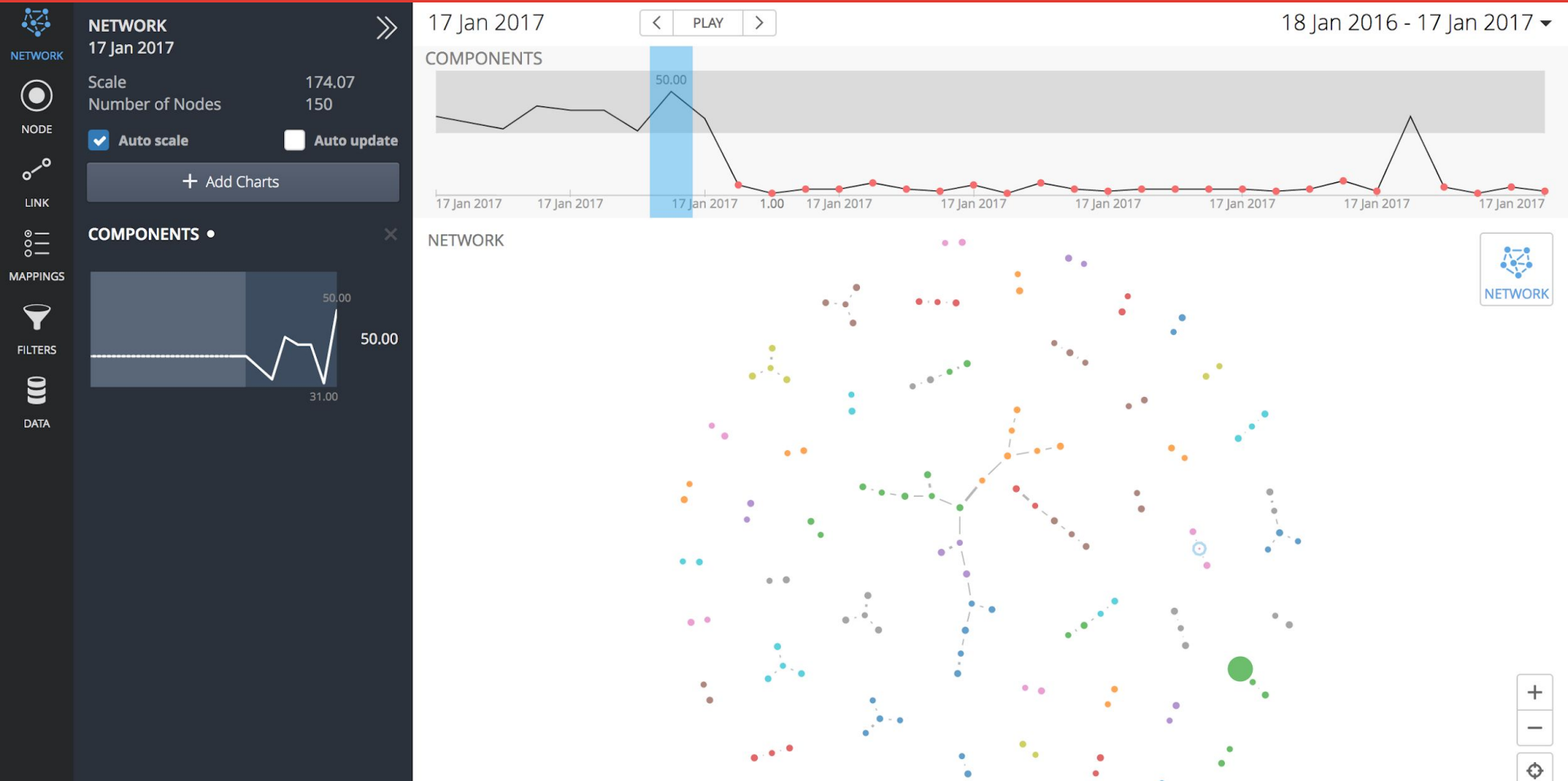
06 MAY 2018, 12:56 EST


FNA
DASHBOARDS
EDITOR
CONTACT US
HELP
DEMOS
MAPPINGS
FILTERS
DATA
100% NODES
100% NODE VALUE
100% LINKS
100% LINK VALUE
Add Filter
Hide disconnected nodes
NODE FILTER
ego_d_CA_UK
Data range
0
3
6 May 2018
NETWORK
NETWORK
FOCUS
BUILD
SCL ELECTIONS LIMITED
SCL SOCIAL LIMITED
CAMBRIDGE ANALYTICA(UK) LIMITED
SCL DIGITAL LIMITED
SCL SOVEREIGN LIMITED
Alexander James Ashburner NIX
Alexander Bruce, Dr TAYLER
SDG SECRETARIES LIMITED
SCL ANALYTICS LIMITED
Ahmad AL KHATIB
Nigel John OAKES
Roger Michael GABB
EMERDATA LIMITED
SCL GROUP LIMITED
Julian David WHEATLAND
Johnson Chun Shun KO
Geoffrey Edwin PATTIE
David Ferguson CHRISTIE
Cheng PENG
Rebekah Anne MERCER
Alexander Waddington OAKES
Alexander James Ashburner NIX
06 MAY 2018, 12:56 EST

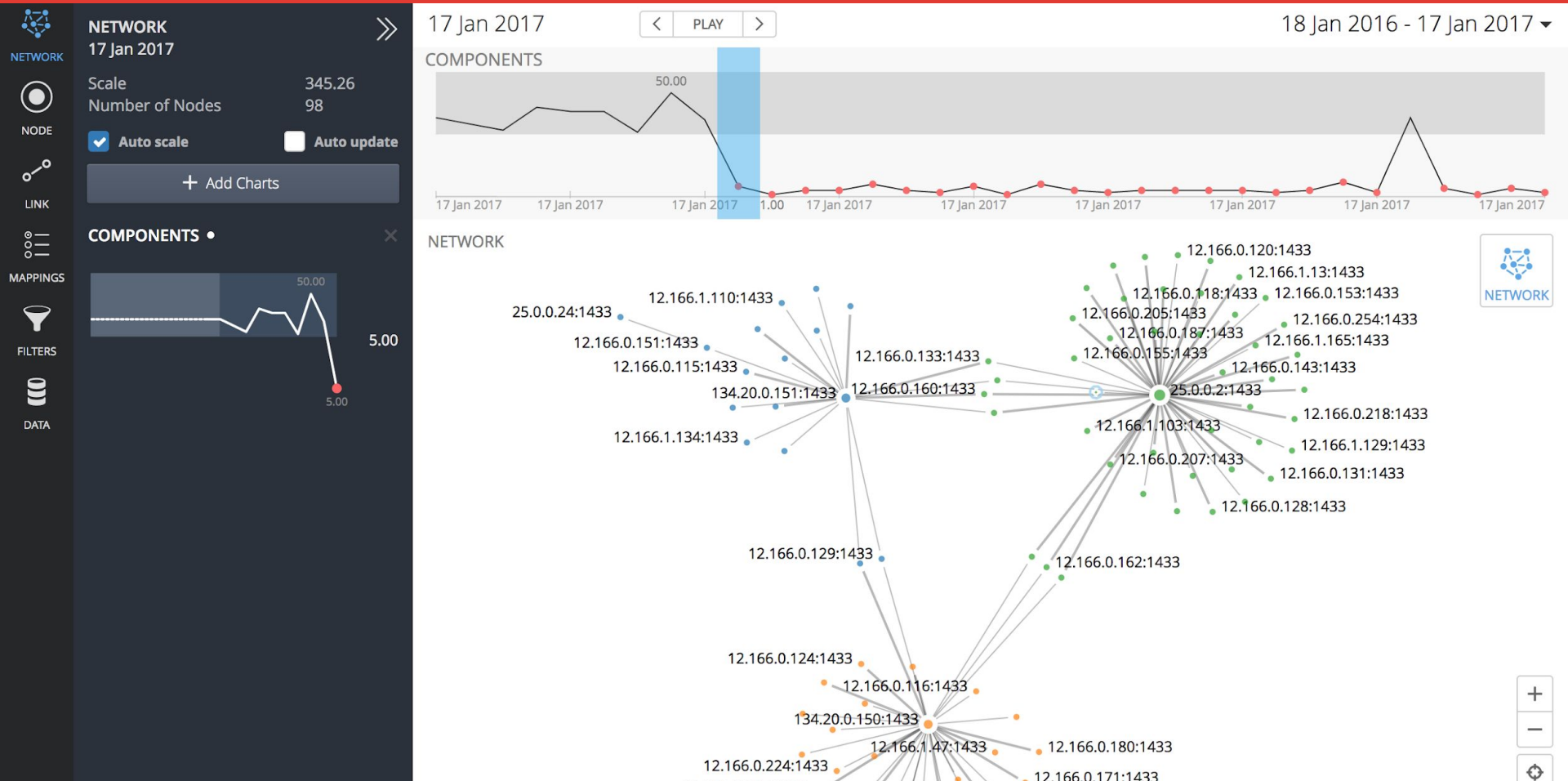
FNA

DDoS Attacks

Detect Anomalies in Cyber Networks in Real-time



Identify Patterns of DDoS attacks



FNA

Supply Chains Networks

www.fna.fi

MAPPINGS

FILTERS

DATA

FNA

LINK PROPERTIES

arc_id	3M CO-SAMSUNG ELECTRON
buyer-country	KR
from_id	3M CO
net_id	2012-01-01
share-of-buyer-cost	0.710
share-of-supplier-r...	2.370
supplier-country	US
to_id	SAMSUNG ELECTRON
value	706.490

1 Jan 2012

NETWORK

15 MAY 2017, 15:27 EST

FNA

Monitoring & Simulating FMIs
& their Members

Use Case: Understanding Interconnectedness

Mapping SWIFTs global payment network

Background
SWIFT message services are used by over 11,000 financial institutions in more than 200 countries. SWIFT was interested what insights could be drawn from the "Big Data" that it collects when transmitting messages between financial institutions.

Objective
Analyse the payment networks created by flows of SWIFT MT103 messages to draw insights about macroeconomic, geo-political and compliance topics.

Insights
Analysis of the SWIFT payment networks revealed a number of insights, including the phenomena of de-risking, payment country blocks relevant for sanctions analysis and how geopolitics shape them, and estimated the cost of the financial crisis at $5 Trillion. The outcome of the research was presented at Sibos 2014 by SWIFT CEO Gottfried Leibbrandt.

SWIFT Institute Research Paper: The global network of payment flows
Research Paper: The Impact of Anti-Money Laundering Regulation on Payment Flows

SWIFT - Big Data Problem

Big data problem: Three billion messages exchanged among banks in 231 countries.

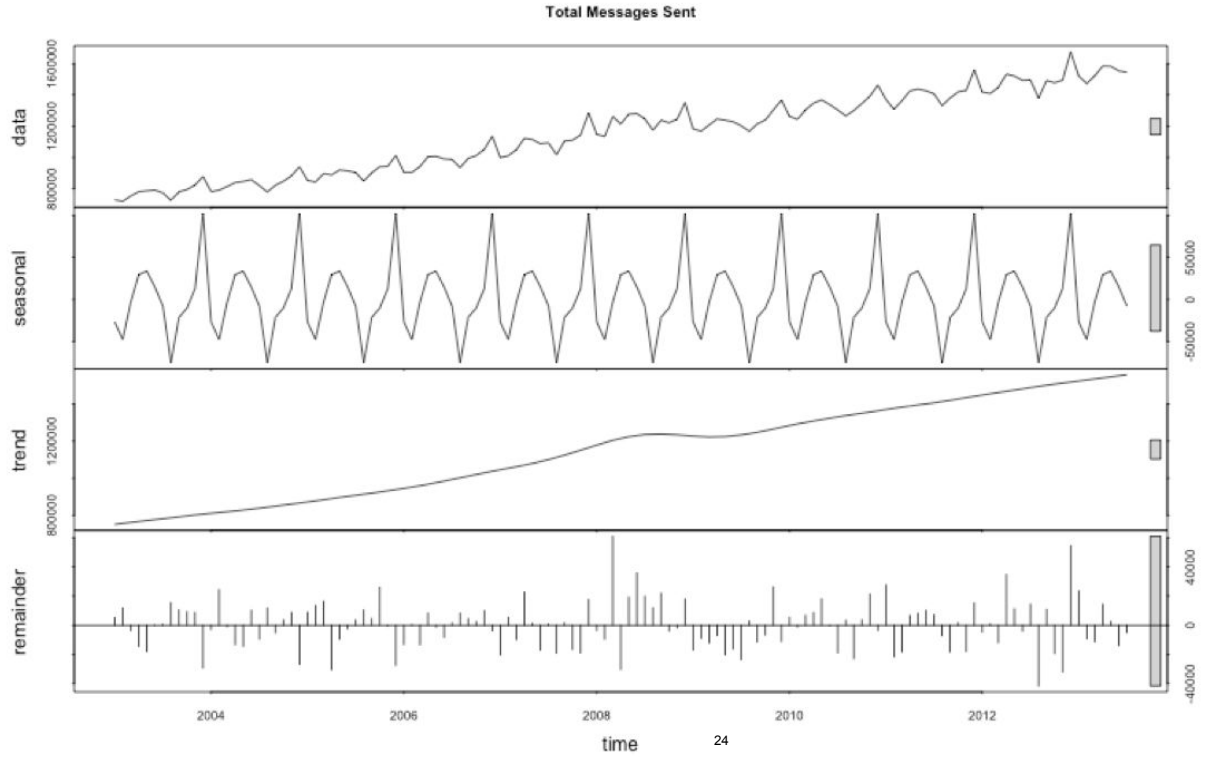
We focus on aggregated links among countries.

Analysis and visualization a challenge. We don’t want to show much information (as on this picture).

FNA

SWIFT - The Cost of Financial Crisis $5tr?

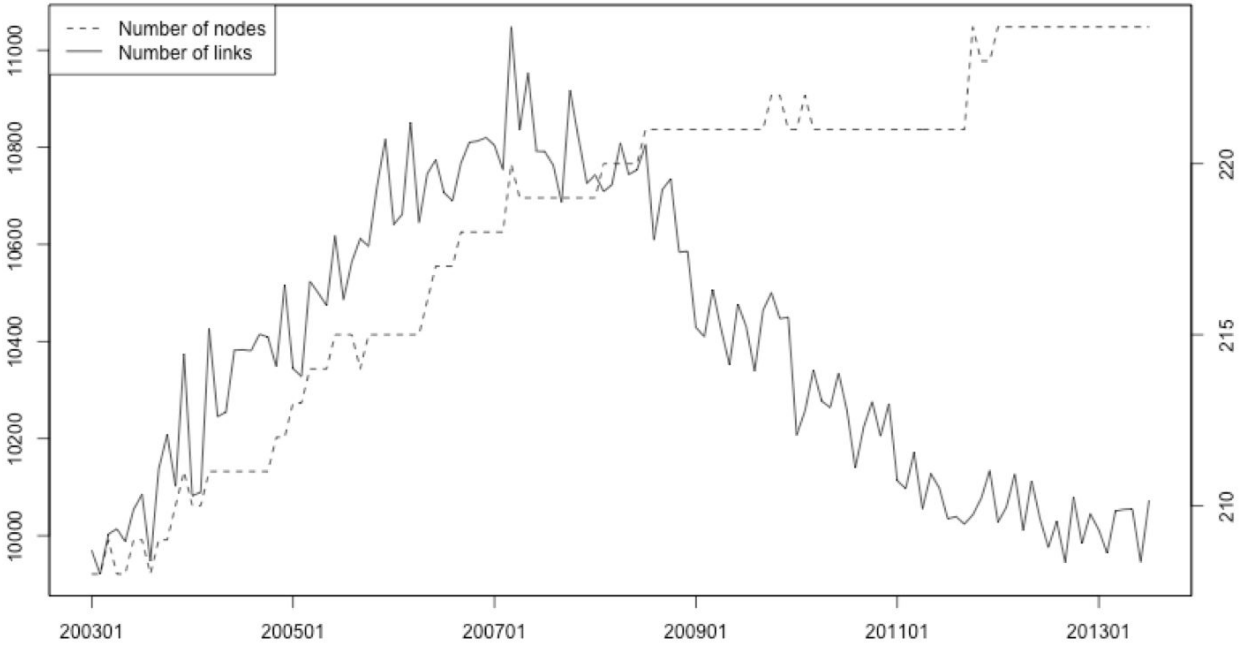
The number of messages is 5.5% lower (post-crisis than they would have been had the pre-crisis trend continued unabated throughout the entire period.



SWIFT - De-risking

Of the 1054 links gained until 2007, in 74% one (or both) were rated as medium or low on the United Nations Human Development Index.

Of the 990 links lost after 2007, 80% involved at least one country listed as an offshore financial center.



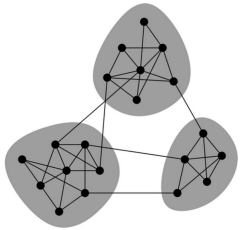
SWIFT - Communities

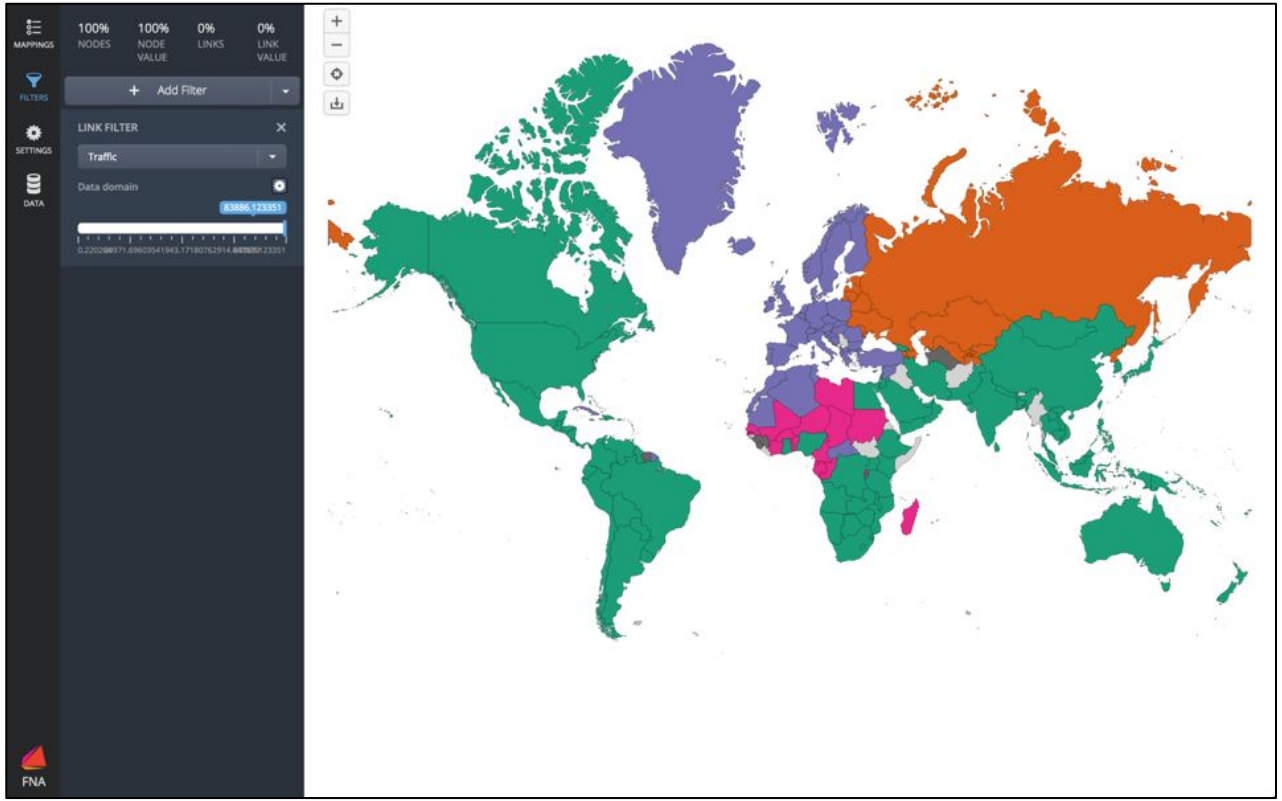
Are there meaningful subgroups among the countries?

Can we group the countries so that messages are sent mostly within groups?

Modularity - measure of concentration of links within communities vs. between communities.

Example Communities

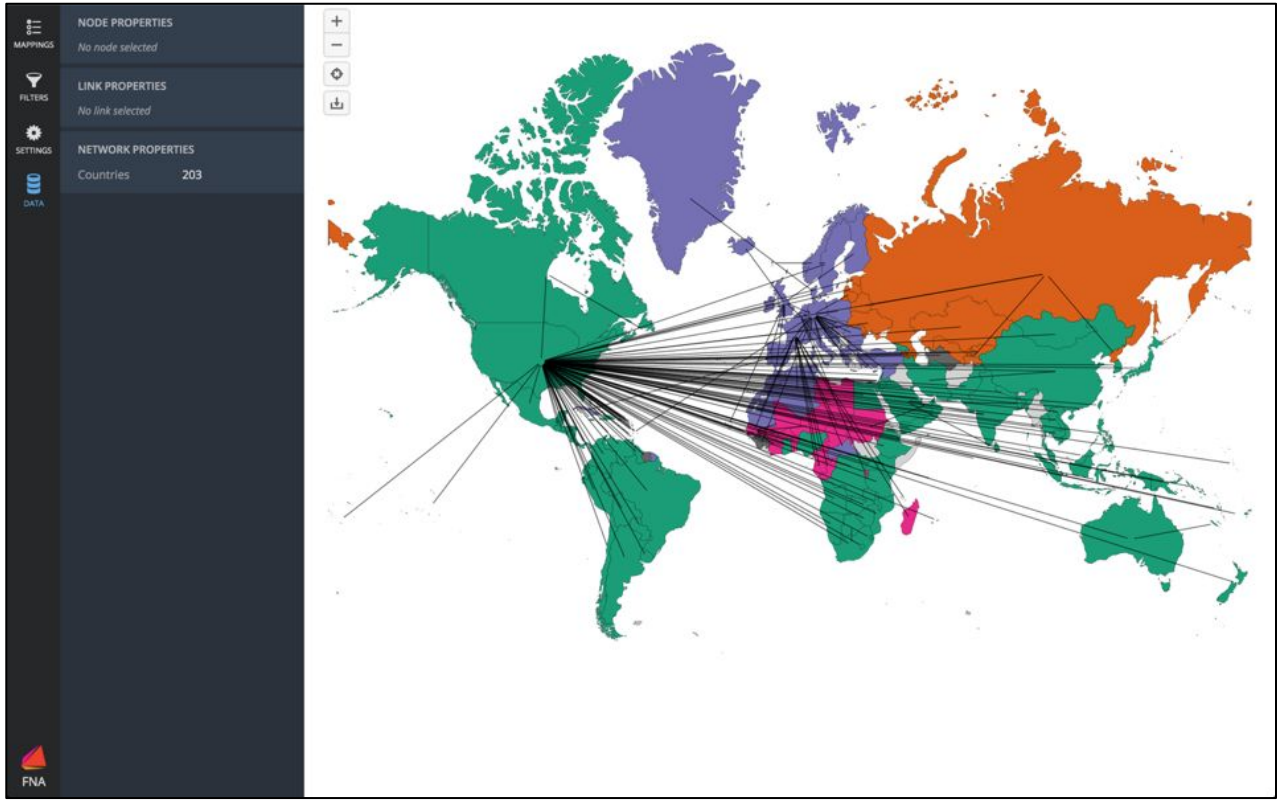




SWIFT - Communities

We overlay the Minimum Spanning Tree showing the strongest links for each country.

We see US, Germany and France as large hubs.



Soramaki and Cook (2014), '[The global network of payment flows]', Journal of Financial Market Infrastructure

Use Case: Monitoring Liquidity and Solvency of FIs

Background

The Central Bank of Colombia has been using balance sheet and regulatory reporting data to understand the liquidity and solvency of participants in the Colombian financial system. However, the analysis is time consuming and the data comes months late.

Objective

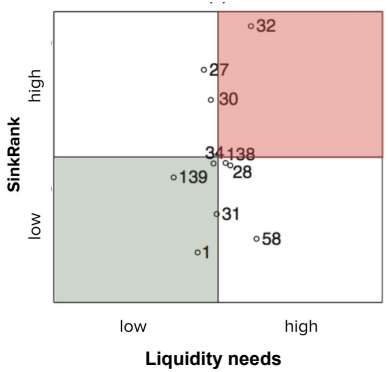
Using network analysis of data from the interbank payment system would allow the Bank to get early warning about risks substantially faster.

Outcomes

Using the FNA Platform, the Bank is now able to monitor its banking system in near real time. Automatic alerts notify the bank of any abnormal behavior in the network. Furthermore, automated stress tests where they fail the two largest participants in the network help to understand the riskiness of the system.

Carlos León et al (2015). Assessing Systemic Importance With a Fuzzy Logic Inference System.

Use Case: Monitoring Liquidity and Solvency of FIs



Background

Bank of Korea, South Korea's central bank, was looking for ways to have early warning about intraday liquidity problems in its systemically important BoK-Wire+ interbank payment system.

Objective

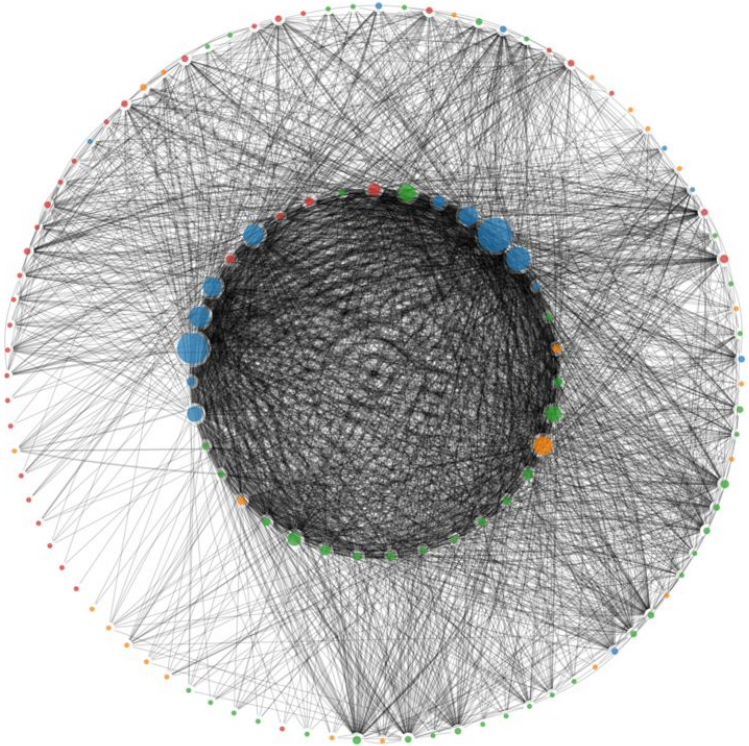
To develop methods to predict the liquidity position of each member in BoK-Wire+ in real-time, as well as measure the importance of member in terms of the liquidity and operational risk a liquidity shortage would cause.

Outcomes

Bank of Korea and FNA developed a framework for identifying bank's liquidity problems in real time and using FNA's SinkRank algorithm to identify most critical banks. The results were published as a research paper.

BoK Research paper: Network Indicators for Monitoring Intraday Liquidity in BOK-Wire+
Journal article: SinkRank: An Algorithm for Identifying Systemically Important Banks in Payment Systems

Predicting Liquidity: Problem

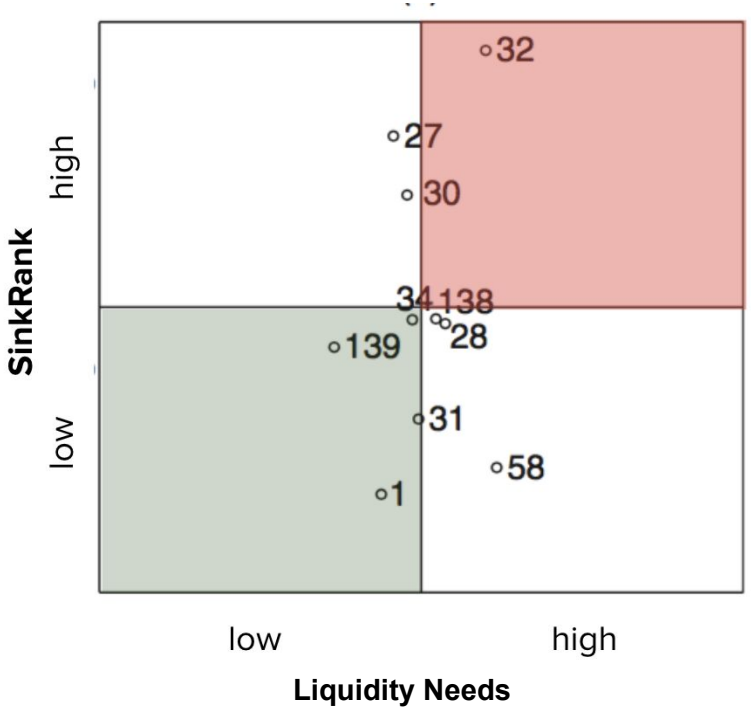


Key issue in payment system is that each bank is dependent on incoming funds to make their own payments.

Objective of this work was to develop measures for ongoing monitoring of systemic risk in payment systems

Baek, Soramäki and Yoon (2014). Network Indicators for Monitoring Intraday Liquidity in BOK-Wire+. Journal of Financial Market Infrastructures 2:3.

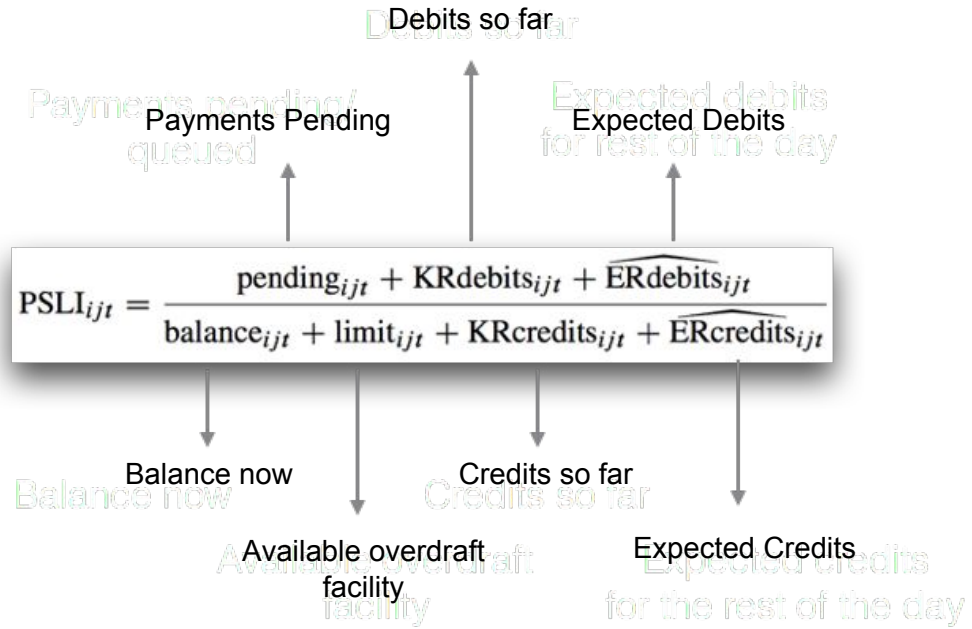
Predicting Liquidity: Framework



Analytics need to be operationalized into a robust and repeatable decision making framework

Predicting Liquidity: PSLI

PSLI (Payment System Liquidity Indicator) is the ratio of projected liquidity demands and projected liquidity supply:



Predicting Liquidity: PSLI

Expected credits and debits are estimated on the basis of a regression model.

The model takes into account the value already settled on the given day, effects related to reserve maintenance and to US holidays and the trade values of bonds and spot exchange.

The model has a good fit.

	Model 1		Model 2		
	Coefficient	*t*		Coefficient	*t*
tue	−0.2939**	−3.09	tue	−0.2692**	−3.49
wed	−0.5075***	−5.05	wed	−0.4879***	−5.67
thu	0.6049***	6.63	thu	0.6054***	7.93
fri	−0.0128	−0.14	—	—	—
reserve_check	−5.2343***	−35.43	reserve_check	−5.2310***	−35.50
us_hol	−1.0795***	−6.53	us_hol	−1.0934***	−6.82
bond	0.0037	0.87	—	—	—
fx	0.0001	0.04	—	—	—
_Ireceiver_1	3.0615***	14.87	_Ireceiver_1	3.1743***	31.34
_Ireceiver_27	12.0550***	38.07	_Ireceiver_27	12.1676***	130.47
_Ireceiver_28	6.7873***	28.69	_Ireceiver_28	6.9051***	80.15
_Ireceiver_30	13.5095***	59.61	_Ireceiver_30	13.6257***	87.87
_Ireceiver_31	2.8790***	34.04	_Ireceiver_31	2.9899***	32.92
_Ireceiver_32	19.3134***	56.84	_Ireceiver_32	19.4082***	89.10
_Ireceiver_34	8.2016***	14.30	_Ireceiver_34	8.3231***	118.77
_Ireceiver_58	2.3454***	68.63	_Ireceiver_58	2.4588***	26.63
_Ireceiver_138	7.6201***	42.56	_Ireceiver_138	7.7360***	87.08
_Ireceiver_139	6.0048***	11.62	_Ireceiver_139	6.1261***	56.87
Number of obs = 2 480			Number of obs = 2 490		
$F(18,2462) = 4\,159.70$			$F(15,2475) = 5\,031.22$		
Prob > F = 0.0000			Prob > F = 0.0000		
R-squared = 0.9682			R-squared = 0.9682		
Adj R-squared = 0.9679			Adj R-squared = 0.9681		

*, ** and *** represent statistical significance at the 5%, 1% and 0.1% levels, respectively.

Measuring Importance: SinkRank

Payments move liquidity in the network.

Payments take place on links at some given frequency that can be measured (eg based on historical or projected flows).

We are concerned on operational failures.The sink can receive payments but cannot send any.

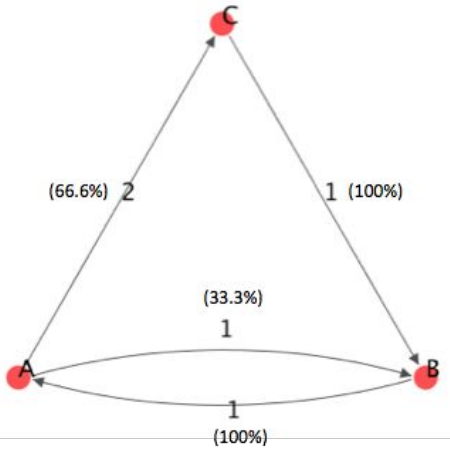
Example:

Let's start by considering one unit of liquidity that is moved by payments in a simple system of three banks.

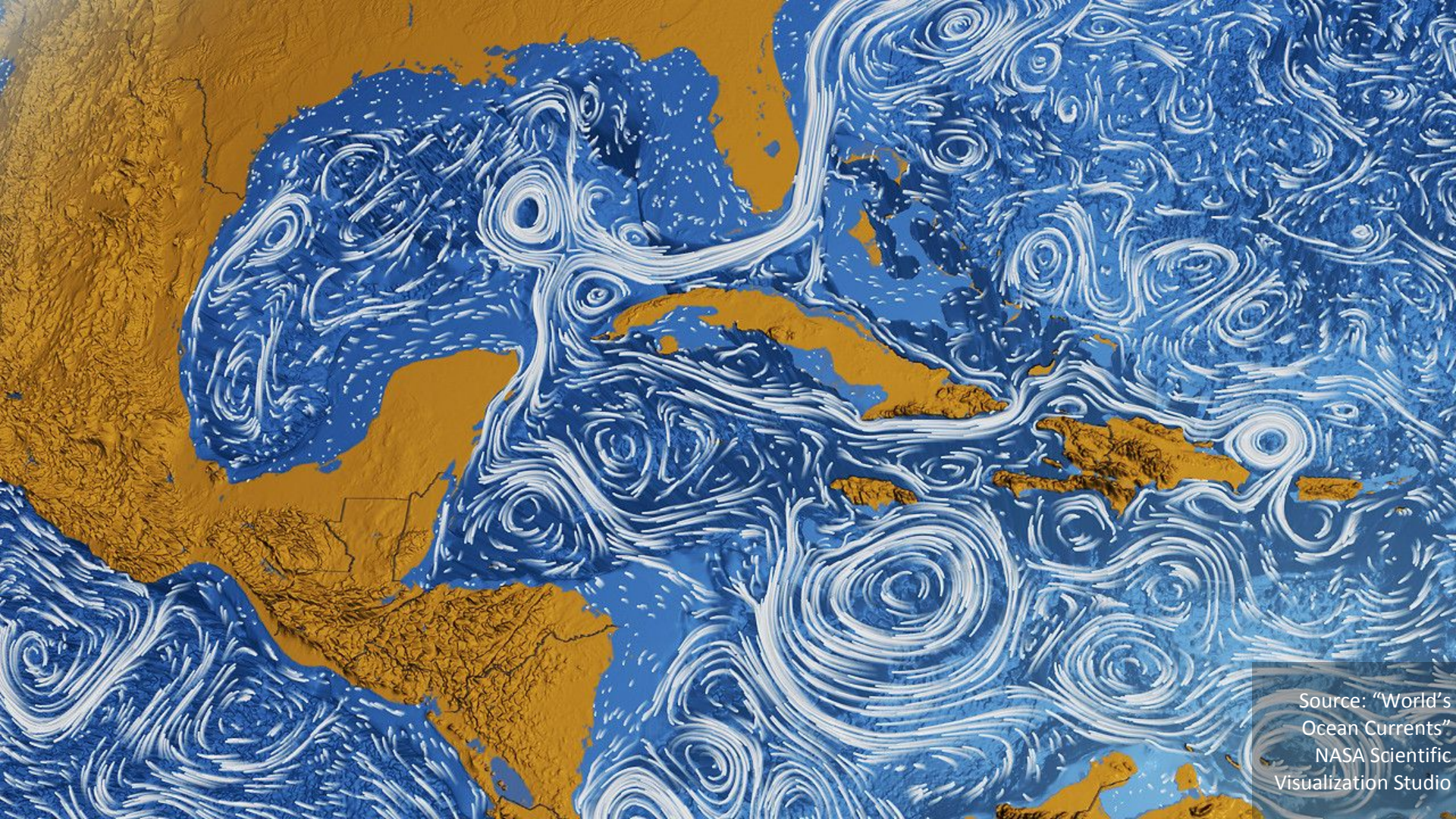
At the time of analysis, the unit of liquidity can be at either A, B or C.

What is the distance of the unit to the different 'sink nodes'?

Soramaki and Cook (2013) SinkRank: An Algorithm for Identifying Systemically Important Banks in Payment Systems



To A	From B	1
	From C	2
To B	From A	$\frac{2}{3}\cdot 2+\frac{1}{3}\cdot 1=\frac{5}{3}$
	From C	1
To C	From A	$\sum_{i=1}^{\infty}(2i-i)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^{i-1}=2$
	From B	$\sum_{i=1}^{\infty}(2i)\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^{i-1}=3$



Source: “World’s Ocean Currents” NASA Scientific Visualization Studio

Measuring Importance: SinkRank

SinkRank is suited for Predictive Modeling

Given an observed distribution of liquidity, and a historical pattern of payment flows

- What is the distribution if bank A has an operational disruption at noon?
- Who is affected first?
- Who is affected most?
- How is Bank C affected in an hour?

Valuable information for decision making

- Crisis management
- Participant behavior

FNA

Using Network Simulations to Design FMIs

What are Simulations

Methodology to understand complex systems – systems that are large with many interacting elements and or non-linearities (such as payment systems)

In contrast to traditional statistical models, which attempt to find analytical solutions

Usually a special purpose computer program is used that takes granular inputs, applies the simulation rules and generates outputs

Take into account second rounds effects, third round, …

Inputs can be stochastic or deterministic. Behavior can be static, pre-programmed, evolving or co-learning

Short History of FMI Simulations

1997 : Bank of Finland

Evaluate liquidity needs of banks when Finland's RTGS system was joined with TARGET

2000 : Bank of Japan and FRBNY

Test LSM features for BoJ-Net/Fedwire

2001 - : CLS approval process and ongoing oversight

Test CLS risk management

Evaluate settlement' members capacity for pay-ins

Understand how the system works

Since: Bank of Canada, Banque de France, Nederlandsche Bank, Norges Bank, TARGET2, and many others

2010 - : Bank of England, CHAPS

Evaluate alternative liquidity saving mechanisms

Use as platform for discussions with banks

Agent Based Modeling

Analytical models need to make many simplifying assumptions.

Problem with static simulations based on historical records is that behavior of banks is not taken into account.

This behavior may have material impact on results in most simulation questions, eg:

- When system features are changed
- In stress situations
- As a reaction to other behavioral changes

-> Agent Based Modeling

Agent Based Models

Each agent has a set of rules that define its behavior
-> system level emergent behavior

Choices

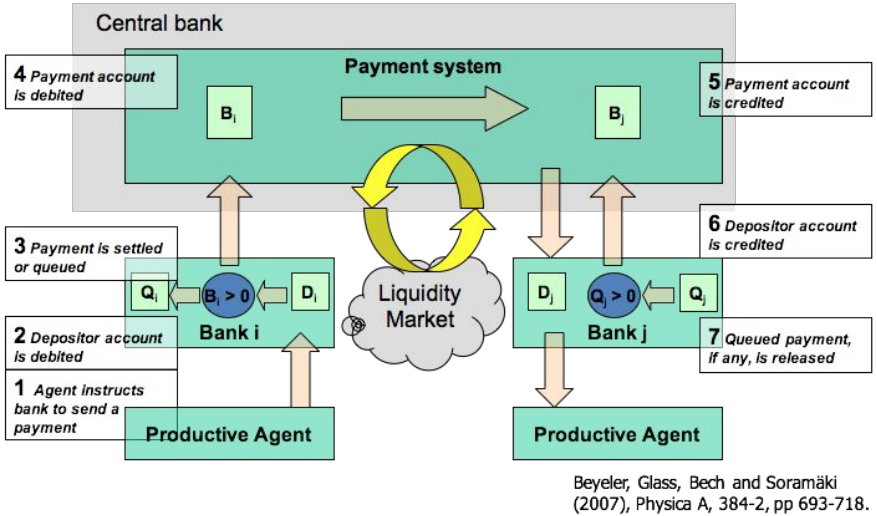
- design of rules
- homogeneous vs heterogeneous agents
- static vs learning agents

Pros

- ability to model complex behaviors
- flexible and realistic
- real systems are sensitive to details of implementation

Cons

- time consuming to set up
- need many input parameters
- results very sensitive to modeling assumptions

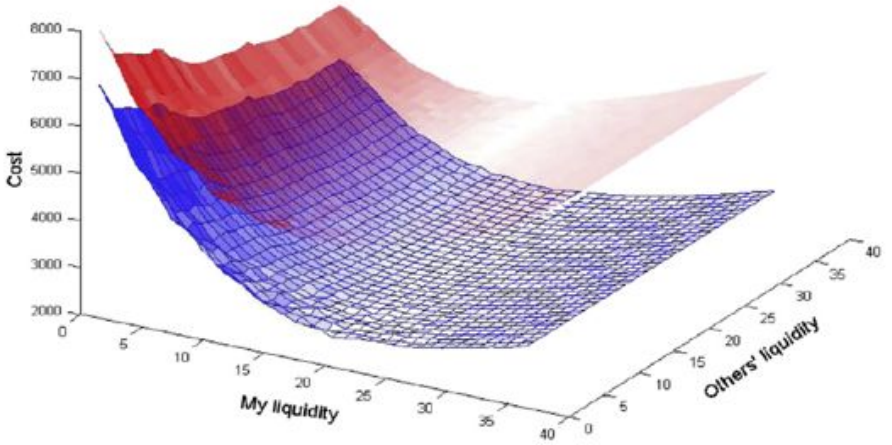


Beyeler, Glass, Bech and Soramäki (2007), Physica A, 384-2, pp 693-718.

Agent Based Models

Existing literature very short

- Galbiati and Soramäki (2008, 2010)
- Arciero et al (2009)
- McLafferty-Denbee (2013)
- Soramäki and Cook (2015)



Results

- Behavior has material impact on results
- Behavior increases delays (or moves away from social liquidity/delay optimum)

Questions

- Money market model
- One vs multiperiod, learning vs fixed populations
- Which payments are discretionary / known
- What is the cost of liquidity/delay tradeoff
- Human vs machine behavior

FNA

Data Needs

Data Needs

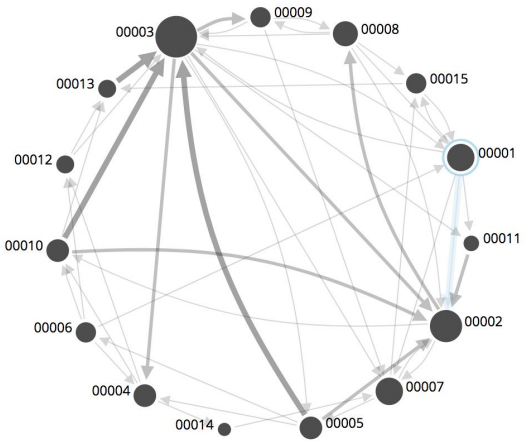
Historical transaction data

- From interbank payment systems
- At minimum: date, time, sender, receiver, value
- More data on type of payment, economic purpose, second tier (if any), type of institution, etc. useful

Representative transaction data

- Based on aggregates or sampling of real data
- Based on a network model (defining bilateral flows)
- Assumptions about:
  - Timing of payments
  - Value distribution
  - Correlations (eg do larger participant send larger payments)
- System stability (net flows over longer times)

FNA R&D: Generating Representative Transaction Data




```
date,time,value,sender,receiver
2017-06-03,08:03:36,5,A,B
2017-06-03,08:06:12,7,A,C
2017-06-03,09:13:35,11,D,A
2017-06-03,11:19:26,1,C,B
2017-06-03,13:25:11,4,B,D
```

Background

Real transaction data held by FMI's and Banks is highly confidential and hard to get access to. Also as historical records, it cannot be used as input data in simulations about future infrastructures that may process very different flows.

Method

FNA has developed and vetted in several client projects a method for generating representative transaction data that contains all known network and statistical properties of the real transaction data.

Outcome

The cost of simulations is lower and the speed at which projects can be completed is higher - lowering the entry barriers to start simulations. Often results with representative data prove the value of the simulations and real data can be used for sensitivity analysis.

Summary of Research (slides)

FMI Simulations: Application Areas

1. Evaluate Changes in Environment
2. Stress Testing & Scenarios
3. Payment System Design
4. Model Validation
5. Monitoring

FNA

Framework for evaluating trade-off between liquidity and delay

Motivation

Settlement in RTGS consumes large amounts of cash

Cash/liquidity is not free

Customers' increasing demands for faster payments means delays cost too

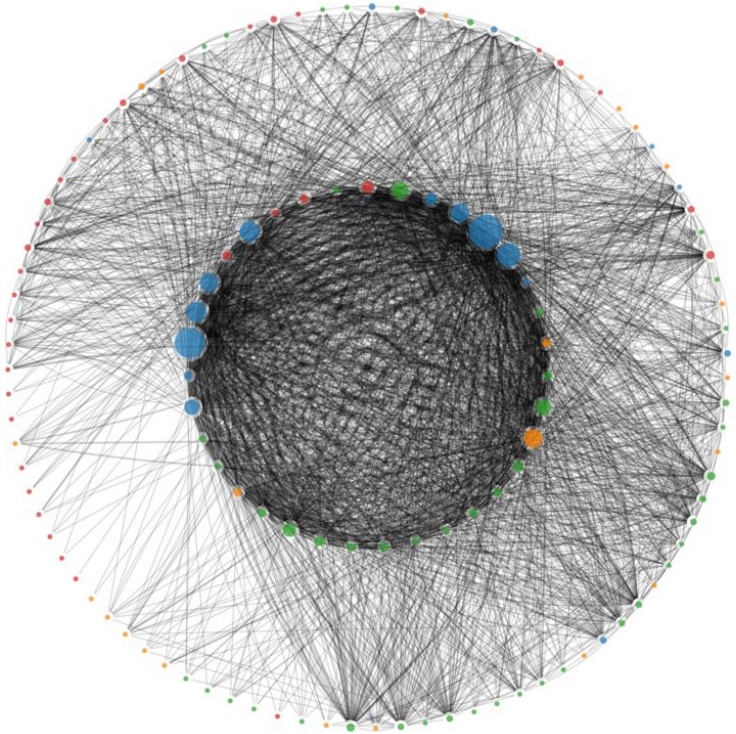
The tradeoff is not going away even with Blockchain

There is no natural co-operative outcome

A complex system, hard to analyse

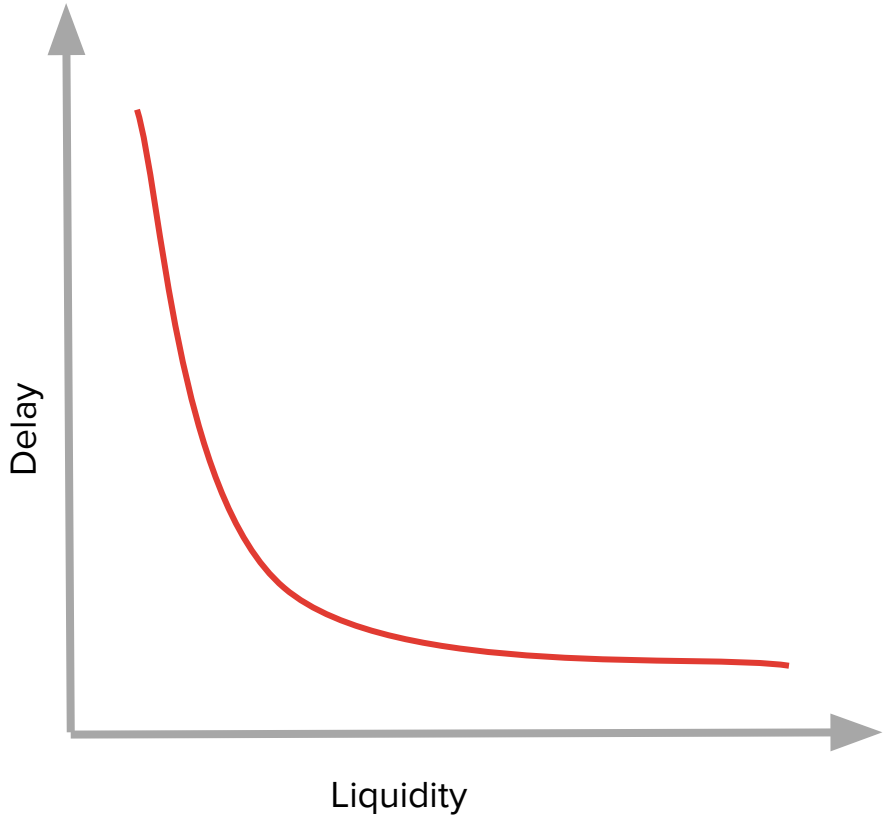
Bottom line impact

Baek, Soramäki and Yoon (2014).
J. of Financial Market Infrastructures



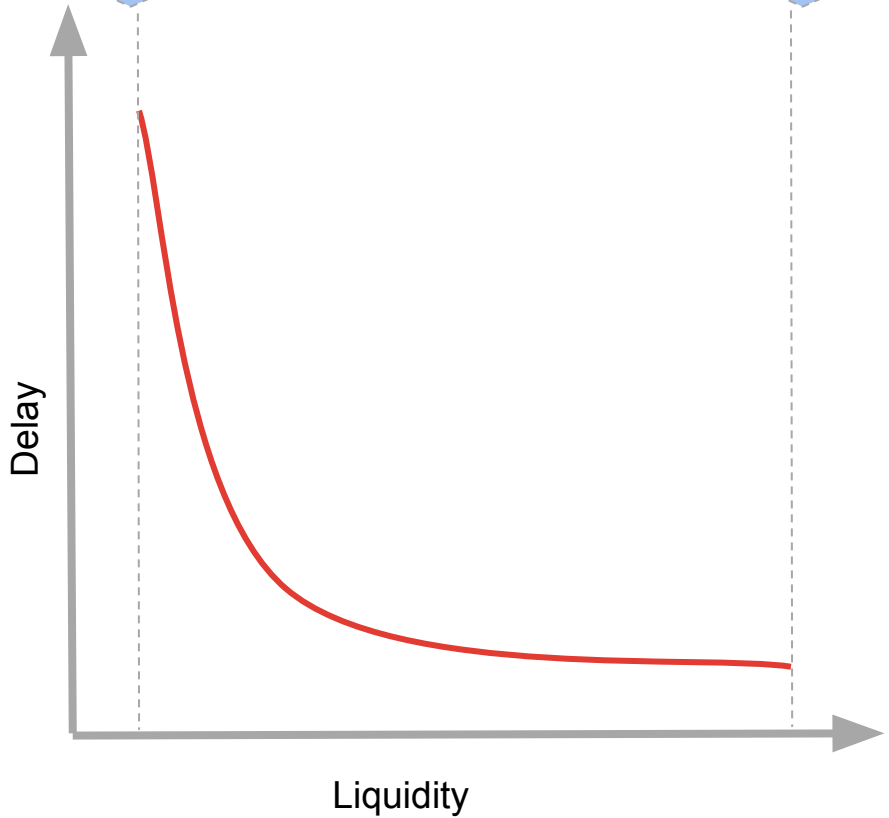
BoK-Wire+

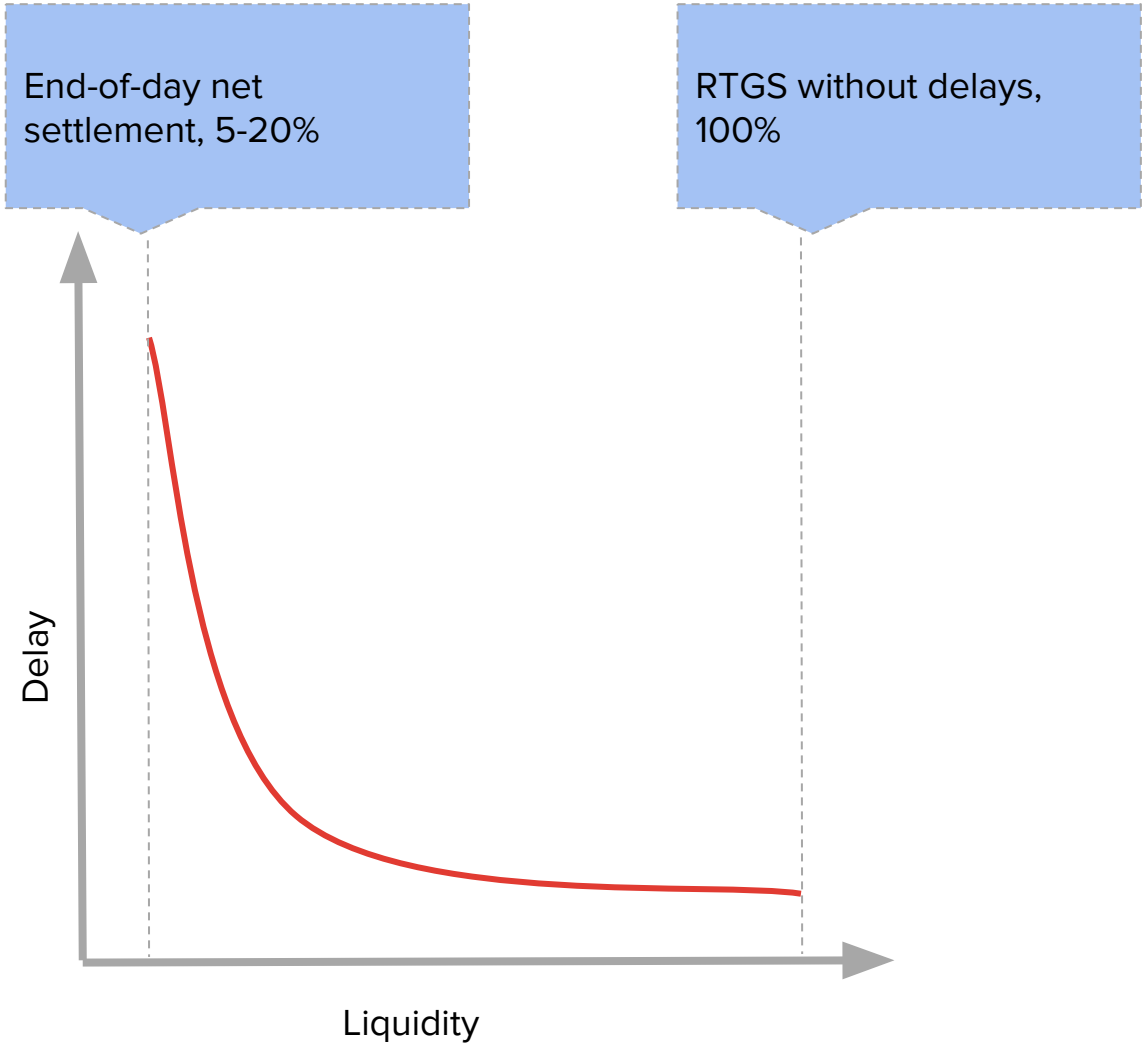
Koponen and Soramäki (1998).
BoF monograph.



There is an amount of liquidity each bank must have to complete settlement

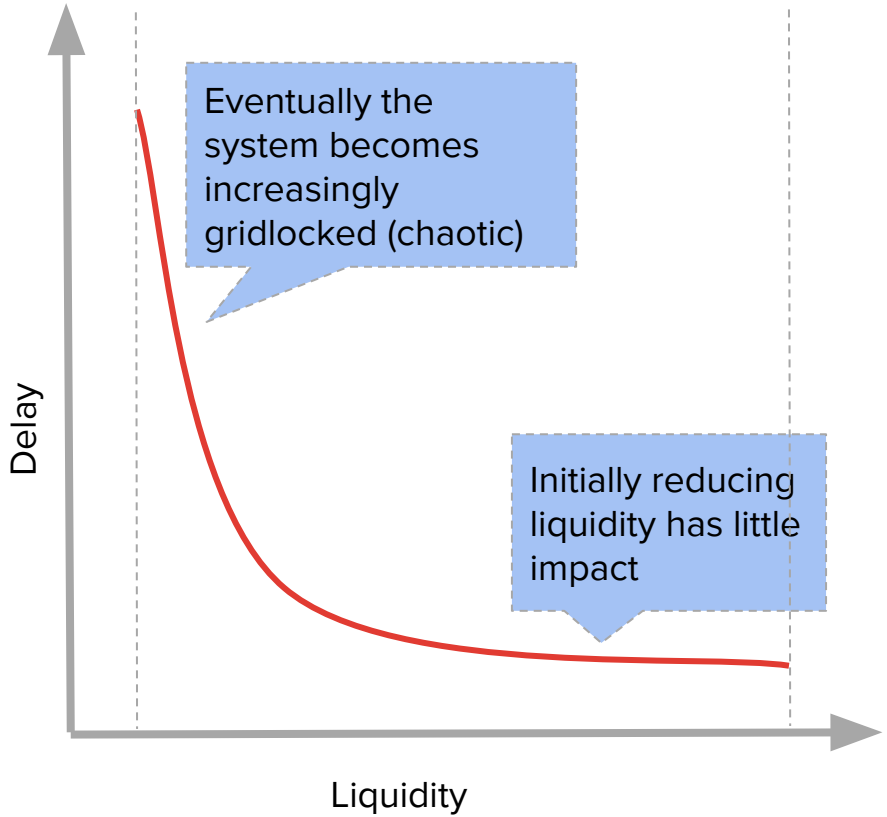
And another amount above which adding more has not impact



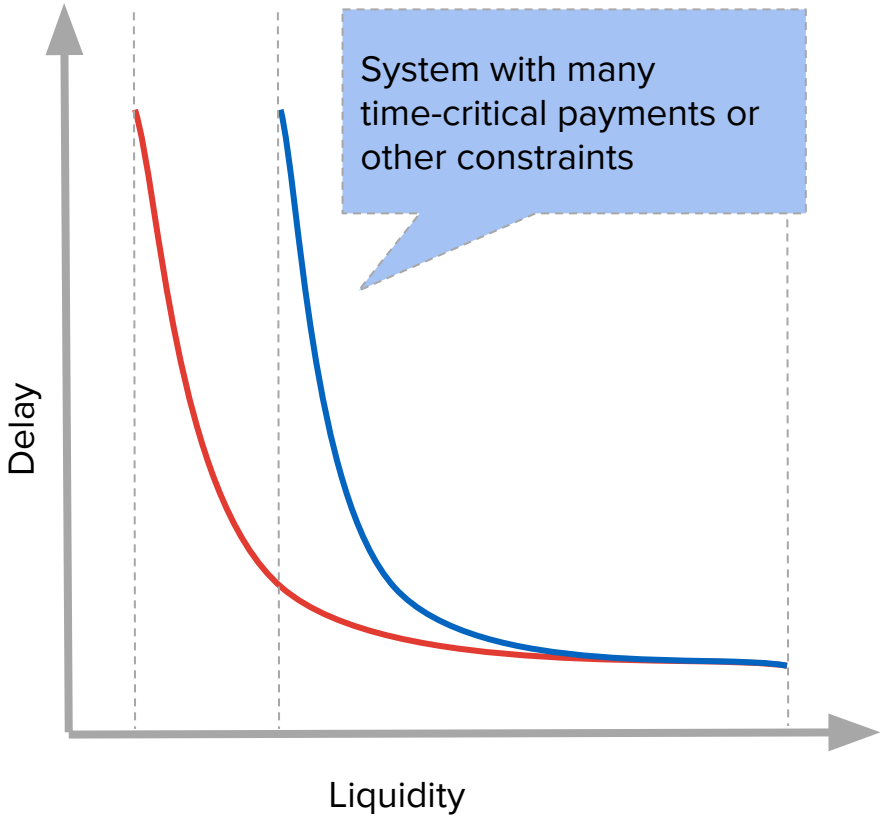

End-of-day net settlement, 5-20%
RTGS without delays, 100%
Delay
Liquidity

A convex shape for trade-off

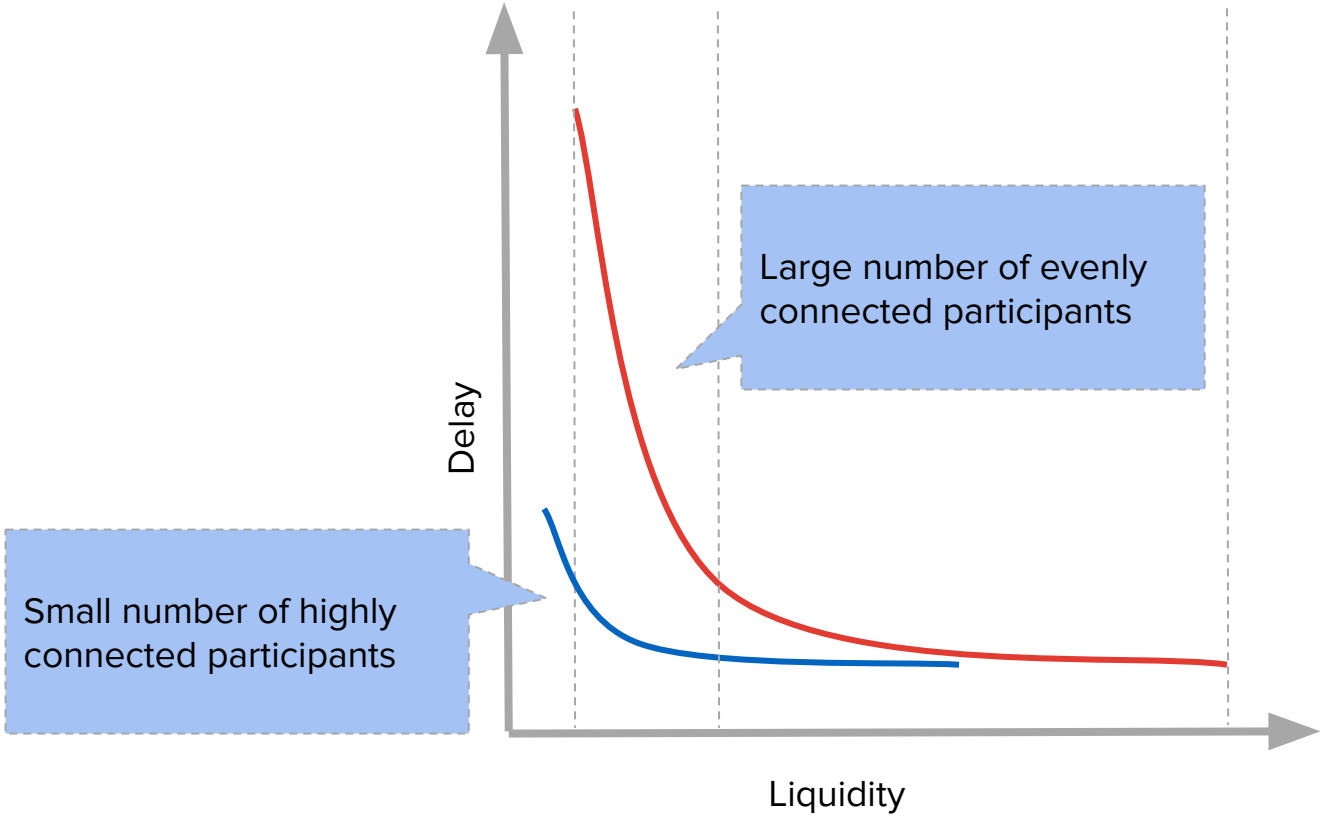
Bayeler, Glass, Bech, Soramäki (2007). Physica A.



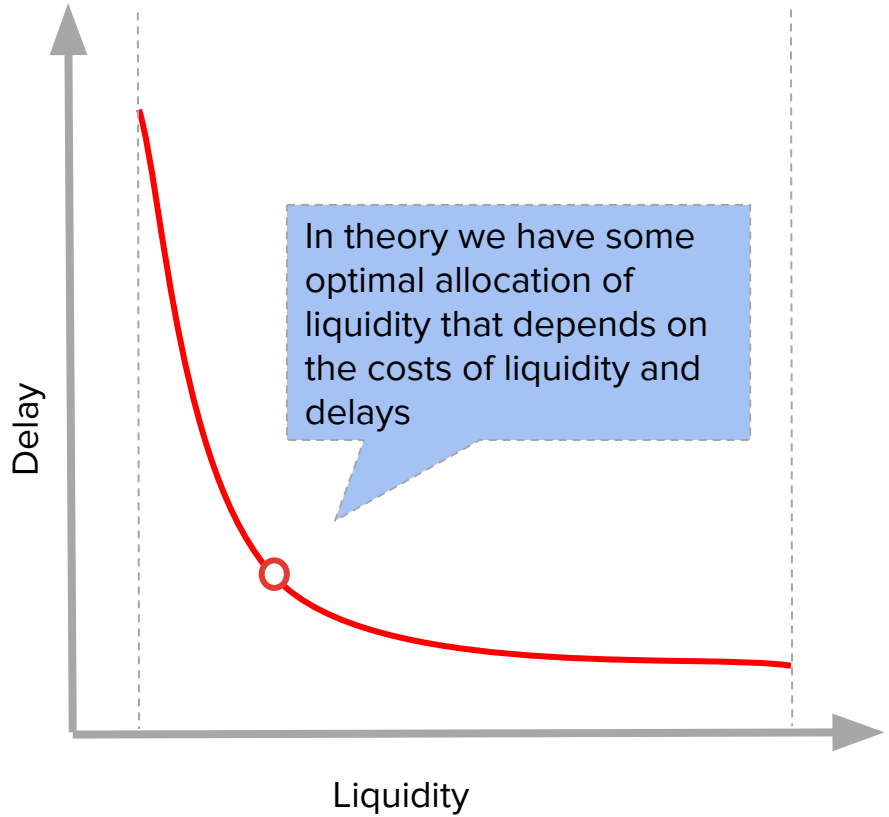
Institutional aspects matter



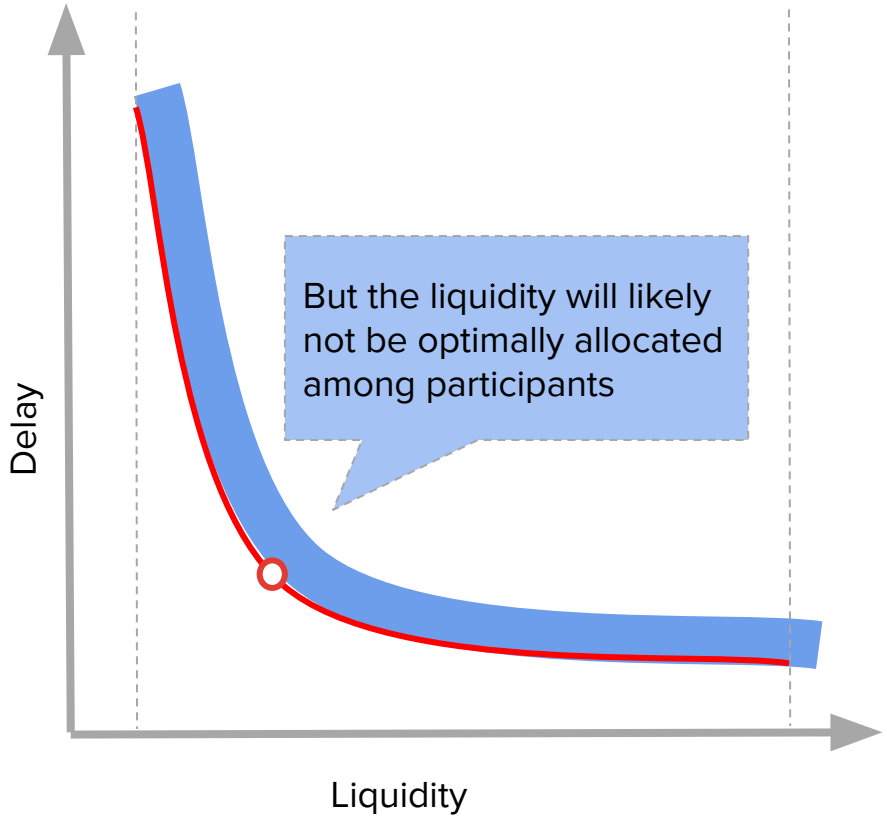
Network topology matters



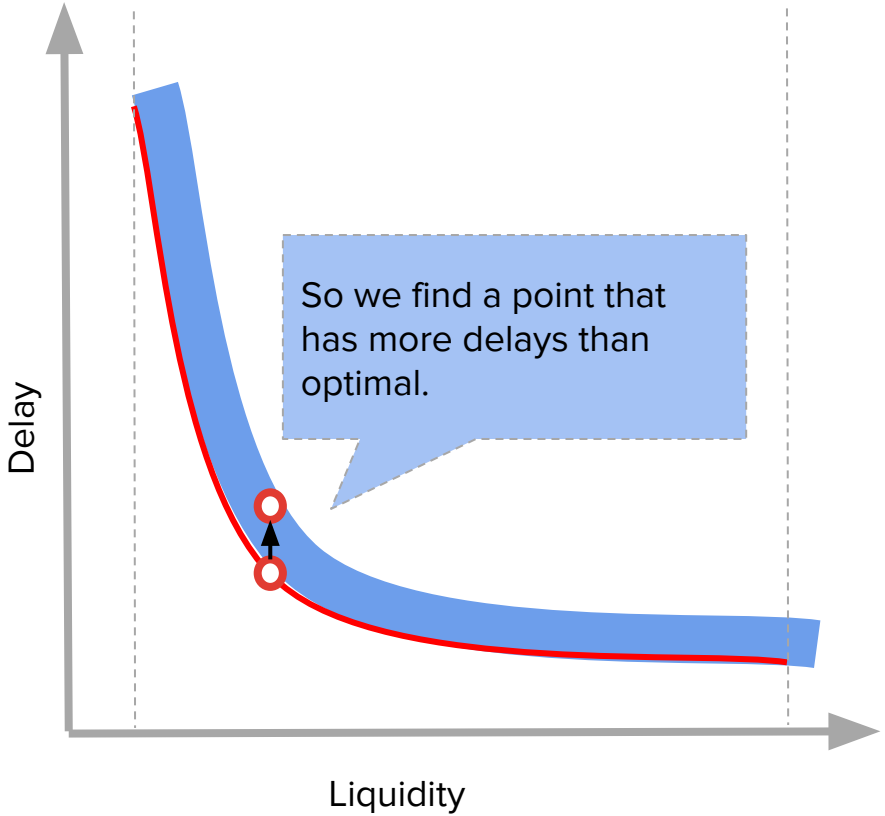
Where should we be on this curve?



Liquidity distribution matters

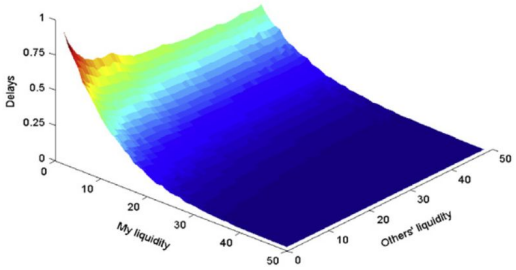


Liquidity distribution matters

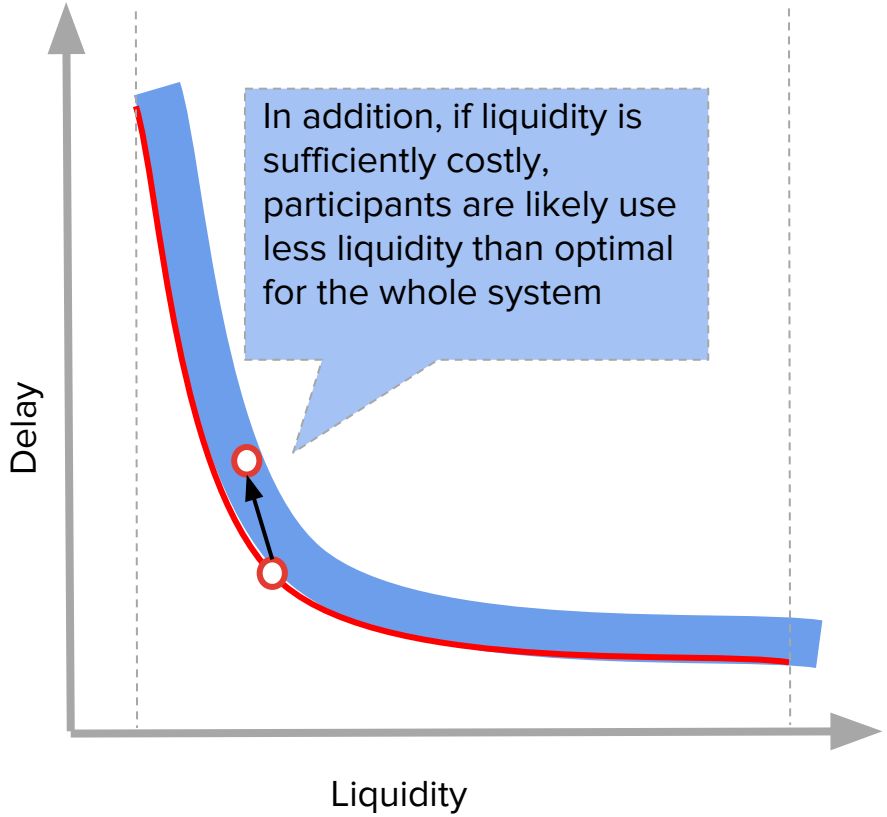


Incentives matter

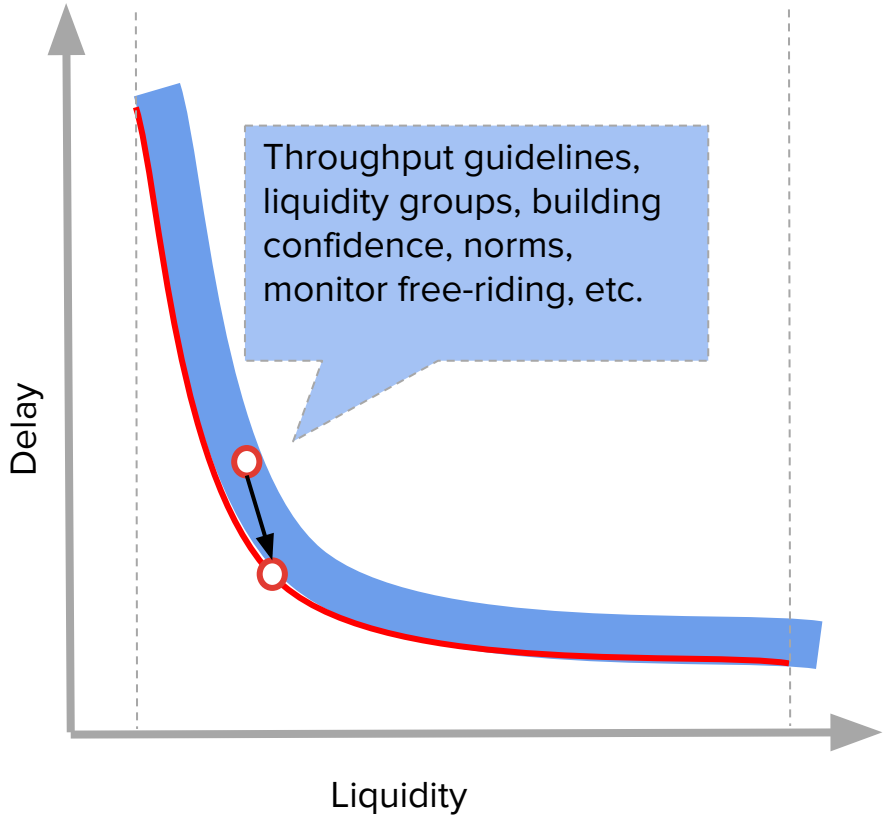
Galbiati and Soramaki (2011). J. of Econ. Dynamics and Control



Bech and Garratt (2003, 2006)

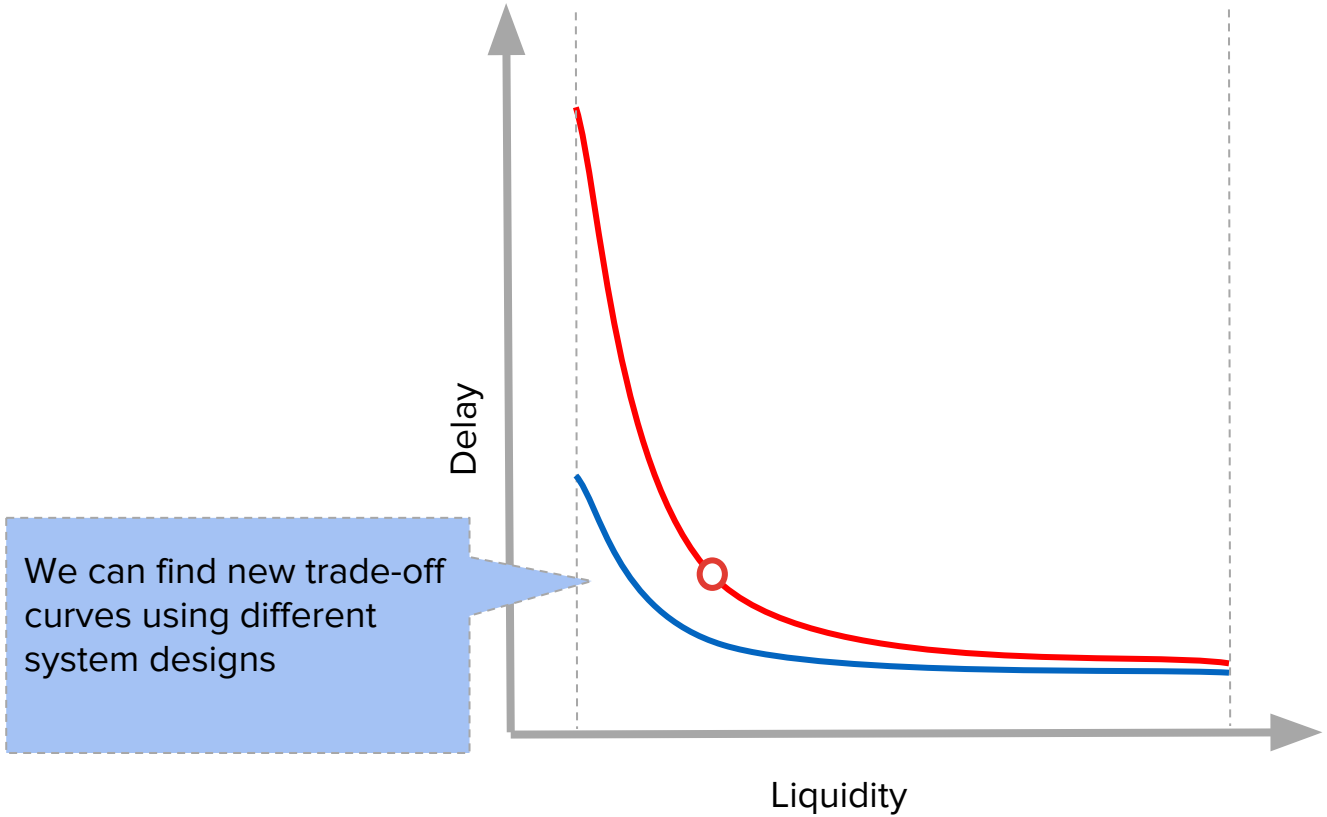


How do we foster co-operation?

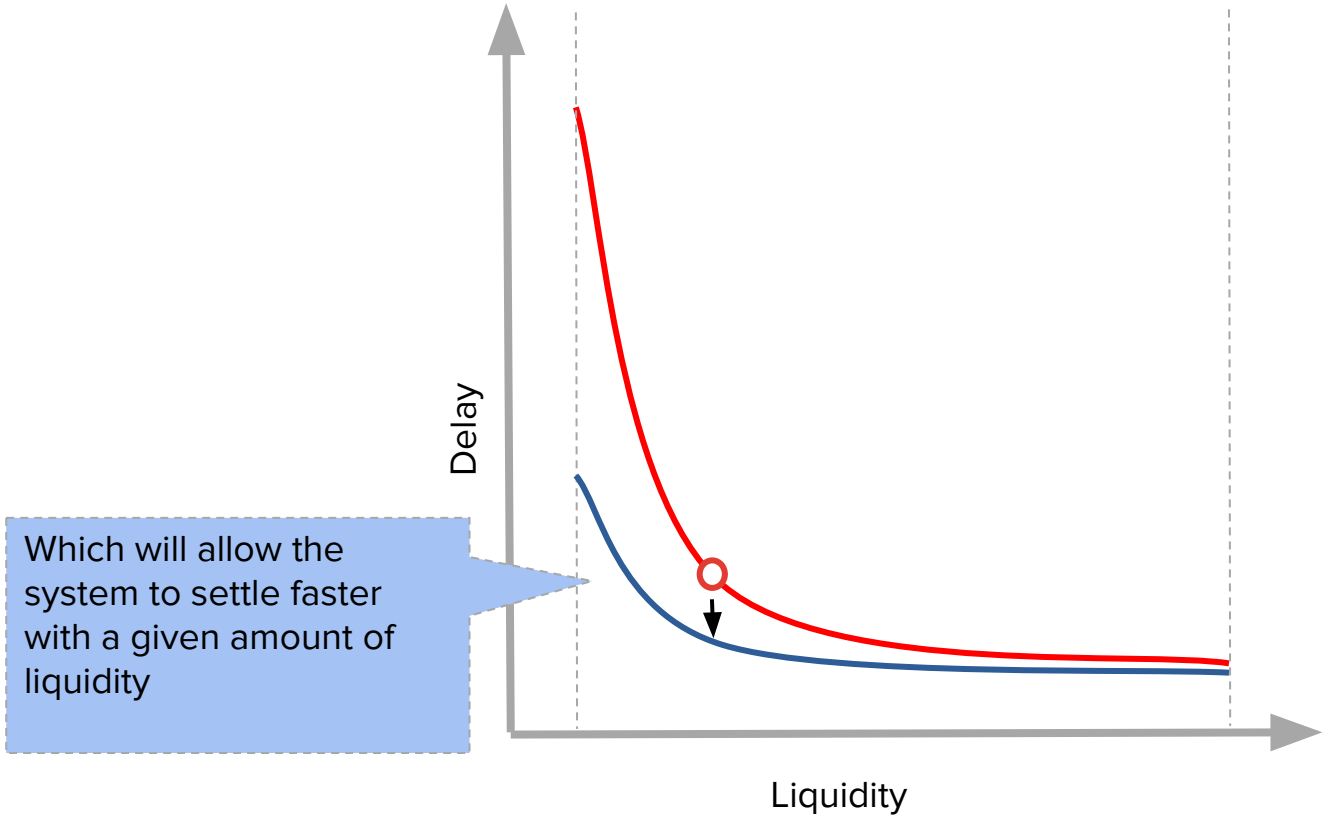


Liquidity Saving Mechanisms (LSMs)

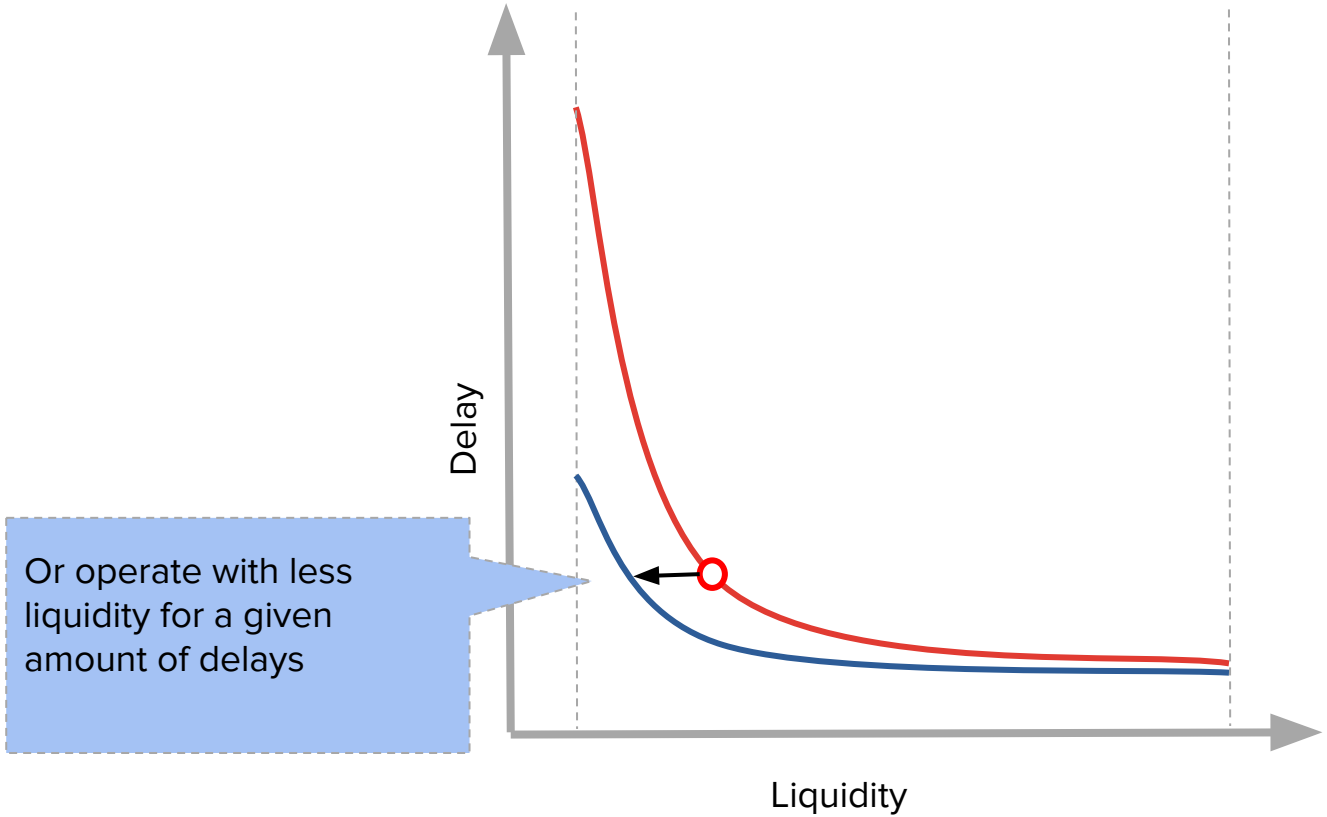
Leinonen and Soramaki (2011).
Bank of Finland WP



Liquidity Saving Mechanisms (LSMs)

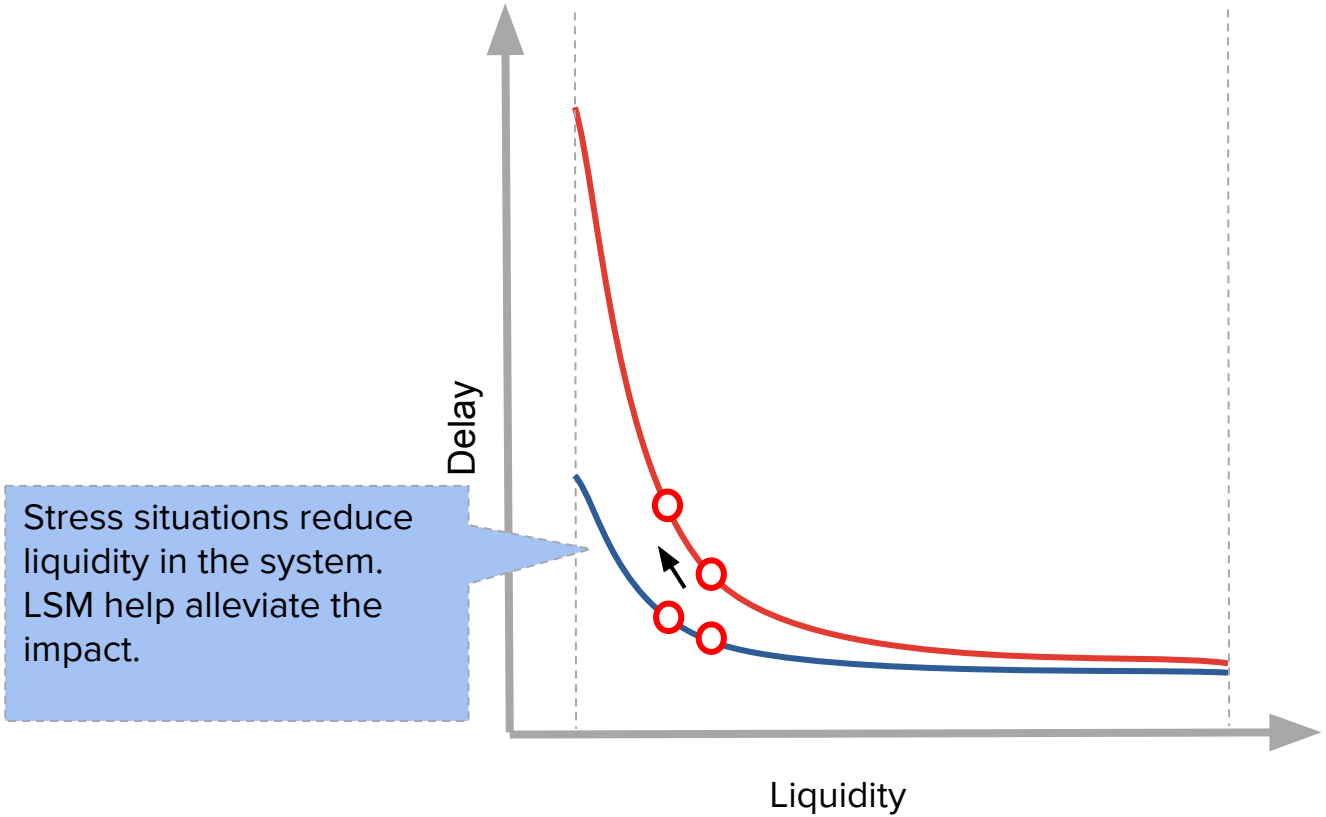


Liquidity Saving Mechanisms (LSMs)



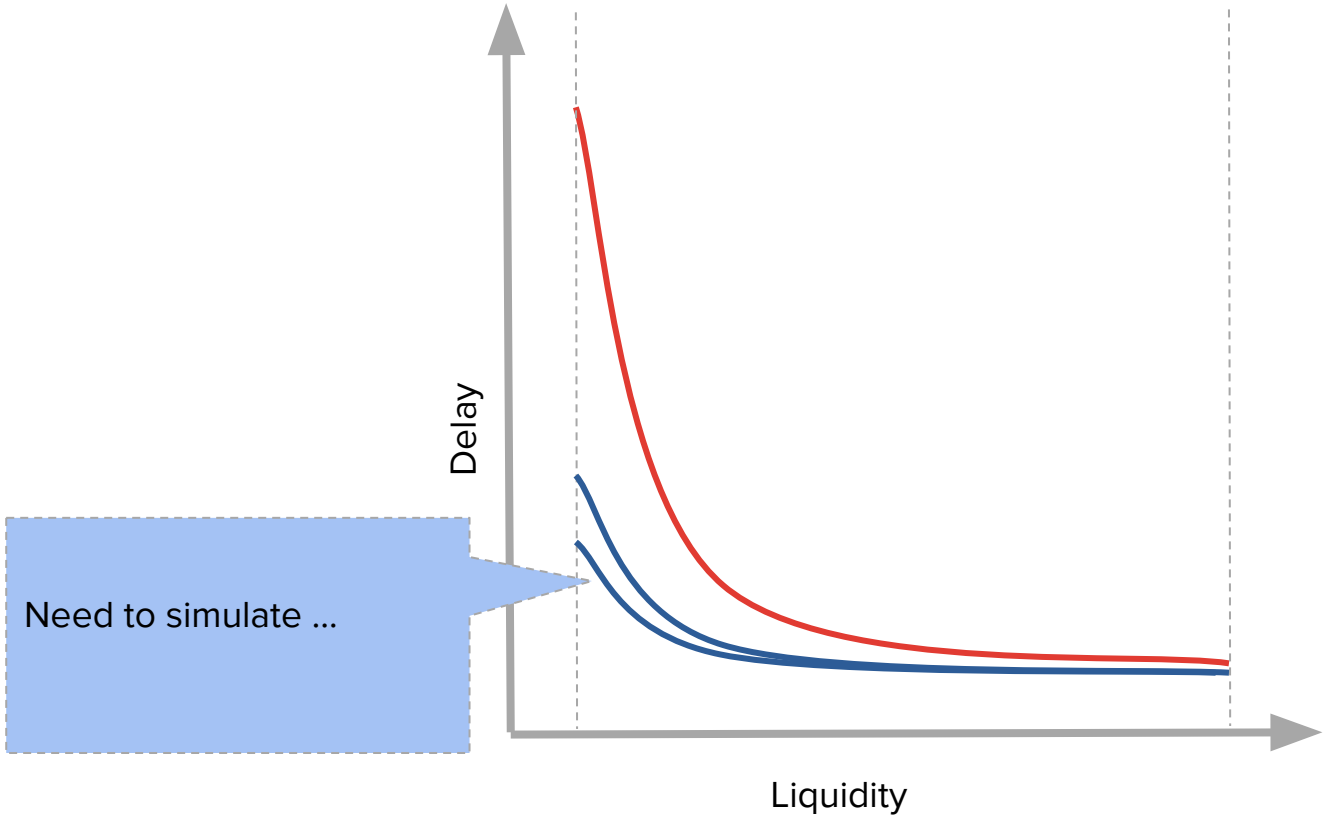
How about stress situations?

Bech and Soramaki (2002). E-money & Payment Systems Review



Stress situations reduce liquidity in the system. LSM help alleviate the impact.

Which LSM is the best?



Bypass FIFO

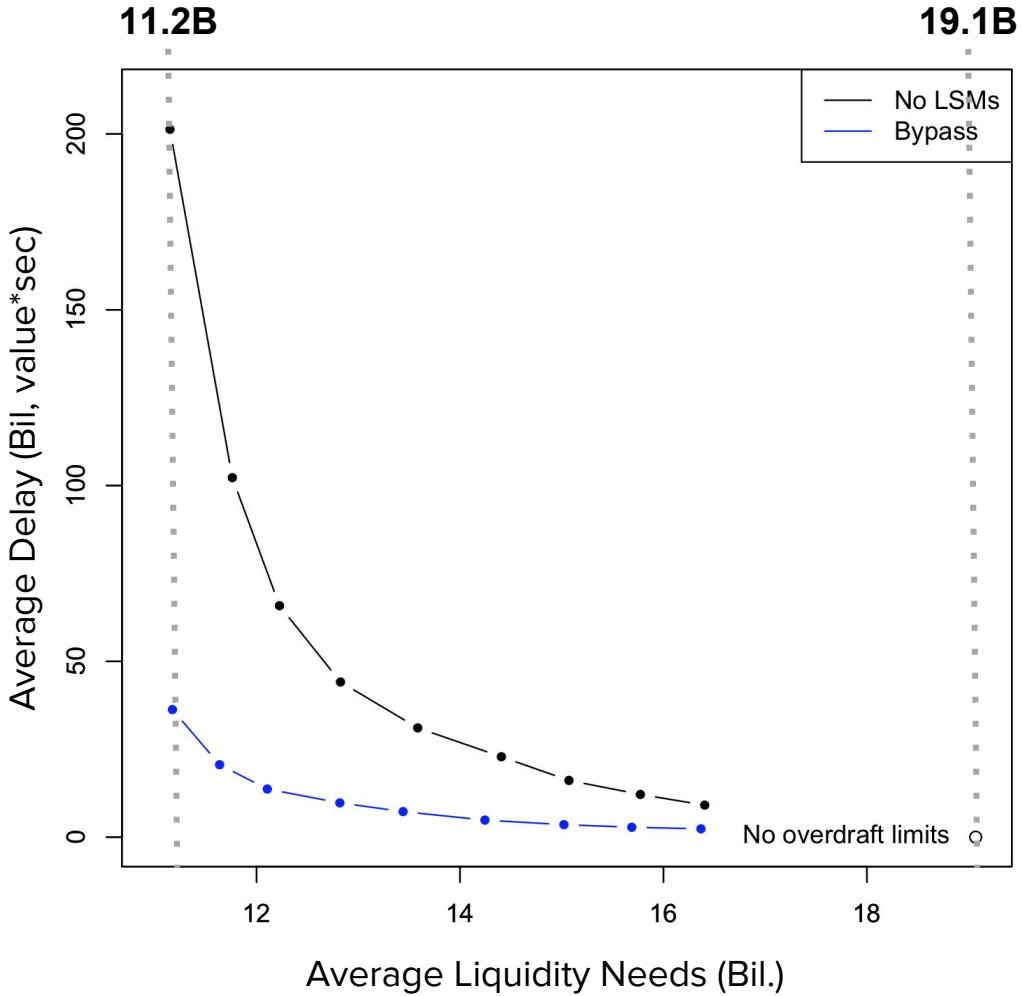
Simulations with Payments Canada for modernization of Canada's Payment System

Problem: FIFO order may 'block' settlement if a large payment is at the front of the queue.

Bypass FIFO tries to settle payments down the queue and selects the first one that it finds.

Example: A has liquidity available 200. A has queued payment: 300, 150 and 100.

Payment 150 can be settled.



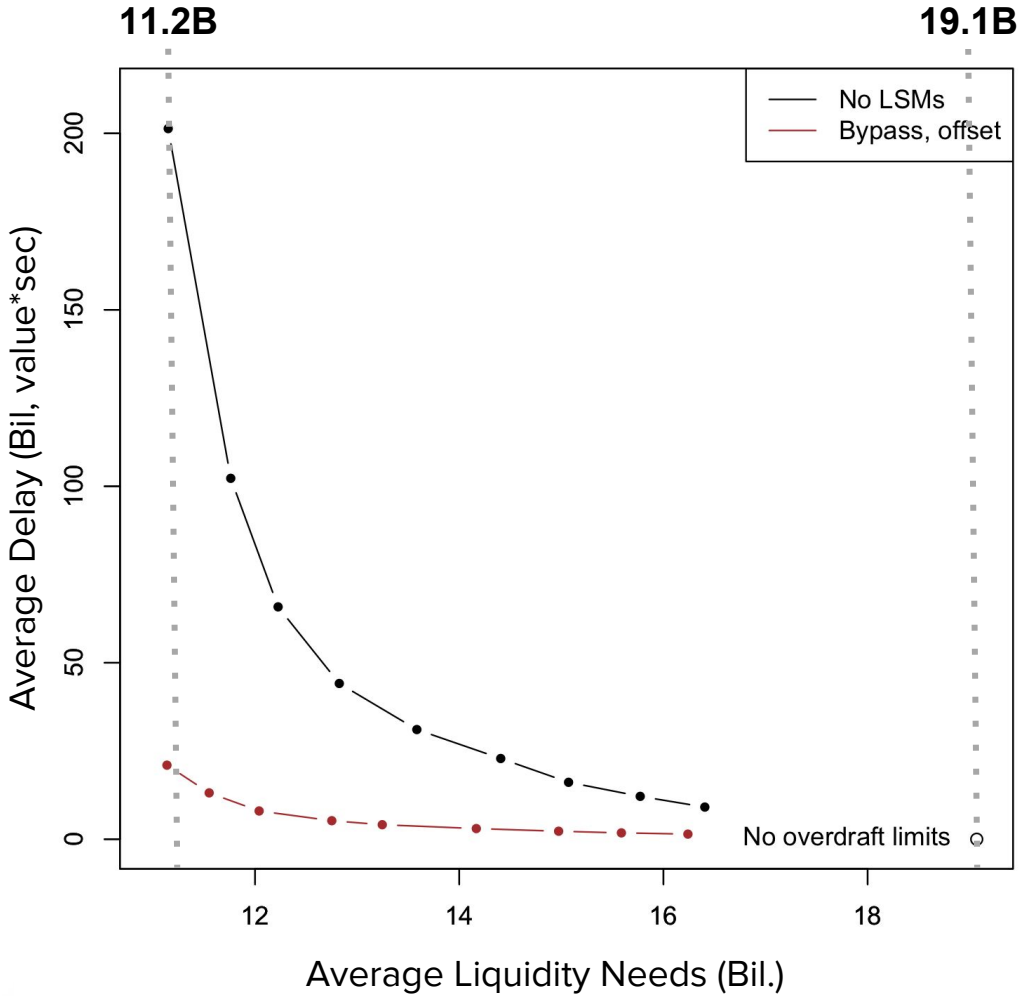
Bilateral Offset

Problem: Liquidity may be unnecessarily used when receiver has payments to sender in its own queue.

Bilateral Offset finds payments from receiver's queue that can be offset with sender's payment.

Example: A has 200 liquidity available and a payment of 500 to B. B has payments to A in queue: 300, 150 and 100.

Payment 500 can be settled offsetting 300 of the 500 against B's payment and the remaining 200 with available liquidity.

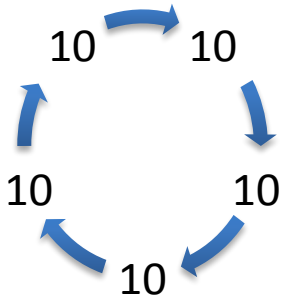


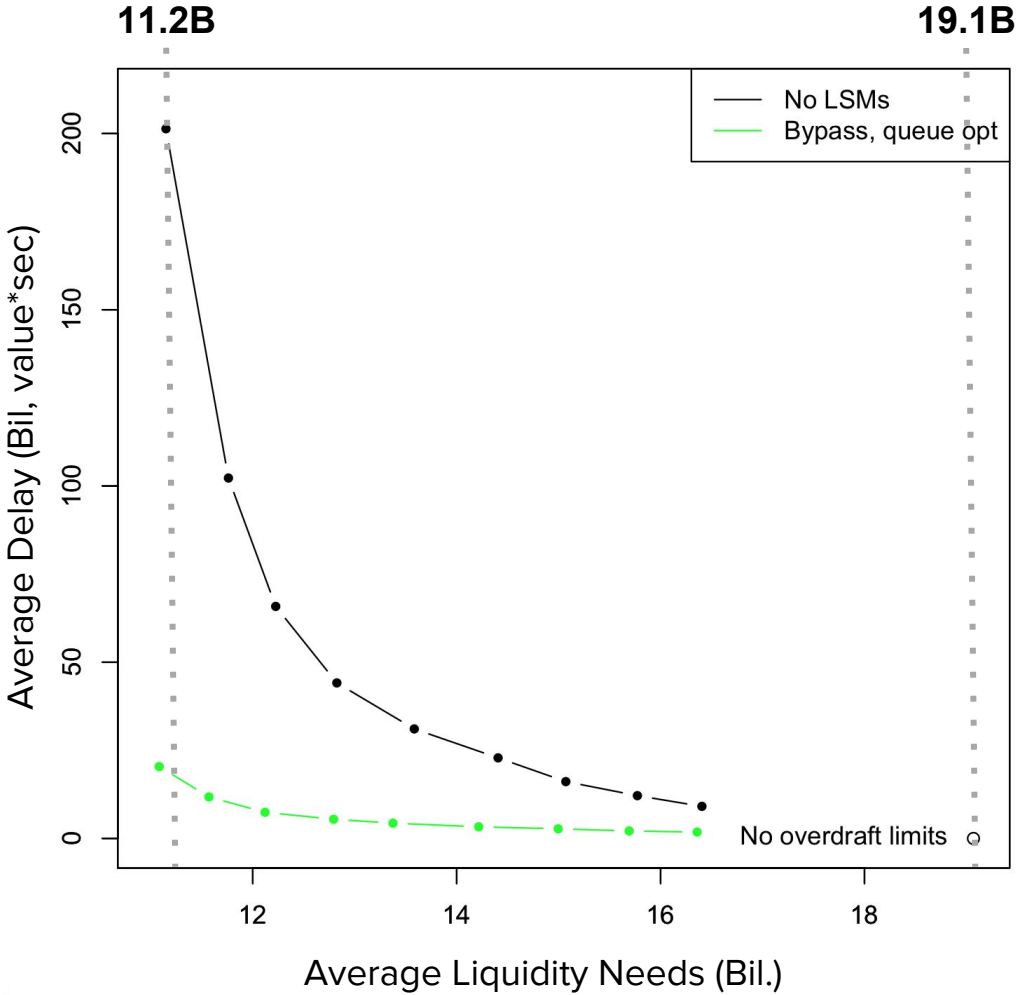
Queue Optimization

Problem: A system may become gridlocked

Queue optimization tries to find a subset of payments that can be settled by all banks with available liquidity through multilateral netting.

Example: A cycle where no bank has liquidity but all payments could be made simultaneously.





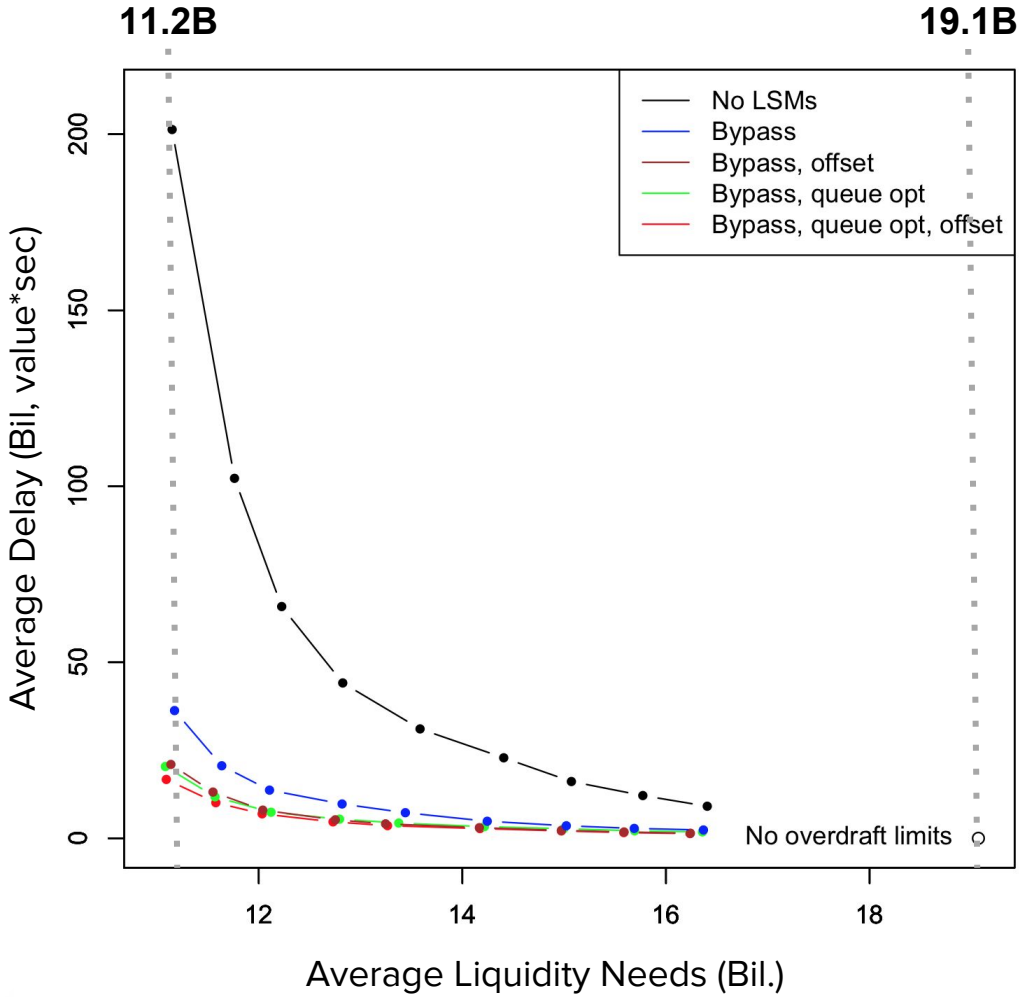
Summary

Bypass alone brings good benefits in reducing delays (or liquidity at a given delay level).

Other LSM's further improve on it.

Bypass + **Bilateral offset** and **Bypass** + **Queue Optimization** are equally efficient.

Having all LSM's running, brings best outcome from a liquidity-delay perspective



FNA

Summary

Top Down Analysis

THE WALL STREET JOURNAL.

Subscribe | Sign In

BUSINESS | JOURNAL REPORTS: LEADERSHIP

New Tools Give Better Picture, Literally, of Financial-System Risk

Researchers are using network analytics and advanced data modeling to identify weak spots in the system that otherwise might go unnoticed

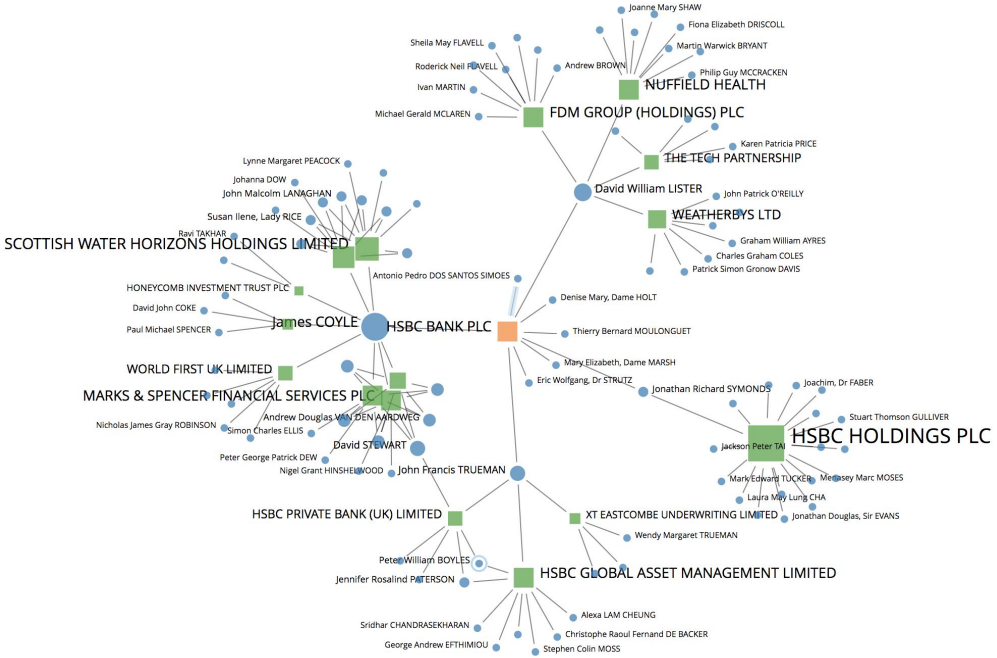
This sprawling tree shows housing prices in U.S. markets moving with little correlation in 2000. The tree has gotten shorter and shorter since, indicating higher correlation between markets.

PHOTO: FINANCIAL NETWORK ANALYTICS

Typical use cases:

- Systemic risk analysis
- System monitoring
- System design
- System stress testing
- Clustering/Classification
- Early warning
- Anomaly detection

Bottom Up Analysis



Typical use cases:

- Criminal investigation
- Terrorist networks
- Money laundering
- KYC & KYCC
- Fundamental investment analysis
- Supply chain analysis

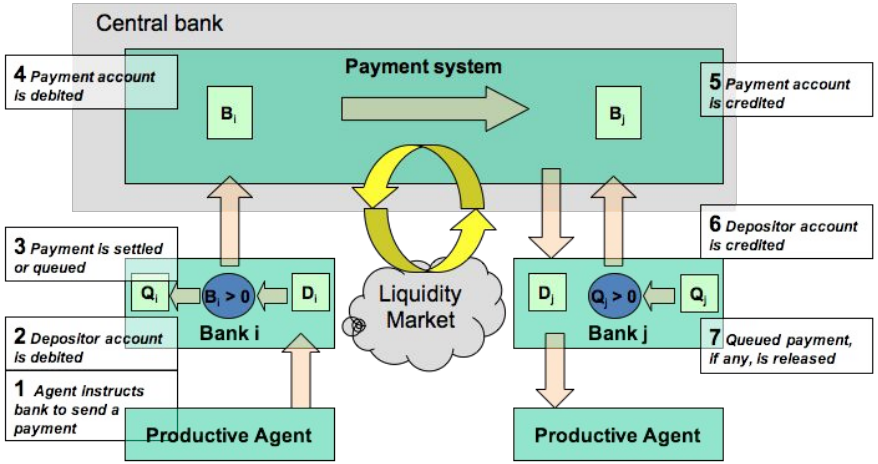
Network Features of Data

Typical use cases:

- AI/ML
- Fraud algorithms
- Recommendation engines
- Algorithmic investment

FNA Research: Comparison of Graph Computing Platform Performance

Agent Based Models



Beyeler, Glass, Bech and Soramäki (2007), Physica A, 384-2, pp 693-718.

Typical use cases:

- Central Counterparty Clearing
- Payment Systems
- FX Settlement
- Financial Markets
- Housing Markets

FNA

Dr. Kimmo Soramäki
Founder & CEO
FNA - Financial Network Analysis Ltd.

kimmo@fna.fi
tel. +44 20 3286 1111

Address
4-8 Crown Place
London EC2A 4BT
United Kingdom

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

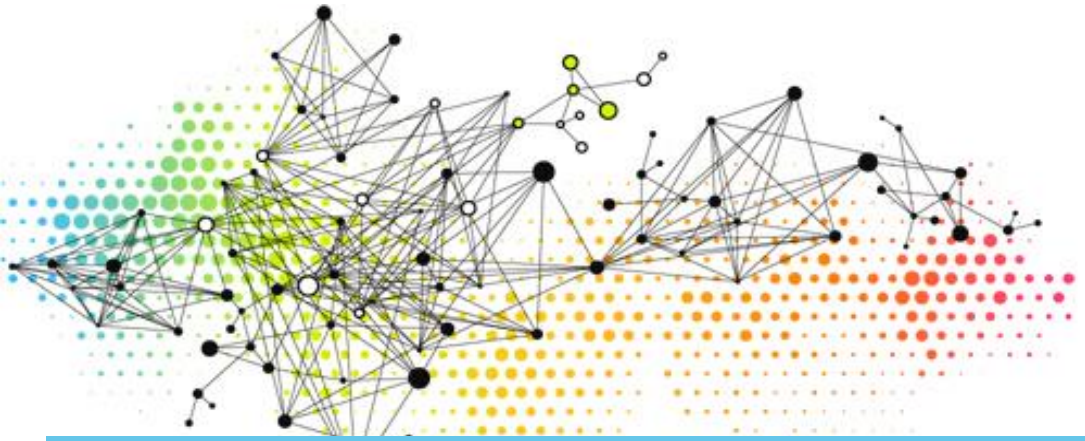
Bali, Indonesia, 23-26 July 2018

Big data: new insights for economic policy[1]

Gabriel Quirós-Romero,

International Monetary Fund

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.



Statistics Department

Big Data
New Insights for Economic Policy

Gabriel Quirós-Romero

Deputy Director, STA, IMF

BI-IFC/BIS INTERNATIONAL SEMINAR

BANK INDONESIA

Bali, 26 July 2018

Outline

I. Background

II. No straightforward definition; administrative data (?)

III. Potential

IV. Challenges

V. Statistical implications domain by domain

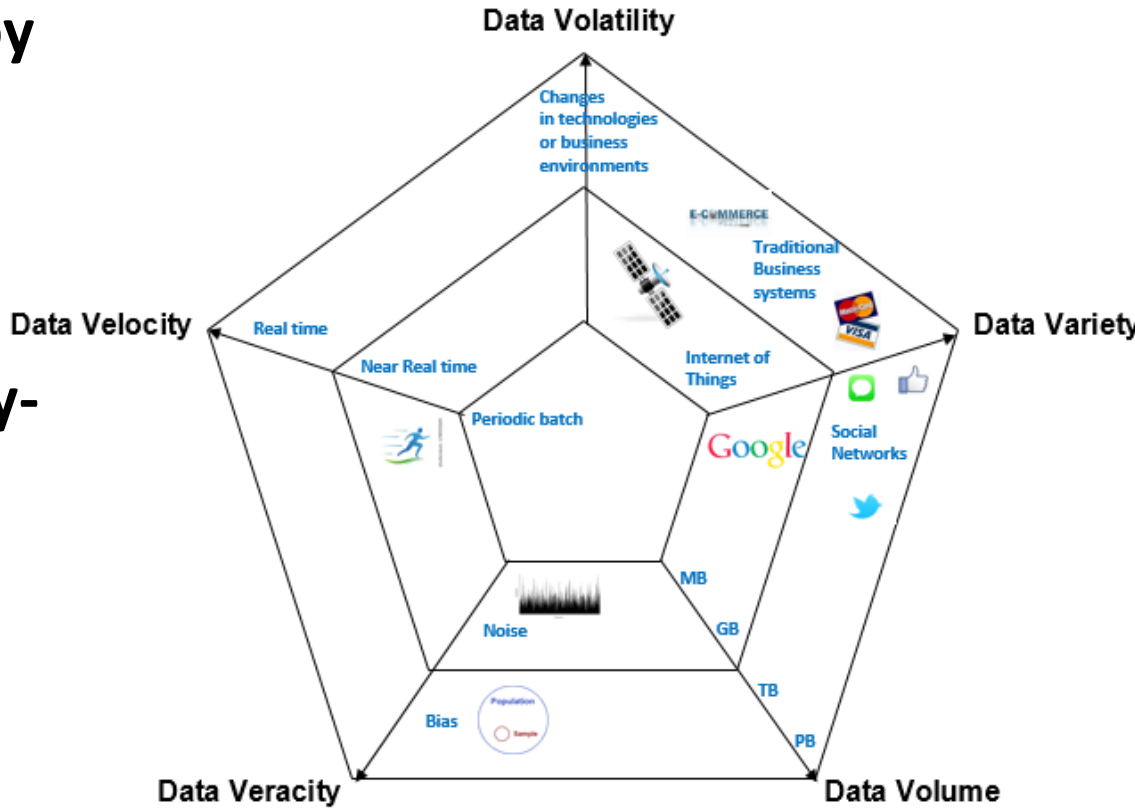
VI. Dos and Don’ts of Big Data for statistics

I. Background

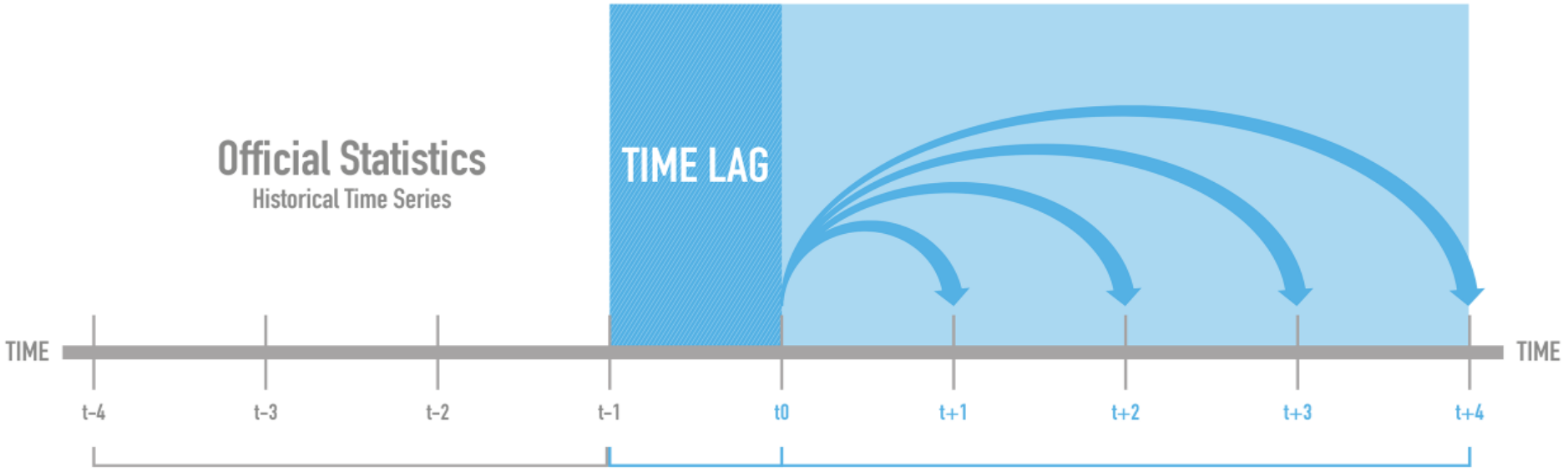
- **IMF STA is researching big data as a new data source for statistics, beyond administrative data:**
  - *potential* of big data to benefit macro-economic and financial statistics?
  - organizational, budgetary, and, in particular, methodological *challenges* that come with incorporating big data?
  - and strategic *statistical implications* for national and international organizations moving forward

II. No straightforward definition administrative data (?)

- Doug Laney, 2001
- **Big data characterized by the "5Vs"**
  - High-volume, high-velocity, high-variety
  - Veracity and volatility
- **Big data ("found") as 'by-products'**
  - Social networks
  - Business operations
  - The Internet of Things



III. Where, how is the Potential of Big data for statistics?



IV. Challenges

Data Quality

- quality assessments of indicators will be crucial to minimize governance, political, and reputational risks
- statistical techniques and methodologies best practices are needed to specifically address *veracity* and *volatility*
- big data for uncovering meaningful insights, trends, and sentiments may underlie different quality assessments compared to using big data in official statistics
  - ❖ continuation of consistent and harmonized historical time series is still needed
- metadata are key to assess and interpret new data sources

IV. Challenges

Data Access

- Proprietary data held by the private sector
  - Public-Private Partnerships that safeguard independence, privacy and confidentiality
- Data that companies own may evolve from a byproduct to becoming a major asset
- Regular licensing costs come in addition to substantial investments into processing and storage solutions
- Risk of volatility persists
- Best practices for building lasting relationships between data owners and data users are needed (UN Global Working Group)

IV. Challenges

- **New Skill Profiles**
  - Special career stream for data scientists
- **Multi-disciplinary project teams needed to make big data speak**
  - Experts from different professional backgrounds work together
- **IT implications**
  - Sharing of software codes and algorithms; open-source software; cloud-computing

V. Implications by Statistical Domain: potential (1/2)

Data Origin+	Data Type	Data Source and Techniques	Potential Indicators Derived	Statistical Domains	What May be the Potential?*
Social Networks	Social networks, blogs and comments 1100. Social Networks: Facebook, Twitter, LinkedIn 1200. Blogs and comments 1600. Internet searches on search engines (Google) 1700. Mobile data content: text messages, Call Detail Record, Data Detail Record, Location update, Radio coverage updates Online news	**Google** trends and search data	now-cast GDP now-cast unemployment consumer sentiment car and property sales	National accounts External sector statistics Financial Statistics Price statistics	**2**
		Mobile phone system data (electronic money schemes, e.g. M-Pesa) Peer-to-peer transactions	financial inclusion indicators remittances, regional disposable income, consumption patterns poverty reduction SDG "Gender Equality" economic growth	National accounts External sector statistics Financial Statistics Price statistics	**1,3**
		Twitter tweets	consumer confidence index border mobility, tourism, transitioning of migrants now-cast food prices sentiment and topic trend analysis	Mobility and urban statistics Price statistics Demographic and social statistics	**1,2,3**
		Web-scraping of **Facebook posts, Wikipedia articles**	geopolitical risk indicators price changes civil protests/labor strikes and national security events consumer sentiment inclusive infrastructure for sustainable development	Price statistics National accounts Demographic and social statistics Labor Statistics	**1,2**
		Call Detail Record data	SDGs indicators, travel/tourism, transport, migration	Mobility and urban statistics	**1,3**
Traditional Business Systems	**Data produced by public agencies** Administrative data	**Taxation registers**	consumer spending small business' income nonresident businesses controlled by resident parent corporations business profiling flight reservation system	National accounts Price Statistics External sector statistics Labor statistics Tourism statistics Transportation statistics	**2, 3**
		Population/business registers	multi-sourcing to derive population and housing census population structure	National accounts Demographic and Social Statistics	**3**
	Data produced by businesses 2210. Commercial transactions	**SWIFT data** on transaction quantities and financial market prices	global financial flows network concentration cross-border transactions export/import indicators	National accounts Price statistics External sector statistics Financial Statistics	**2,3**

V. Implications by Statistical Domain: potential (2/2)

Data Origin+	Data Type	Data Source and Techniques	Potential Indicators Derived	Statistical Domains	What May be the Potential?*
	2220. Banking/stock records 2230. E-commerce 2240. Credit cards Business websites Scanner data		withdrawal of correspondent banking relationships trade financing		
		Web-scraping to collect **price data** from **online retailers**	daily inflation turning points in inflationary trends e-commerce index	Price statistics Financial statistics	2,3
		Web-scraping **business websites**	enterprise profiling job vacancies	National accounts Financial statistics Labor statistics	2,3
		Scanner data Prices and quantities	national and regional consumer prices household income and expenditure	Price statistics National accounts Financial statistics	2,3
		Credit card data	consumer spending growth trends of the retail sales	National accounts External sector statistics	
Internet of Things (machine-generated data)	**Data from sensors** **311. Fixed sensors** 3111. Home automation 3112. Weather/pollution sensors 3113. Traffic sensors/webcam 3114. Scientific sensors **312. Mobile sensors (tracking)** 3121. Mobile phone location 3122. Cars 3123. Satellite images	GPS **positioning/tracking data**	travel services exports/imports trip duration inbound/outbound international travelers remoteness index traffic intensity	National accounts External sector statistics Demographic statistics Transport statistics Urban statistics Tourism statistics Population statistics	1,2,3
		Traffic/Road sensors	proxy of economic growth/health commuting time traffic intensity incoming/outgoing traffic travel/tourism	National accounts External sector statistics Transport statistics Tourism statistics Mobility statistics	1,2,3
		Satellite imagery Research and mapping of weather and climate data	improved geographical localization of statistical units and assets spatial sampling frame for output measurement land use and geostatistical cartography crop planting area, land use and agricultural output population and asset location as proxy for SDG "Gender Equality"	National accounts Price statistics External sector statistics Demographic and social statistics Transport statistics Agricultural statistics Demographic and urban statistics	1,3
		Smart meters (energy consumption measures)	non-occupancy rates household consumption electricity supply and consumption price differentials household structure and size	Environmental and Energy statistics National Accounts Price statistics Demographic and Social Statistics Transportation Statistics Geo-Spatial Statistics Agricultural Statistics Rural and Population Statistics	1,2,3

VI. Dos and Don'ts of Big Data

- In connection to the respective statistical domains, a number of Dos and Don'ts from big data can be identified, which are unevenly distributed across statistical domains
- The Dos and Don'ts are largely driven by the essence of big data: by-products of private technological and business models that capture behavior of consumers, corporates, banks, individuals or government agencies
- Big data are particularly promising to enhance directly or indirectly statistics on transactions, less so on stocks

VI. Dos and don'ts of big data

Dos

Big data, particularly promising at helping measure:

- ✓ "soft" information: sentiment, alerts, reactions…
- ✓ consumer behavior and patterns (e.g. Amazon, Google searches and 'clicks', social networks,…)
- ✓ Tourism and private consumption (e.g. roaming information, Google searches, credit cards, click-stream data, scanner …)
- ✓ Financial flows (e.g. SWIFT, mobile phones, …)
- ✓ Prices (scanner data,…)
- ✓ Job vacancies and labor skills (e.g. LinkedIn,…)
- ✓ Agricultural and construction (satellite images,..)
- ✓ big data provides granular, microdata
- ✓ …..

VI. Dos and Don'ts of Big Data

Don'ts

- Sample representativeness: bias towards more modern and dynamic economic activities and social behavior
- Big data less suited for stocks, i.e. total financial assets and liabilities of firms, households, government, non residents, both at micro and macro levels
- Revaluation and other volume changes, particularly important in monetary and financial statistics
- As by-product, long time-series based on big data are inexistent and will be fragile because instability from business and technological changes, discontinuity in data provision
- Privacy and confidentiality of personal, firm-level data

IMF Publication on Big Data

Big Data: Potential, Challenges, and Statistical Implications

C Hammer, D. Kostroch, G Quirós-Romero, and STA Group, 2017

https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2017/09/13/Big-Data-Potential-Challenges-and-Statistical-Implications-45106

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Big data: new insights for economic policy – The Bank of England experience[1]

Paul Robinson,
Bank of England

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANK OF ENGLAND

Big data: new insights for economic policy – The Bank of England experience

Building Pathways for Policy Making with Big Data

BI – IFC International Seminar on Big Data

Paul Robinson, Bank of England

26 July 2018

Outline

- Opportunities
- Challenges

Opportunities

Policy making is an inexact science

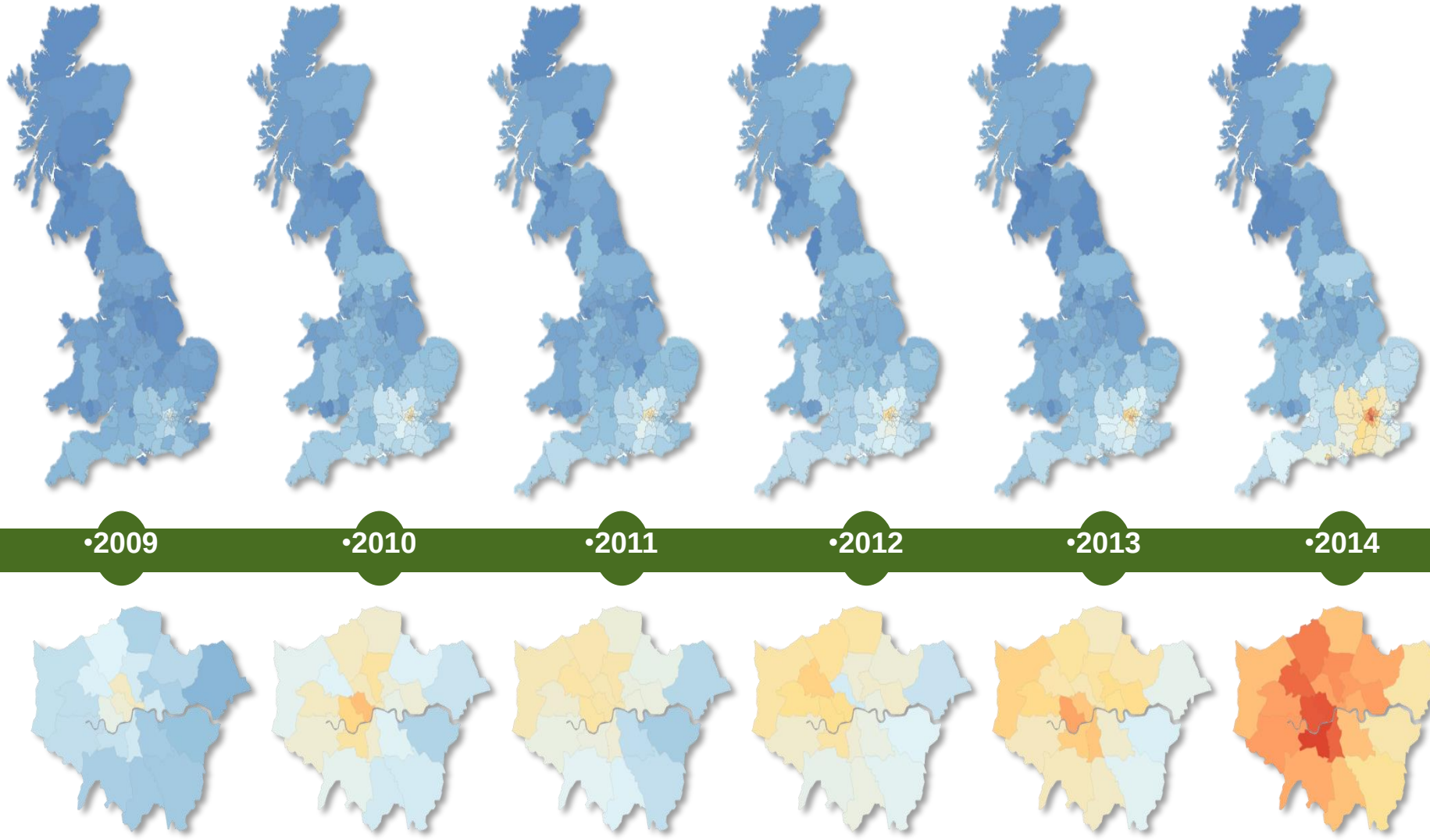
- Imperfect measurement
  - Noise, biases, blind spots, out of date information, (near) simultaneity of cause and effect
- “Too much” data, too little information
- Imperfect theory
- Complex, adaptive system with lots of feedback
  - Leads to “chaotic” behaviour
- Internal frictions

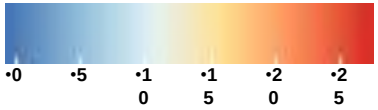
How can Big Data help?

- Imperfect measurement
  - Insight into previously hidden phenomena
  - Combining different types of data
  - Speed and completeness of coverage
- "Too much" data, too little information. Use data science methods to:
  - Improve processing large data sets
  - Help separate the signal from the noise
- Imperfect theory
  - Hypothesis generation
  - Alternative modelling approaches (eg Agent-based models)
- Complex, adaptive system with lots of feedback
  - Difficult to cope with, but more accurate understanding of initial conditions and more frequent updating help a lot
- Internal frictions
  - Improved management information

Big data sets offer significant potential advantages

- Greater **detail** (Volume, Velocity, Variety)
- Allow insights that aggregate numbers might obscure
- Examples:
  - UK housing market
  - Market dynamics around the abolition of the EUR/CHF floor
  - Market liquidity around large market moves



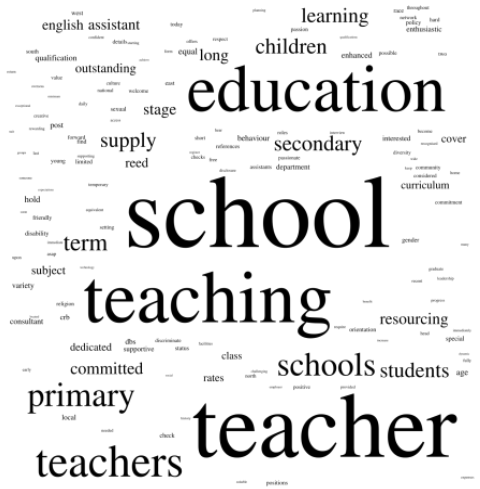


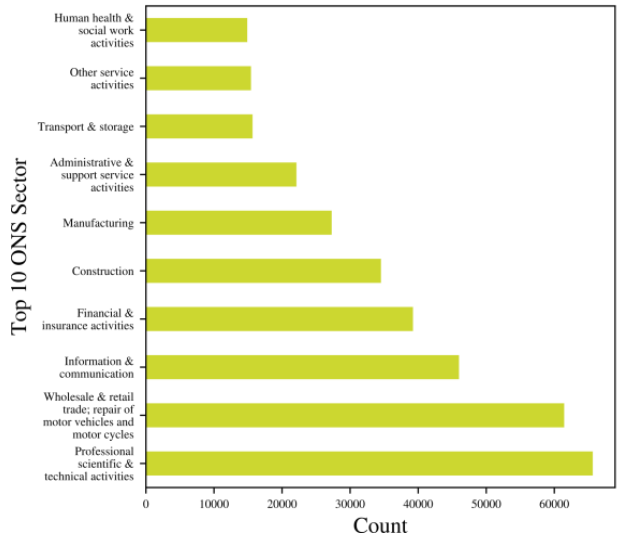
•**Key:** % of mortgages with loan more than 4.5x income

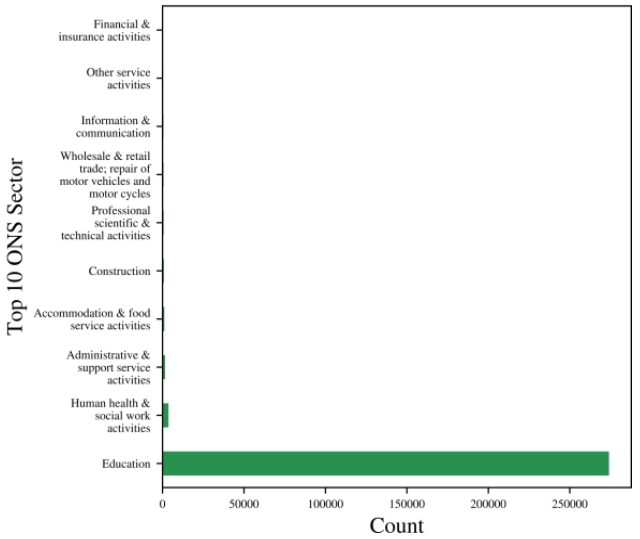
Big data sets offer significant potential advantages

- Greater **flexibility** (Velocity, Variety)
- Gives a window into changing structure of the economy
- Example:
  - Using job adverts to understand changing labour market dynamics

Understanding the labour market using job ads







Big data sets offer significant potential advantages

- Greater **timeliness** (Velocity)
  - ‘Nowcasting’ and ‘nearcasting’
  - Always important, especially in times of crisis
- Greater **efficiency** / value for money (Value)
  - Using administrative data
  - ‘Found’ data

Using Machine learning

- Best used as a complement, rather than a substitute, for other methods
- Objective:
  - The data speak more directly
  - Avoids “unbelievable assumptions of convenience” that underly some alternative modelling strategies
- Inductive rather than deductive approach
- Particularly well-suited to prediction problems
  - Eg statistical issues
  - Nowcasting

The topics

Word clouds of topics found using Latent Dirichlet Allocation.

Topic 0

Topic 1

Topic 2

Topic 3

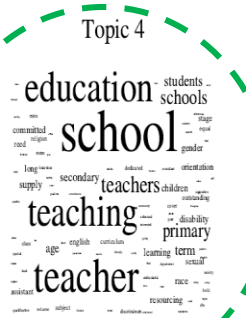
Topic 8

Topic 9

Topic 10

Topic 11

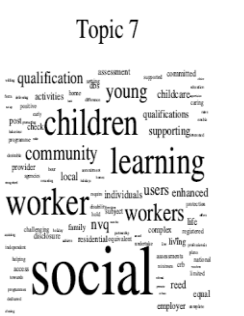
Topic 4



Topic 5

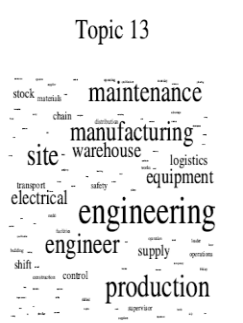
Topic 6

Topic 7



Topic 12

Topic 13



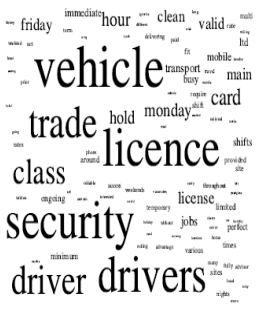
Topic 14

Topic 15

Topic 16

Topic 17

Topic 18



Topic 19

Challenges

Lots of data is not the same as lots of information

- Central banks are typically particularly focussed on avoiding:
  - Monetary dysfunction:
    - High inflation
    - Turbulent currency movements
    - Liquidity traps
  - Financial instability
  - Bank failures
- None of these happens frequently
  - So we often lack a track record

Lots of data is not the same as lots of information

- ML models are often black boxes
  - So policy makers find it hard to ‘inspect the mechnism’
- Together these make it difficult for policy makers to have confidence in the stability of ML models and predictions
- Lots of data does facilitate lots of (potentially spurious) correlations
- And lots of models
  - Similar forecasting ability but very different implications
  - How do we make the right choices?

Correlation versus Causality

- ML focuses on prediction
  - Not on structural models
  - But central banks set policy and a policy intervention may change the structure of the economy
  - Beware the ‘Lucas critique’ (and structural breaks)
- This does not mean that ML is not a good fit for central banks
  - Forecasts often matter
  - Intermediate targets can be useful

“Veracity”

- Big data sets are often populations, not samples
  - Therefore no sampling error
- But the observed population characteristics may not be typical of the underlying data generating process
- Or it may be biased relative to the true population of interest

Confidentiality / ‘Big Brother’ state

- This was not relevant to the CPI work
- In general, the more detailed and granular the data set is, the more likely it is to contain confidential information
- We must ensure that:
  - we only use data for appropriate reasons
  - the minimum number of people are able to see any confidential data given the needs of the situation
  - data are stored securely and professionally

Conclusion

- Big data has much to offer
- But it is not, and likely never will be, a panacea
- Policy judgement will always remain crucial
- So we need to ensure that we move towards:
  - Combination of the best parts of structural and data-driven models
  - Interpretable outputs from big data models

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Promise: measuring from inflation to discrimination[1]

Roberto Rigobon,

Massachusetts Institute of Technology (MIT)

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

PROMISE:

Measuring From Inflation to Discrimination

July 2018

Roberto Rigobon

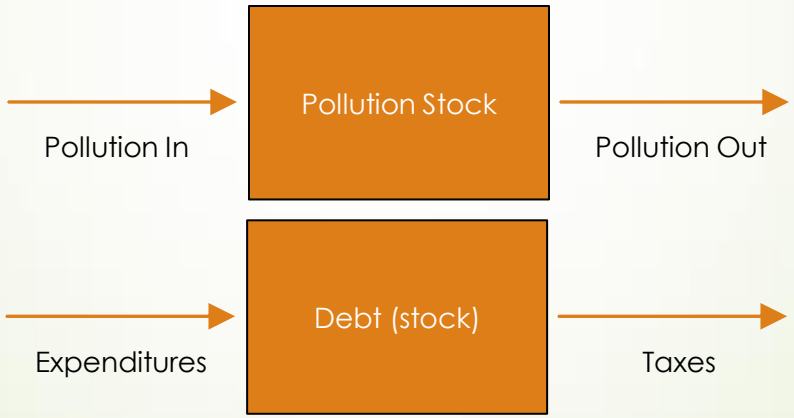
MIT, NBER, CNStat

Dimensions of Social Wellbeing

P	• Personal
R	• Relationships
O	• Organizations, Firms, and Jobs
M	• Markets and Economy
I	• Institutions
S	• Social and Political
E	• Environment

Principles

	P	R	O	M	I	S	E
Principles	Balance	Trust	Future Value	BB-NN	Consistency	SP	EE



Values

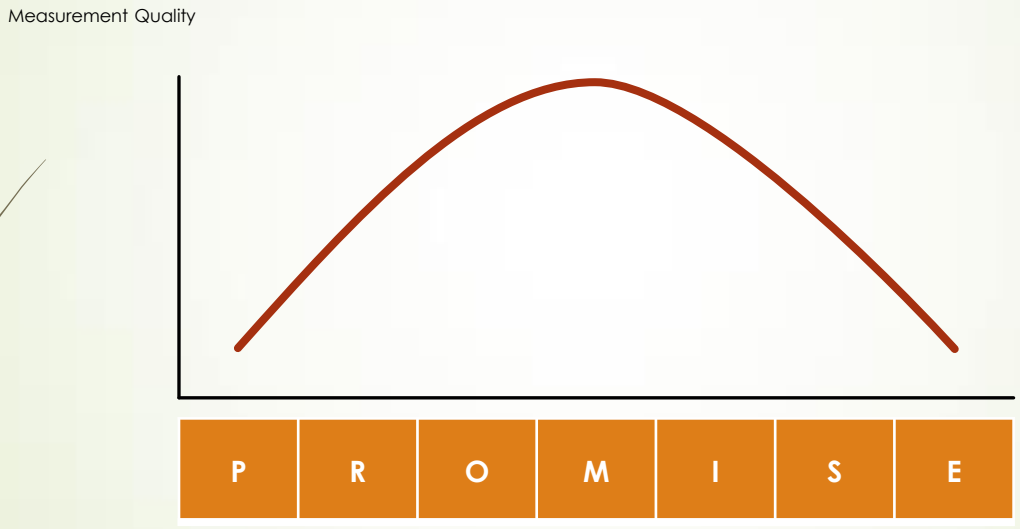
	P	R	O	M	I	S	E
Principles	Balance	Trust	Future Value	BB-NN	Consistency	SP	EE
Values	Weights on each dimension						

- How much do you care about each dimension?

Measurement

	P	R	O	M	I	S	E
Principles	Balance	Trust	Future Value	BB-NN	Consistency	SP	EE
Values	Weights on each dimension						
Measure	Quality of Measurement. How each "stock" is measured?						

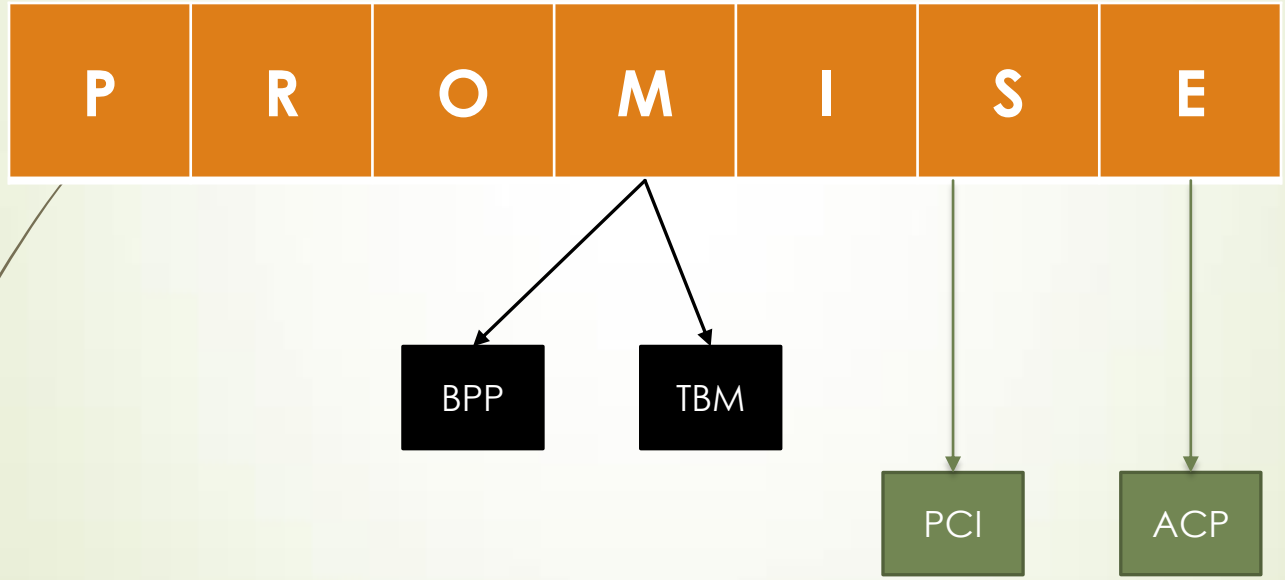
Measurement Quality



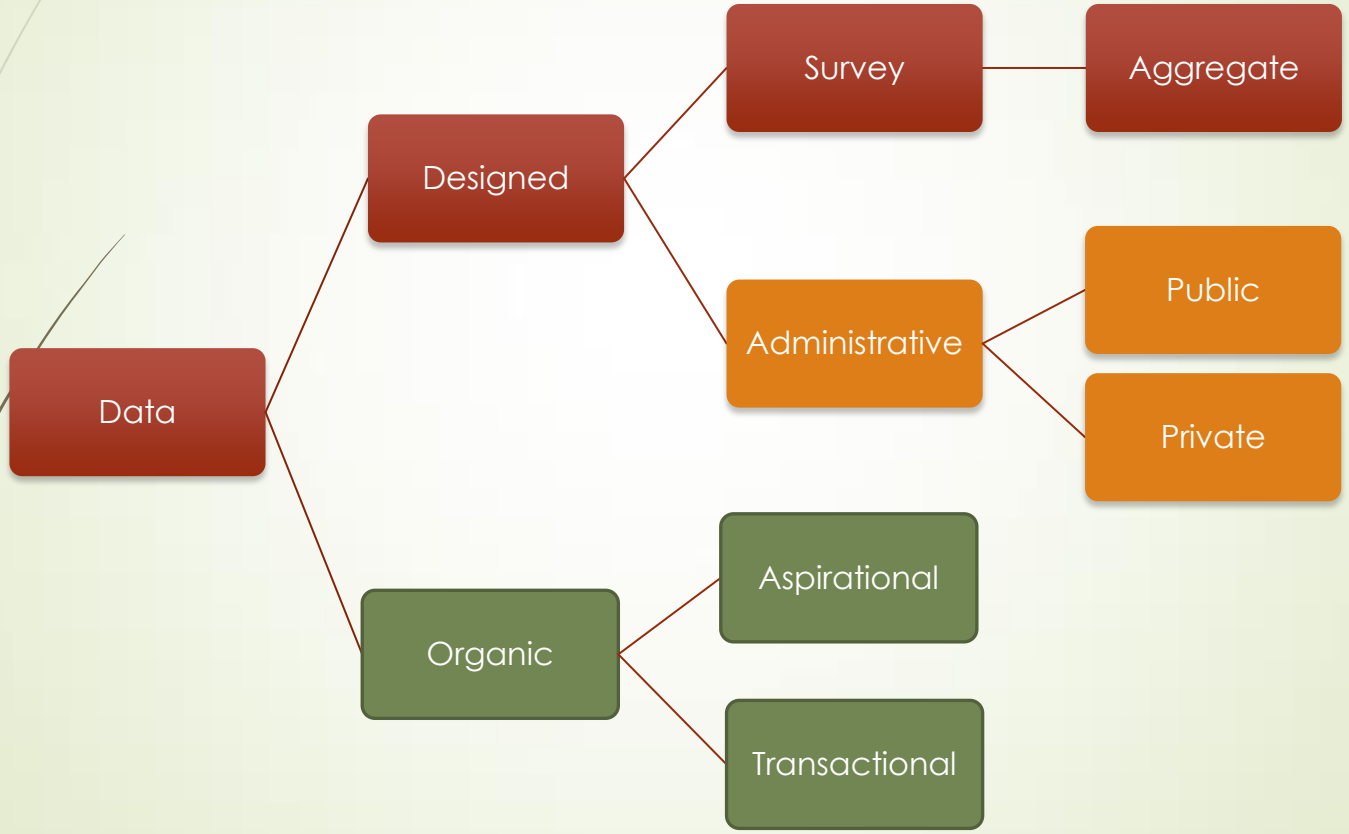
What you measure changes what you do!

- Incentives aligned with measurement
- Actions aligned with measurement

My Research



Evolution of Data Sources

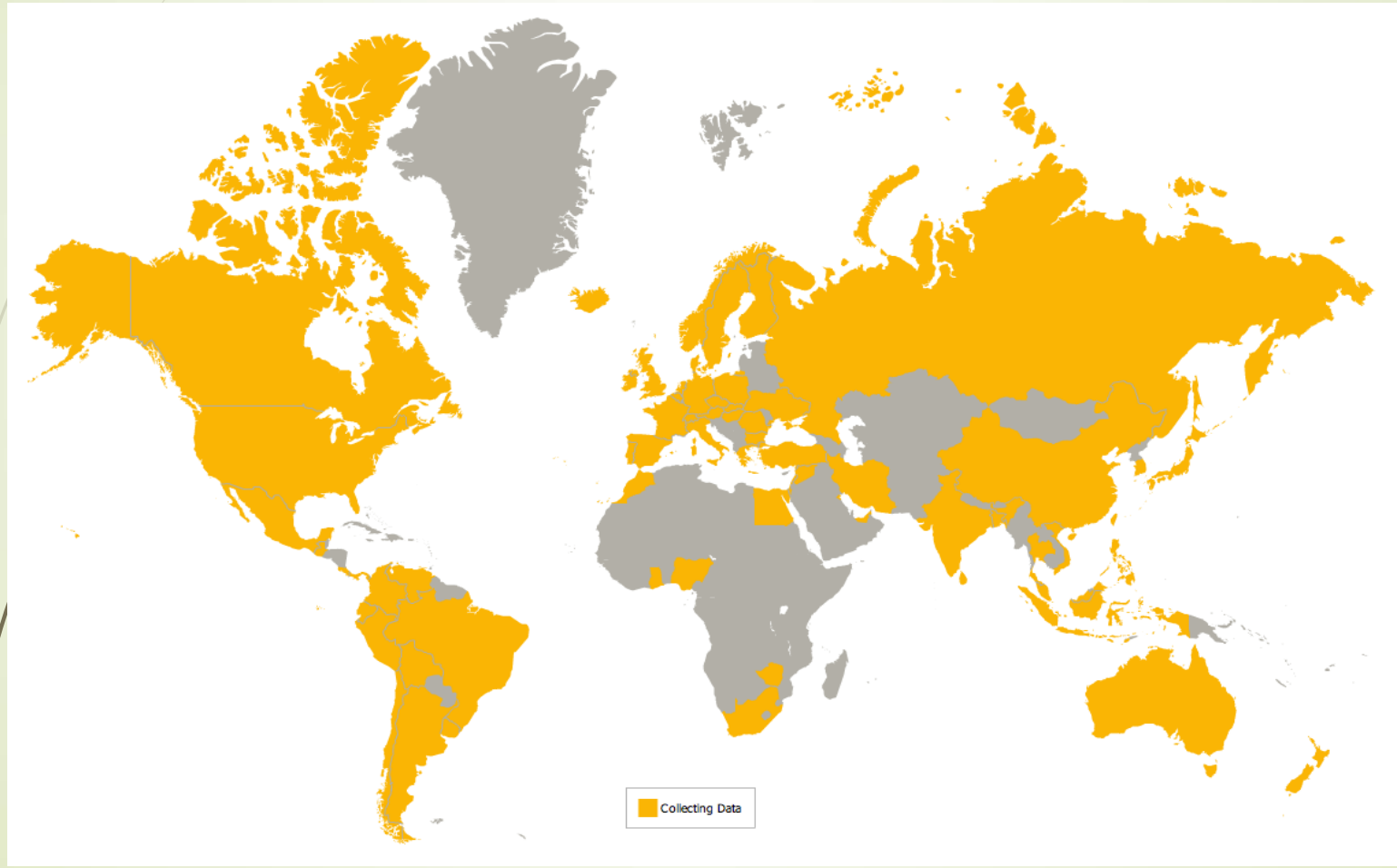


Advantages / Disadvantages

	Designed	Organic
Representative	Yes	No
Sample Selection	Response Rates Deteriorating	Extreme
Intrusive	Extremely Intrusive	Non-intrusive
Cost	Large	Small
Curation	Well-studied	Unclear
Structure	Geography and Socio-Economic	Behavior
Privacy	Well protected	Large Violations of Privacy

Hybrid approach to macro measurement

BPP: Countries covered



Online Information and Indexes

Our Approach to Daily Inflation Statistics

1. *Use scraping technology*
2. *Connect to thousands of online retailers every day*
3. *Find individual items*
4. *Store and process key item information in a database*
5. *Develop daily inflation statistics for ~20 countries*

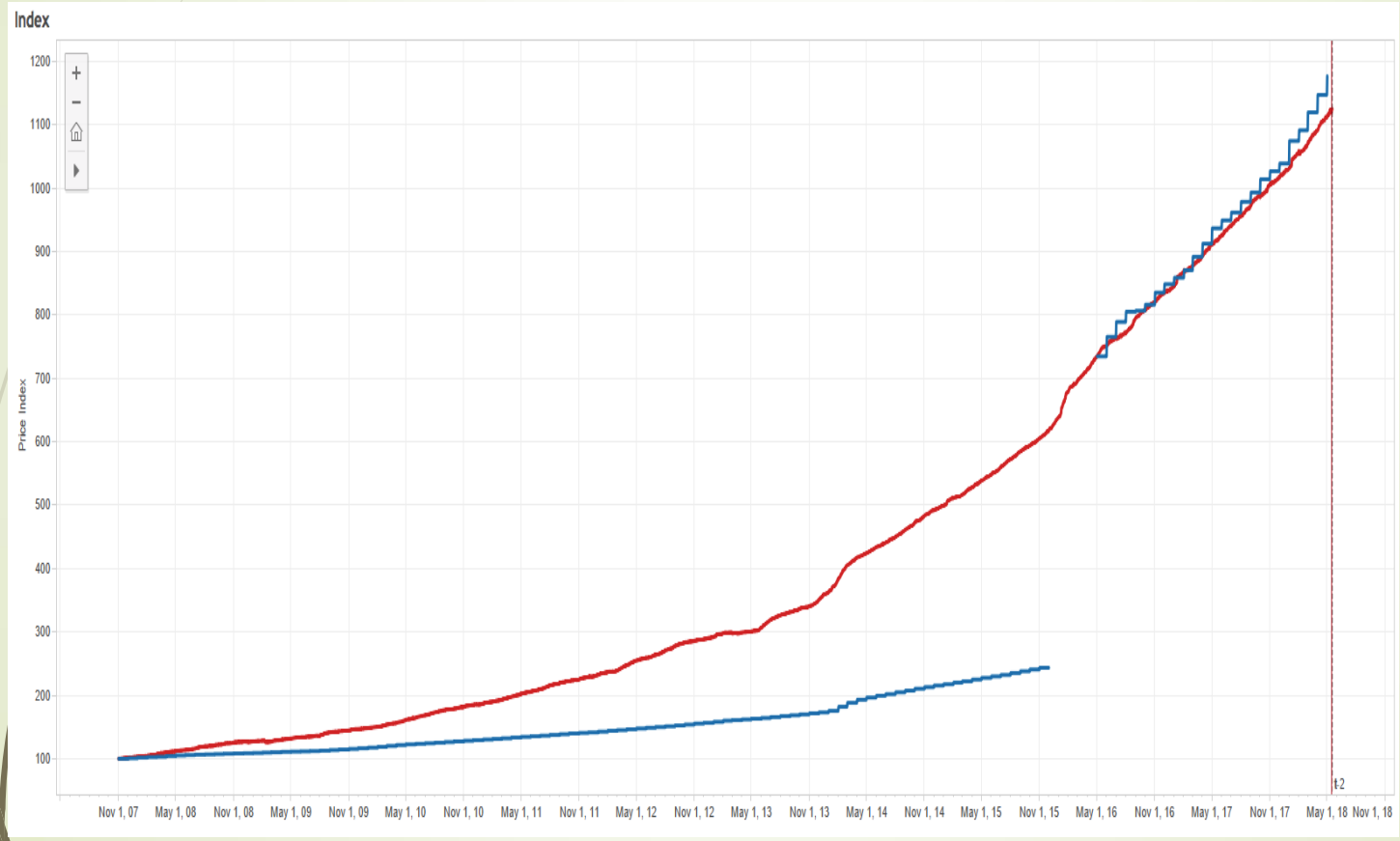
- *Date*
- *Item*
- *Price*
- *Description*

Online versus Offline?

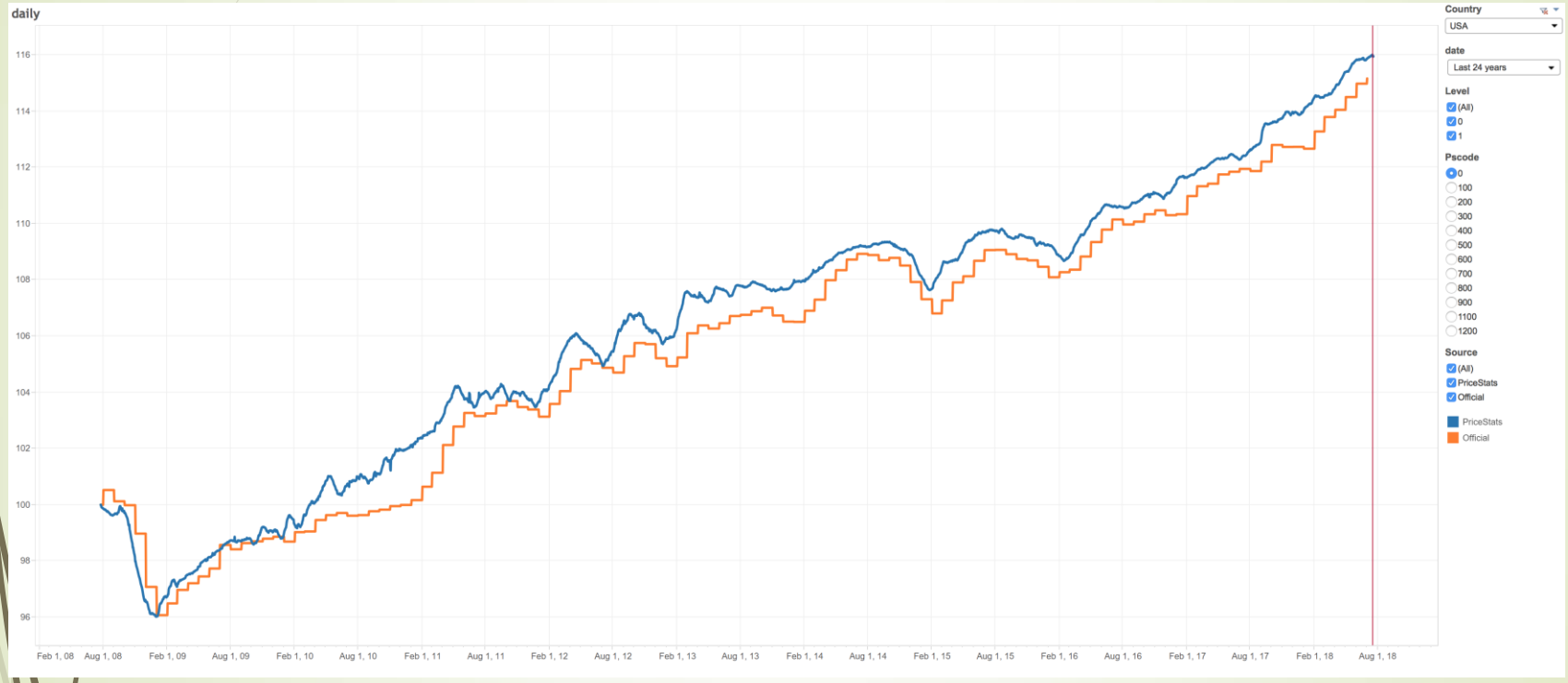
Preference to buy online versus in-store

Online		In-store
60%	Books, music, movies & video games	28%
39%	Toys	37%
43%	Consumer electronics & computers	51%
36%	Sports equipment/outdoor	44%
37%	Health & beauty (cosmetics)	47%
40%	Clothing & footwear	51%
32%	Jewelery/watches	49%
33%	Household appliances	56%
30%	DIY/home improvements	52%
30%	Furniture & homeware	59%
23%	Grocery	70%

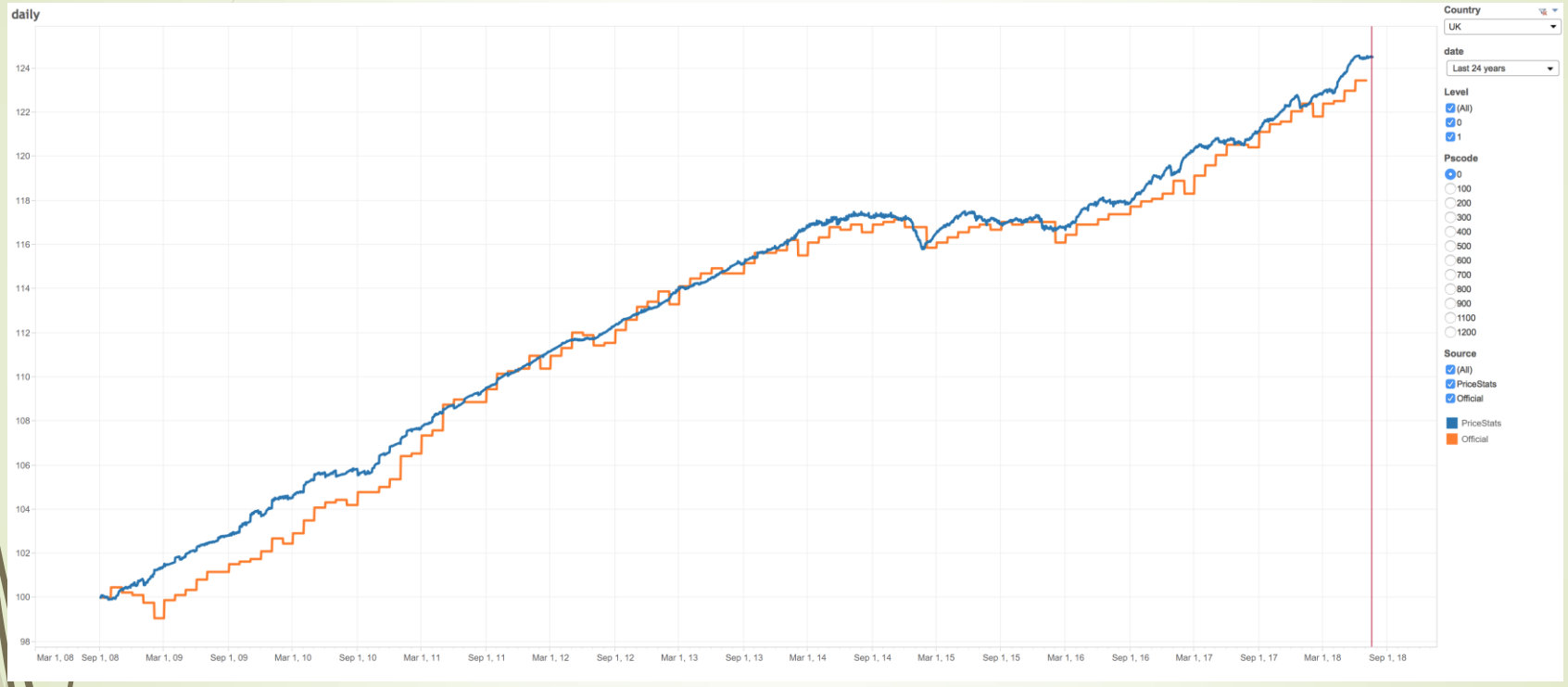
Argentina


Index
Price Index
1200
1100
1000
900
800
700
600
500
400
300
200
100
Nov 1, 07
May 1, 08
Nov 1, 08
May 1, 09
Nov 1, 09
May 1, 10
Nov 1, 10
May 1, 11
Nov 1, 11
May 1, 12
Nov 1, 12
May 1, 13
Nov 1, 13
May 1, 14
Nov 1, 14
May 1, 15
Nov 1, 15
May 1, 16
Nov 1, 16
May 1, 17
Nov 1, 17
May 1, 18
Nov 1, 18

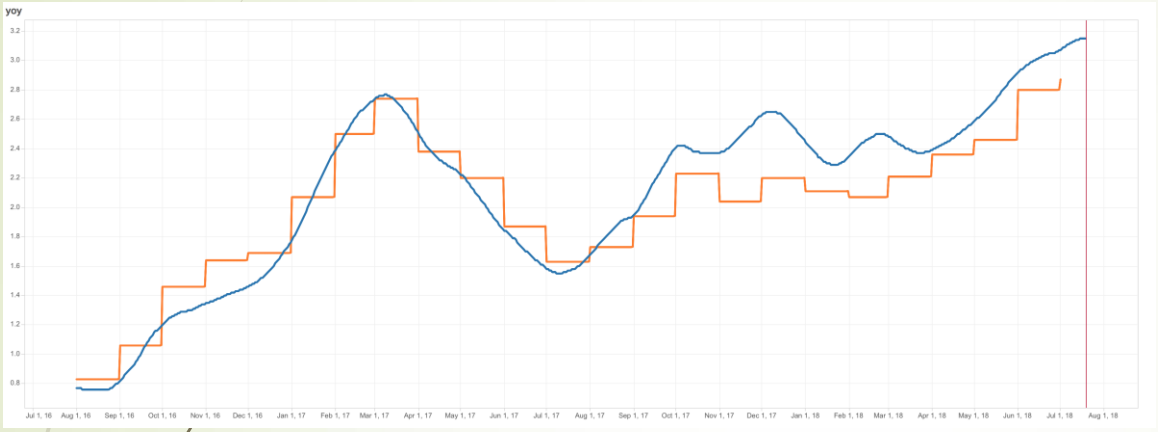
USA


daily
116
114
112
110
108
106
104
102
100
98
96
Feb 1, 08
Aug 1, 08
Feb 1, 09
Aug 1, 09
Feb 1, 10
Aug 1, 10
Feb 1, 11
Aug 1, 11
Feb 1, 12
Aug 1, 12
Feb 1, 13
Aug 1, 13
Feb 1, 14
Aug 1, 14
Feb 1, 15
Aug 1, 15
Feb 1, 16
Aug 1, 16
Feb 1, 17
Aug 1, 17
Feb 1, 18
Aug 1, 18
Country
USA
date
Last 24 years
Level
(All)
0
1
Pscode
0
100
200
300
400
500
600
700
800
900
1100
1200
Source
(All)
PriceStats
Official
PriceStats
Official

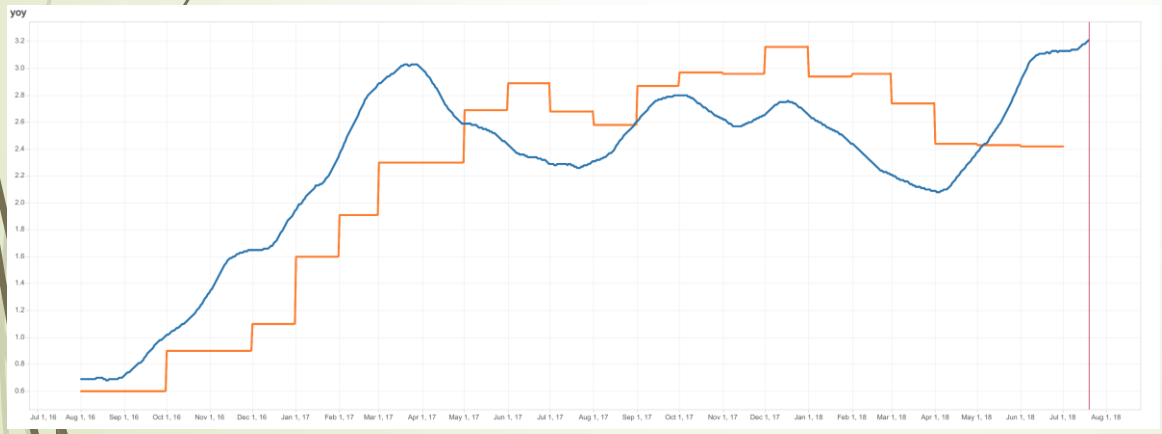
UK


daily
124
122
120
118
116
114
112
110
108
106
104
102
100
98
Mar 1, 08
Sep 1, 08
Mar 1, 09
Sep 1, 09
Mar 1, 10
Sep 1, 10
Mar 1, 11
Sep 1, 11
Mar 1, 12
Sep 1, 12
Mar 1, 13
Sep 1, 13
Mar 1, 14
Sep 1, 14
Mar 1, 15
Sep 1, 15
Mar 1, 16
Sep 1, 16
Mar 1, 17
Sep 1, 17
Mar 1, 18
Sep 1, 18
Country
UK
date
Last 24 years
Level
(All)
0
1
Pscode
0
100
200
300
400
500
600
700
800
900
1100
1200
Source
(All)
PriceStats
Official
PriceStats
Official

Inflation y-o-y



USA



UK

Thousands Big Mac's Project

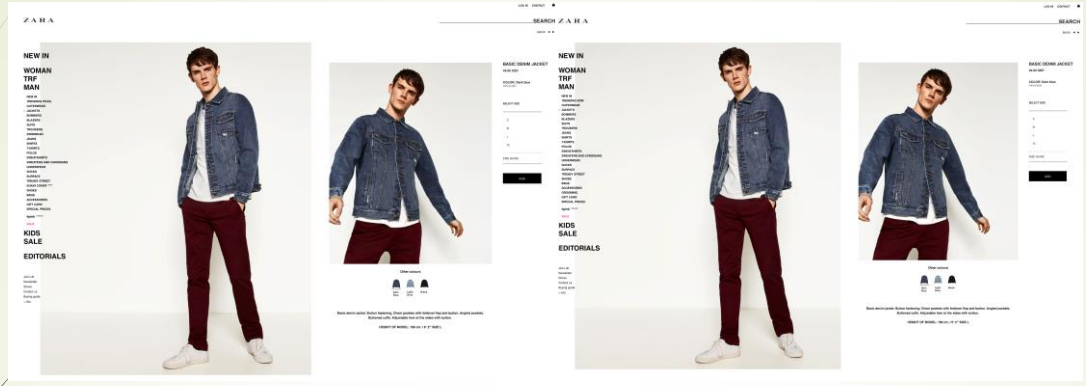
Compare prices for a bottle of Coke across countries

Apply similar approach to hundreds of products on a daily basis

PPP Indices

- Online prices represent an effective tool to measure PPP fluctuations
  - Identical items sold around the world
  - Detailed descriptions to achieve a nearly perfect matching
  - Daily Prices
- PPP indices:
  - More than 300 narrow product categories
  - With thousands individually matched items
  - In food, fuel, and electronics: we are missing clothing, personal care, household products.
  - Cars we will never match

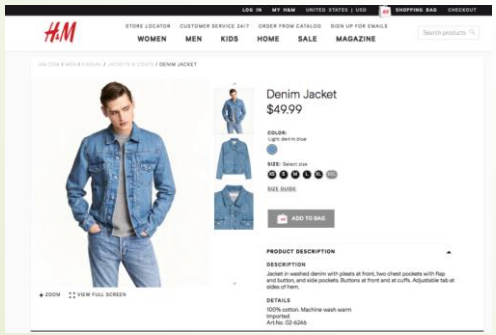
Relative Prices

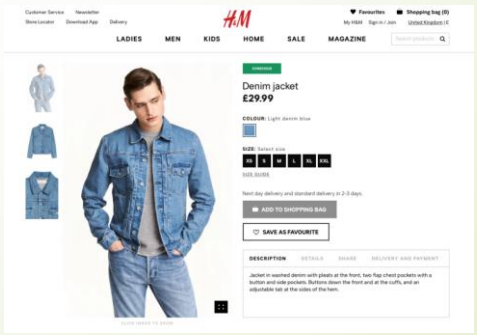


Price: 49.90
Product: 4081762

Price: 29.99
Product: 4081762

$\frac{49.90}{29.99}=1.664$



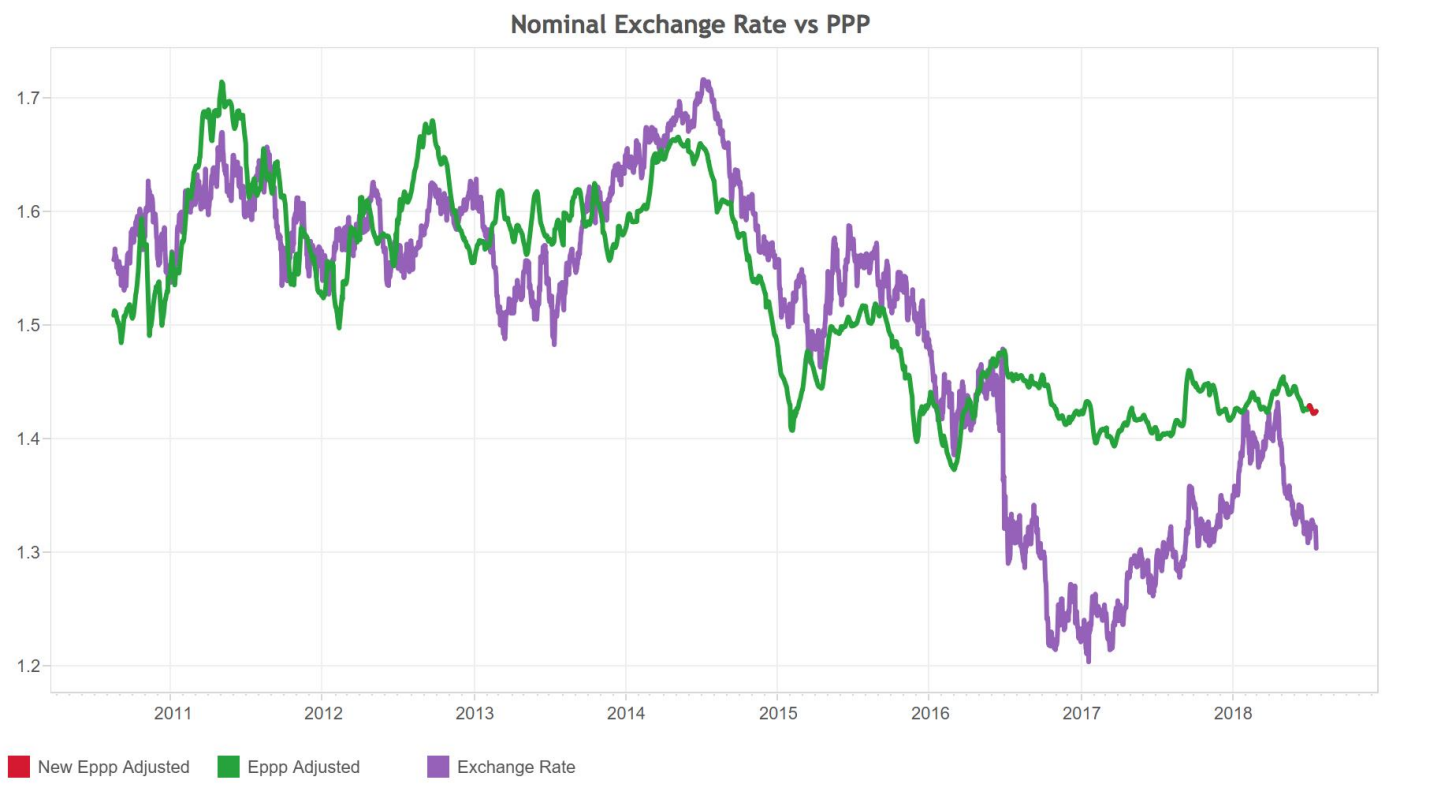


Price: 49.99
Product: 70136

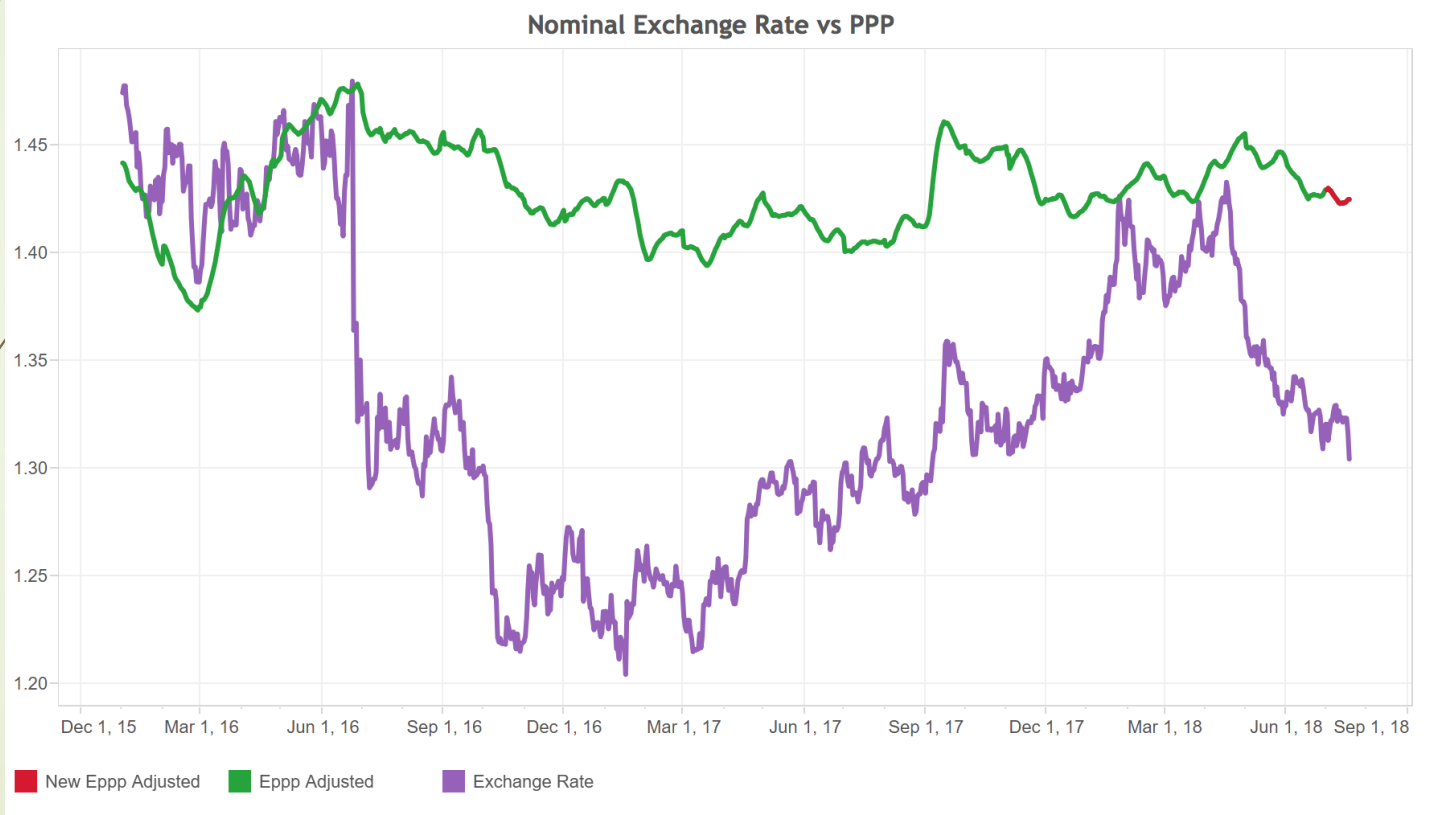
Price: 29.99
Product: 70136

$\frac{49.99}{29.99}=1.667$

UK


Nominal Exchange Rate vs PPP
1.7
1.6
1.5
1.4
1.3
1.2
2011
2012
2013
2014
2015
2016
2017
2018
New Eppp Adjusted
Eppp Adjusted
Exchange Rate

UK (3 years)


Nominal Exchange Rate vs PPP
1.45
1.40
1.35
1.30
1.25
1.20
Dec 1, 15
Mar 1, 16
Jun 1, 16
Sep 1, 16
Dec 1, 16
Mar 1, 17
Jun 1, 17
Sep 1, 17
Dec 1, 17
Mar 1, 18
Jun 1, 18
Sep 1, 18
New Eppp Adjusted
Eppp Adjusted
Exchange Rate

Aggregate Confusion

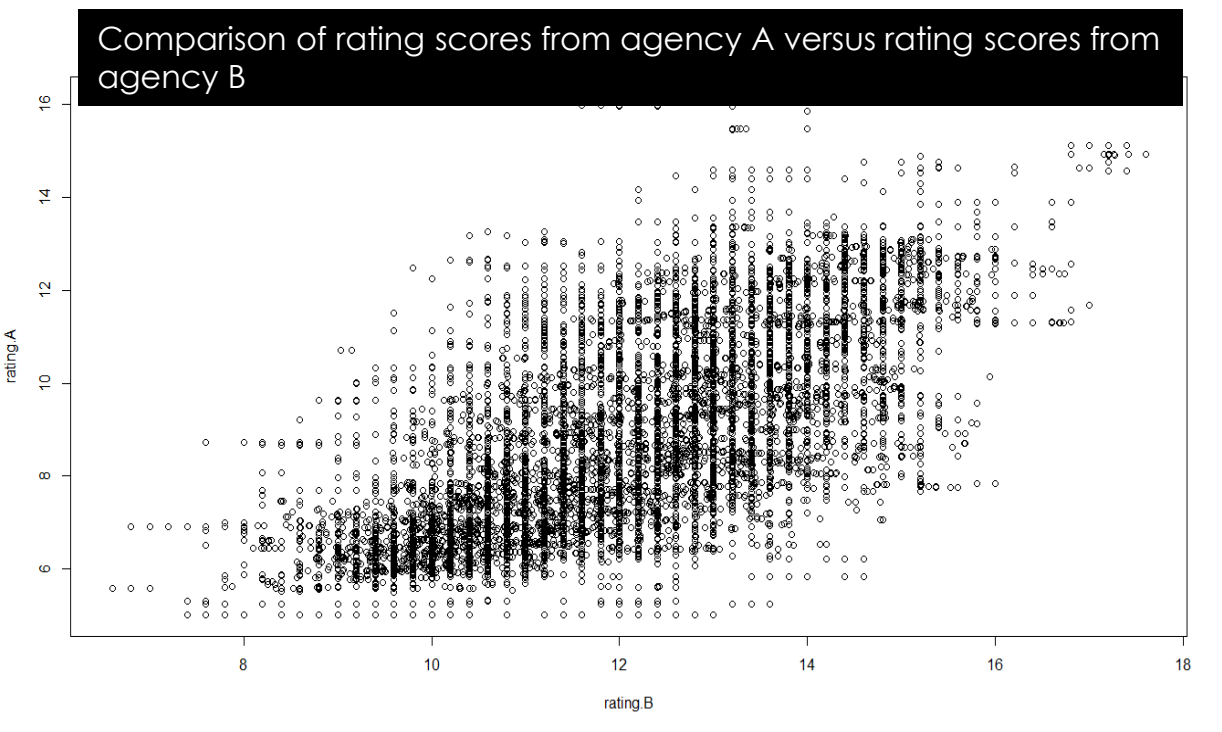
Solving the confusion of ESG ratings

Julian Koelbel, Florian Berg, Roberto Rigobon

MIT MANAGEMENT SLOAN SCHOOL

July, 2017

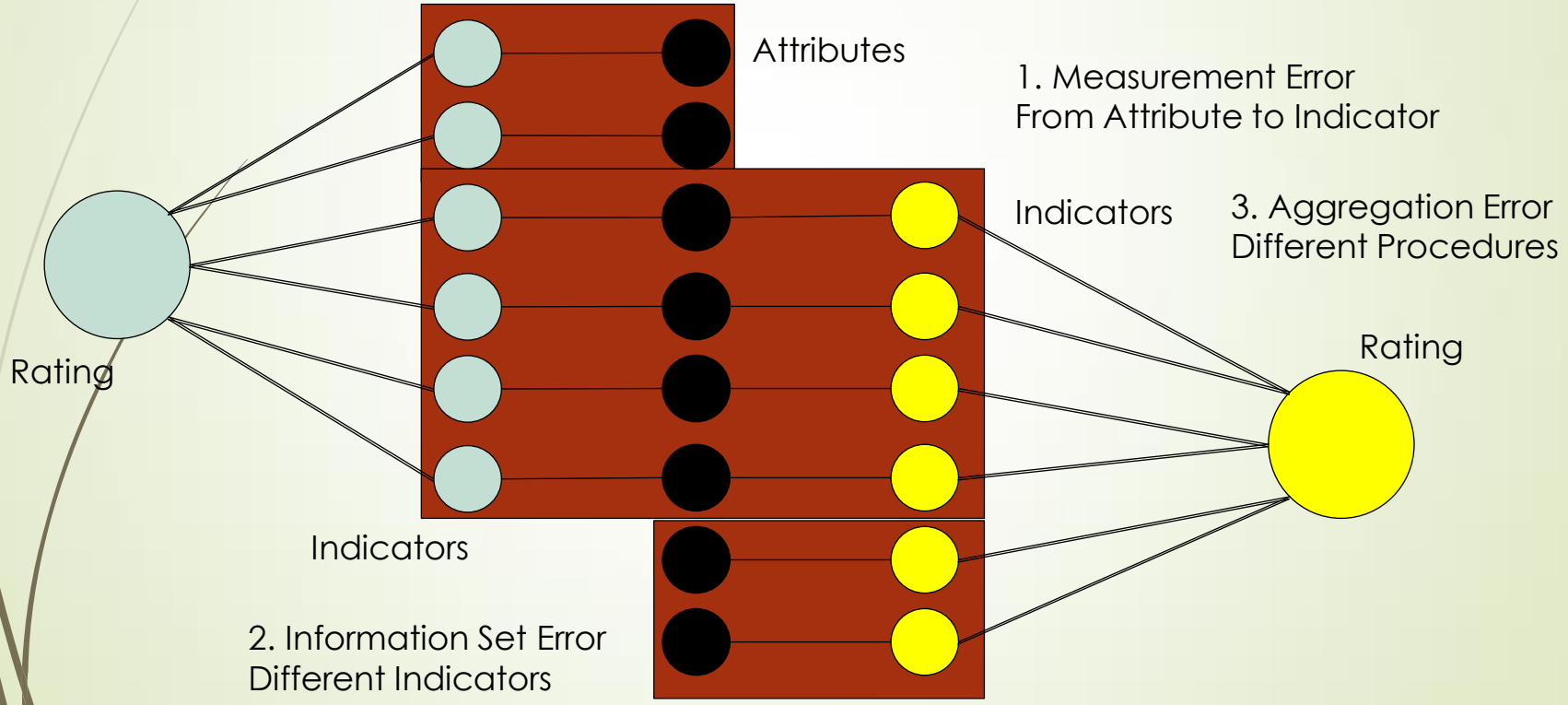
ESG Rating Agencies' Scores diverge



Things that are more correlated than the ratings!

- The number of people that downed every year in swimming pools in the US is correlated with the number of Nicholas Cage movies: 66 percent
- 72 percent of people who dislike licorice understand HTML
- 75 percent of people who can't type without looking at the keyboard prefer thin-crust pizza to deep-dish
- Per capita consumption of mozzarella and PhD's in civil engineering are correlated in 96 percent

Sources of errors



Preliminary Results

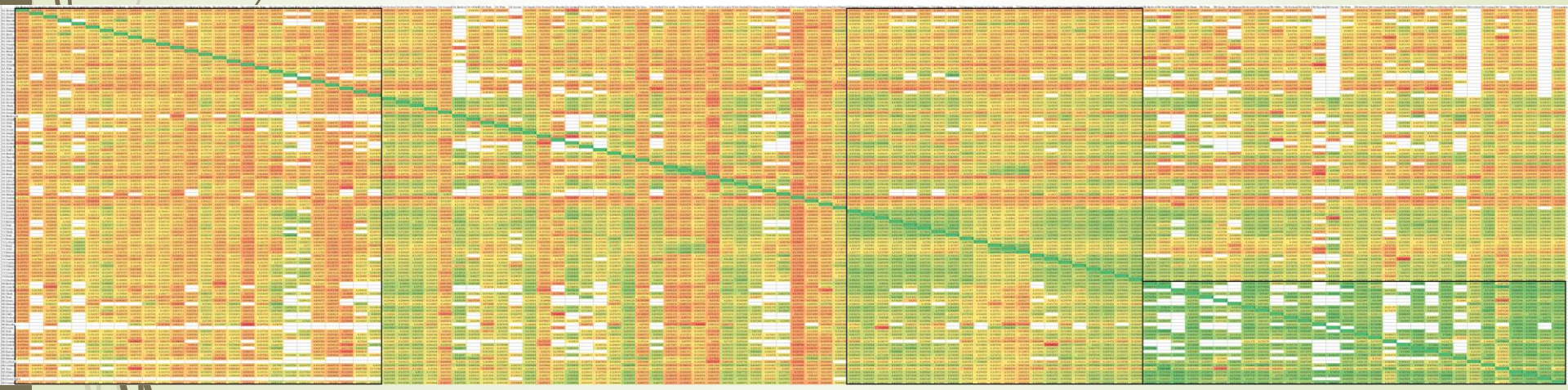
	Discrepancy			
	Aggregation	Measurement	Information and Unmeasured	Straight Explanation
		VI		
MSCI/KLD	3.3%	48.6%	48.1%	19.47%
RS	1.7%	87.7%	10.6%	49.14%
SA	3.8%	22.5%	73.7%	50.38%
		RS		
MSCI/KLD	0.0%	54.9%	45.1%	18.71%
VI	0.3%	82.0%	17.7%	49.14%
SA	3.5%	34.3%	62.1%	43.52%
		SA		
MSCI/KLD	1.5%	29.8%	68.7%	22.89%
VI	6.9%	56.3%	36.8%	50.38%
RS	5.4%	76.3%	18.3%	43.52%
		MSCI/KLD		
VI	0.0%	73.1%	26.9%	19.47%
RS	0.0%	84.1%	15.8%	18.71%
SA	1.5%	43.3%	55.2%	22.89%
		58%	40%	

Preliminary results

- Reverse Engineering implies quasi-linear rules
  - Get more than 98+ percent explanatory power within sample
- Variance Decomposition of Discrepancy
  - 58 percent is measurement
  - 40 percent is information set
  - 2 percent is from aggregation rules

Attribute measurement errors

- At the attribute level things are worse!
  - Some attributes are negatively correlated across rating agencies!
    - Lobbying
    - Responsible Marketing Policies

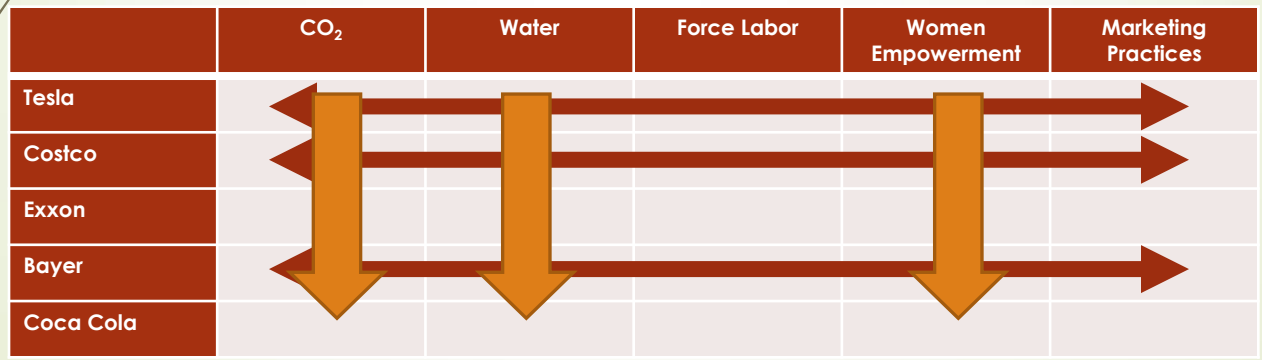


Three Domains for Improvement

- Organizational Structure
- Aggregation Procedure
- Data Source

Organizational Bias

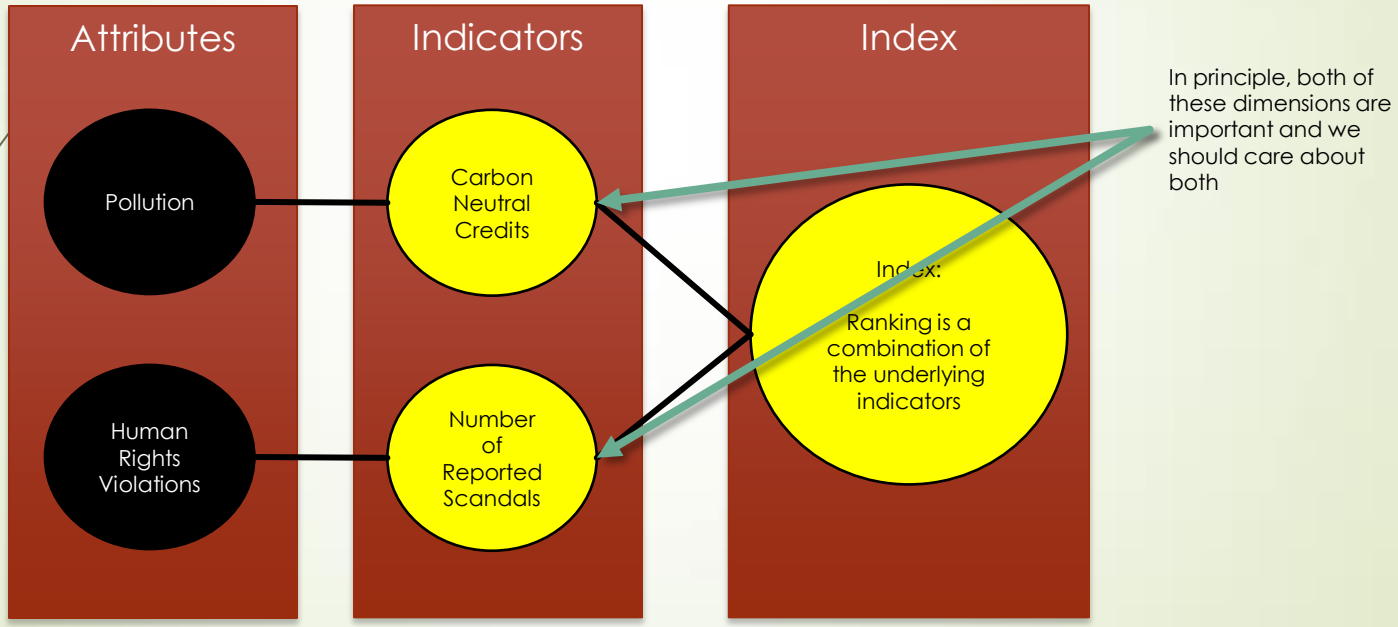
- Organizational Structure
  - Analysts are organized through firms and not across attributes (rows as opposed to columns)
  - Bias is firm (analyst) specific



  - Florian's research (still preliminary)

Quasi-linear Rules

- Ranking Implies that Dimensions that should NOT be substitutable become substitutes



Data Source Bias

- Data mostly comes from Three sources
  - **Statements**: Self-reported or non-mandatory surveys
  - **Outcomes**: Extreme Events
    - Accidents or scandals measured in public documents, courts, or regulator's reports
  - **Perception** : Reporting on extreme events
    - Accidents or scandals as reported by the media
  - Taxonomy might need another dimension

New Measures

- How to measure?
  - Women empowerment
  - Household Stress
  - Mansplaining and Hidden Networks (Team Chemistry and/or Attribution Problem)
  - Human Trafficking
  - Opioid Crises
  - Job Quality and Job Satisfaction
  - Water management practices

Example

- Women Empowerment
  - How we measure today?
    - Counting — Only relevant at the extremes
    - Scandals — Only relevant if media agrees
    - Wage gap — Discrimination on regressors
  - Characteristics?
    - Too late
    - Too concentrated on extreme events
    - Too infrequent
    - Too based on perception (news outlets)
    - Too fixated in the wrong statistic
  - How are we evaluating materiality?

Pillars of the new measures

1. Continuous Measurement of process
   - Timely measures
2. Non-intrusive
   - Can't rely on surveys – needs electronic forms of data collection
3. Open source
   - Many could adopt the methodologies
4. Privacy protecting
   - Violations of privacy can be significantly harmful, especially when estimating hidden behavior that is morally questionable
5. Imperfect Measurement
   - To guarantee the previous 4 characteristics the measures need to be noisy.

Example

- Women Empowerment
  - Measure interruptions
    - How frequently men interrupt women and other men?
    - Airtime: how long each one talks?
    - Mansplaining: assertiveness and apologies
    - Culture, hierarchy, language, media matter – but why gender and minority status would?

How much of the unconscious biases are unmeasured biases?

Data sources of organic data

- Digital documentation of transactions with a service or manufacturing process
  - Credit card transaction, retail sales scanner data
- Data from social network communication
  - Facebook, Twitter, Instagram
- Data transmitted from software agents within mobile devices
  - GPS
- Data from the internet of things
  - e-commerce: Web pages, price aggregators
  - Utility meter data, sensor data for traffic, air, water, soil quality
- Biometric data
  - DNA
- Human communication digital data
  - Emails, blogs, text
- Digital video data

Weaknesses depend on use

	Survey	Estimation	Forecasting	Measurement
Representativeness				✓
Selection Bias		✓		✓
Reliability and Consistency	✓		✓	✓
Transparency on Data Collection and Treatment	✓	✓		✓
Errors-in-variables		✓	✓	✓
Aspirational (Transactional)	✓	✓		✓
Private (as opposed to public)	✓		✓	✓
Model Uncertainty and Behavioral Changes			✓	✓

Warnings

- The Promiscuous Pursuit of Data
  - Big Data reduces variance of the estimates, not their bias
  - Little sample uncertainty but Large model uncertainty
  - Do not fall in love with the Data: Transactional versus Aspirational
- The Human Element of the Data
  - Problem of identification
    - Who drowns more regularly at sea?
    - Happiness Index: Correlation and Causation
  - Measurement not Forecasting
- The Potential for Irrelevance
  - Problem of representativeness
    - Restaurant Reviews: Sampling Bias
    - Parental Guidance through Facebook
- Privacy is a First Order Problem
  - Aggregation does not solve the problem of privacy

The future of NSI's and CB's Data Collection

- Hybrid Approach: Change in National Statistics
  - Relevance and Timeliness
    - Demand driven (not supply driven)
  - Cost, Granularity, and Reducing Response Rates
    - Organic Data based indicators
  - Privacy
    - Aggregation is not enough
  - Organizational Paradigm
    - Geographical and Socio-economically organized; versus network and behaviorally organized.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

A personal view on big data and policymaking[1]

Naruki Mori,

Bank of Japan

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

"Building Pathways for Policy Making with Big Data"
IFC-Bank Indonesia Satellite Seminar on Bali, 26 July 2018

日本銀行
BANK OF JAPAN

A Personal View on Big Data and Policymaking

Naruki Mori,
Associate Director-General,
Research and Statistics Department
Bank of Japan
naruki.mori@boj.or.jp

The views in this presentation are my own and do not necessarily reflect those of the Bank of Japan.

Today's topics

1. Analytical use of big data

2. Possible use of big data for statistics

- A case study: big data for the price index

3. Useful for policymaking?

1. Analytical use of big data at the Bank of Japan

- Financial markets analysis
  - ✓ Analysis of the Japanese government bond (JGB) markets
    - Intraday market liquidity of JGB futures

      https://www.imes.boj.or.jp/research/papers/english/me34-3.pdf
    - Indicators measuring liquidity in the JGB markets (cash and futures)

      http://www.boj.or.jp/en/research/brp/ron_2018/data/ron180329a.pdf
- Payment and settlement system analysis
  - Effects of the improvement in payment and settlement system

    http://www.boj.or.jp/research/brp/psr/psrb160629.pdf (Japanese only)

⇒ High-frequency data such as tick data are increasingly used in the analysis of financial markets.

2-1. Possible use of big data for statistics

- We are exploring possibilities of using big data in compiling the existing statistics: for example,

✓ Securities statistics in the flow of funds accounts

- Use of micro (security-by-security) data sets with description on individual issues (e.g. interest rate, currency, maturity) available on the website of Japan Securities Depository Center (JASDEC)

✓ Producer Price Index

- Use of web scraped data for splitting price differences between old and new products into those due to quality changes and those due to pure price changes

New data sources are complementary to traditional data sources such as those compiled by national statistical agencies.

2–2. A case study: big data for the price index – Traditional and Non–traditional approaches –

Based on our staff's draft paper available at:
https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Japan.pdf

- We combine features of both traditional approach and non–traditional approach by applying machine learning methods in order to pair old and new products accurately. We use big data obtained from Japan's leading price comparison website *Kakaku.com* and **machine learning methods.**
- We can call our new method "Webscraped Prices Comparison Method (WSM)."

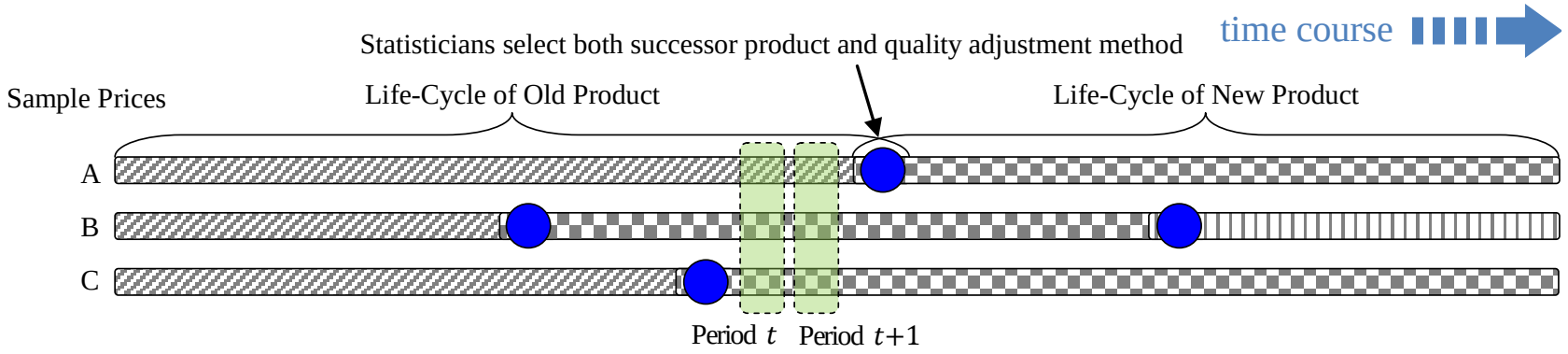
***Traditional* approach**: Price index which is created by carrying out changes of sample prices reflecting the product life cycles and quality adjustments between old and new products.

and

***Non–traditional* approach**: Price index which is compiled by making use of big data and computing capabilities.

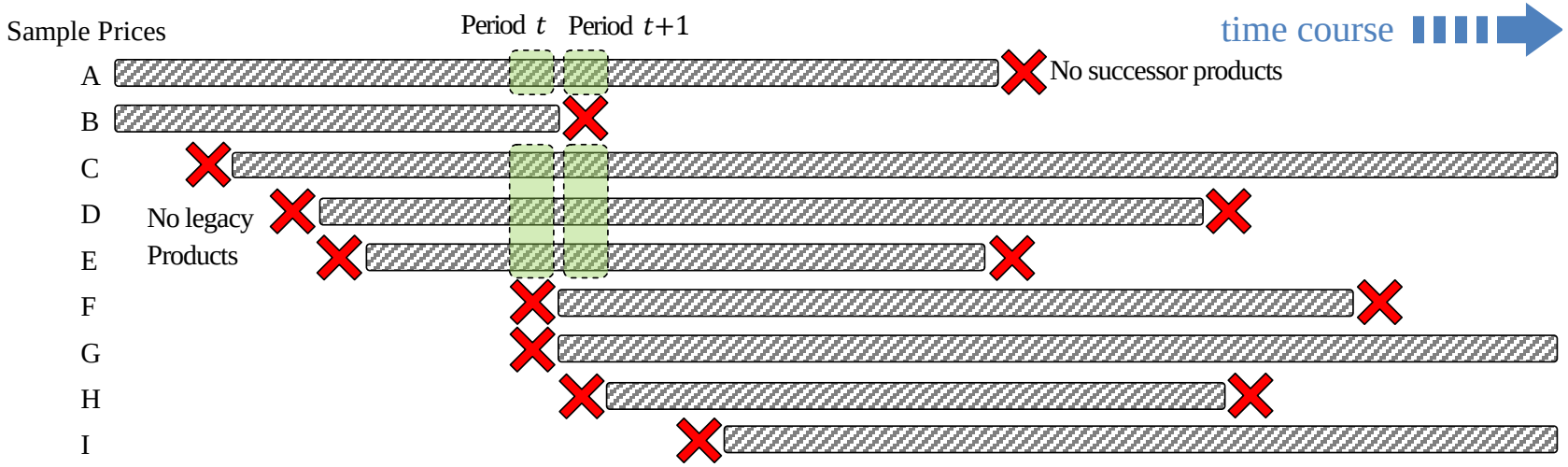
2–2. How to make old and new product pairs
– Traditional approach adopted by statistics agencies–

- Price statisticians select representative products to be surveyed, considering product specifications and data availability. At time of changing sample prices, they apply optimal quality adjustment to remove price change arising from changes of quality.
- Resources constraints at statistics agencies and reporting burden normally lead to a small number of sample prices.



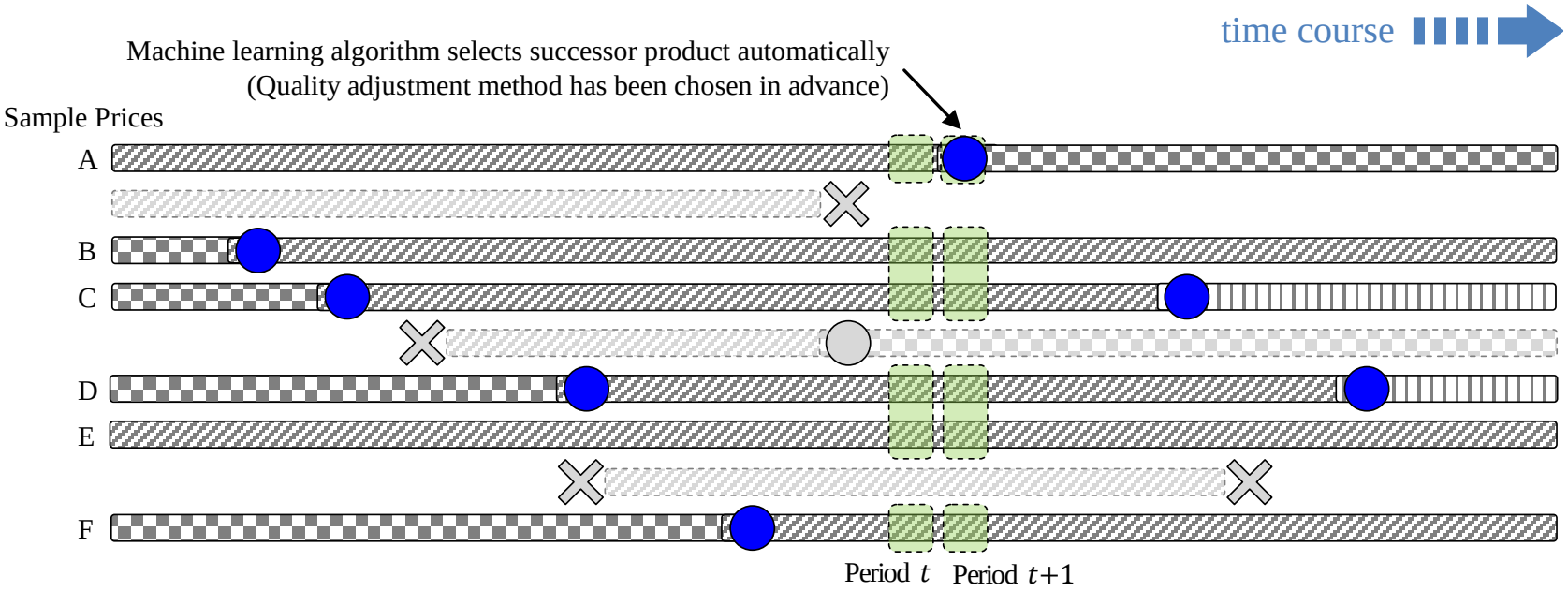
2–2. How to make old and new product pairs
– Non–traditional approach using big data –

- Compiling price index by using big data and **Matched–Model Method** (MMM) which calculates the percentage change of price for products which exist in both *survey period* and *following period*.
- If price pushbacks are constantly conducted when launching new products, the index cannot properly reflect the impact of such price pushbacks, and may cause a downward bias.



2–2. How to make old and new product pairs – A new method which adopts machine learning –

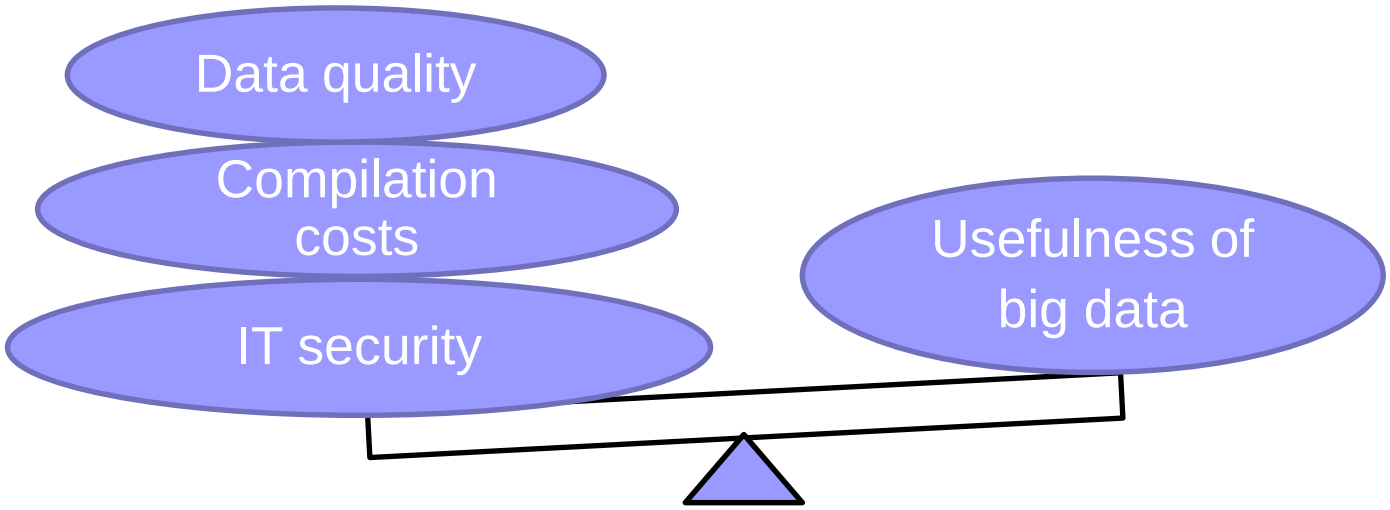
- We developed a **supervised machine learning** algorithm which pairs legacy and successor products with high precision to conduct appropriate quality adjustment for given big data.
- This index properly reflects the impact of price pushbacks.



2–3. Challenges for BOJ

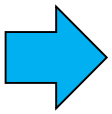
- Constraints on human and financial resources

- ✓ In order to utilize big data for economic research and statistics, we need to have staff with necessary expertise (e.g. data scientists who can make good use of machine learning methods).
- ✓ Handling big data requires not only significant resources but also data quality (reliability in data sources), methodological soundness, IT security, and proper arrangements for the protection of privacy and confidentiality. We need to strike a balance between costs and benefits.



3. Big data for policymaking?

- Inputs for Policy Board members
  - ✓ High-frequency data of prices (e.g. weekly, daily)
    - Daily price index compiled by scanner data (T-point price index)
    - Weekly purchasing price index of households
  - ✓ Text mining to gauge business sentiment
    - Text mining of Economy Watchers Survey compiled by Cabinet Office
    - Google Trends
- Readily available and high-frequency economic data may help policymakers to assess current economic developments timelier (e.g. nowcasting).



However, does the policymaking framework quickly adopt them? Otherwise, the use of big data is limited to providing background information for Policy Board members in making their assessments and decisions.

3. General remark: useful for policymaking?

- How can we fit big data into the policy reaction function? For example, how about the Taylor rule?

- At major central banks, price stability target (π^*) is headline CPI inflation.
- Taylor rule: $i_t = \pi_t + \alpha(\pi_t - \pi^*) + \beta y_t$

  IS curve: $y_{t+1} = \theta y_t - \sigma(i_t - E_t\pi_{t+1})$

  Phillips curve: $\pi_{t+1} = \gamma\pi_t + \delta y_t$

  where i = policy interest rate, π = CPI inflation, y = GDP gap
- It may suggest that CB's policy reaction function continue to be based on traditional data sources.

- Appropriate policymaking needs to be based on accurate economic data. Thus, policymakers and statistics agencies have been making efforts to improve accuracy of economic indicators (e.g. preliminary estimates (QE) of GDP, price index). What role will big data play in these efforts?

Thank you!

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Nowcasting New Zealand GDP using machine learning algorithms[1]

Adam Richardson, Thomas van Florenstein Mulder, Tugrul Vehbi,

Reserve Bank of New Zealand

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Nowcasting New Zealand GDP using machine learning algorithms

Adam Richardson[1], Thomas van Florenstein Mulder[2], Tuğrul Vehbi[3]

Abstract

We examine whether machine learning algorithms can improve nowcasts of real GDP growth in New Zealand. We use a large real-time dataset of around 550 New Zealand and international macroeconomic indicators. We train a range of popular machine learning algorithms over an expanding window and replicate an actual nowcasting situation starting from 2009 Q1 and moving forward a quarter at a time through to 2018 Q1. We compare the predictive accuracy of these nowcasts with that of other benchmarks such as a simple autoregressive model, a factor model, a large Bayesian VAR and a suite of statistical models used at the Reserve Bank of New Zealand. We find that the machine learning algorithms outperform the statistical benchmarks. Moreover, combining the nowcasts of the machine learning models leads to further improvements in performance. The results indicate that there are gains in nowcasting accuracy from using machine learning methods.

Keywords: Nowcasting, Machine learning, Forecast evaluation

JEL classification: C52, C53, C55,

[1] Manager, Modelling Team, Reserve Bank of New Zealand. Email: Adam.Richardson@rbnz.govt.nz

[2] Analyst, Modelling Team, Reserve Bank of New Zealand. Email: Thomas.vanFlorensteinMulder@rbnz.govt.nz

[3] Adviser, Modelling Team, Reserve Bank of New Zealand. Email: Tugrul.Vehbi@rbnz.govt.nz

We would like to thank the seminar participants at the Reserve Bank of New Zealand for their valuable comments. The views expressed here are the views of the authors and do not necessarily reflect the views of the Reserve Bank of New Zealand.

Contents

1. Introduction

Policy makers typically make decisions in real time using incomplete information on current economic conditions. Many key statistics are released with lags and are subject to frequent revisions. Nowcasting models have been increasingly popular tools developed to mitigate some of these uncertainties and they have been widely used by forecasters at many central banks and other institutions (Giannone et al. 2008, Banbura et al. 2013, Jansen et al. 2016, Bloor 2009).

Prompted by advances in computing power, machine learning (ML hereafter) methods have recently been proposed as alternatives to time-series regression models typically used by central banks for forecasting key macroeconomic variables. The ML models are particularly suited for handling large datasets when the number of potential regressors is larger than that of available observations.

In this paper, we investigate the performance of different ML algorithms in obtaining accurate nowcasts of the current quarter real gross domestic product (GDP) growth for New Zealand. We use multiple vintages of historical GDP data and multiple vintages of a large features set - comprising approximately 550 domestic and international variables - to evaluate the real-time performance of these algorithms over the 2009-2018 period. We then compare the forecasts obtained from these algorithms with the forecasting accuracy of a naive autoregressive benchmark as well as other data-rich methods such as a factor model, a small Bayesian VAR (BVAR) and a suite of statistical models used at the RBNZ. To our knowledge, our study is the first to evaluate the relative nowcast performance of alternative ML methods using real-time data.

Our results show that the majority of the ML models produce point nowcasts that are superior to the simple AR benchmark. The top-performing models such as the support vector machines, Lasso (Least Absolute Shrinkage and Selection Operator) and neural networks are able to reduce the average nowcast errors by approximately 16-18 per cent relative to the AR benchmark. Moreover, combining the nowcasts of the ML models using various weighting schemes leads to further improvements in performance. The majority of the ML algorithms also outperform the other two commonly used statistical benchmarks, namely the factor model and the small Bayesian VAR model.

Our contributions in this paper are twofold. First, we provide new evidence on the *real-time* nowcast performance of commonly used ML algorithms by comparing them to other useful benchmark models commonly used at central banks. Second, our results have important repercussions for policy as the accuracy of GDP nowcasts play a big role in correctly assessing the overall health of the economy.

This paper joins a growing literature that evaluates the relative success of the ML models in forecasting over the more traditional time-series techniques. However, to our knowledge, none of these papers focusses on the real-time forecasting performance of the models. Makridakis et al. (2018) compares the forecast accuracy of various popular ML algorithms with eight types of traditional statistical benchmarks and finds that the out-of-sample forecasting accuracy of ML models is lower than that of more traditional statistical methods. Chakraborty and Joseph (2017), on the other hand, conduct an out-of-sample forecasting exercise using UK data and argue that ML models generally outperform traditional modelling approaches in prediction tasks. Similarly, Kim (2003) finds that support vector machines are a promising alternative for predicting stock market variables.

To the extent, we use a large dataset for New Zealand for our analysis, our paper is also related to Eickmeier and Ng (2011) who use elastic net and ridge regression (amongst other shrinkage methods) to produce macroeconomic forecasts from a large number of domestic and international predictors. They find data-rich methods result in gains in forecast accuracy over common statistical methods using small data sets. Also, Matheson (2006) uses a factor model to produce macroeconomic forecasts from a large number of predictors. This model results in good forecast performance at longer-term horizons when compared to other statistical models and the RBNZ's own forecasts.

The remainder of this paper is as follows. Section 2 explains the models and the data used. Section 3 presents the results and Section 4 concludes.

2. Empirical Application

In this section, we provide a brief description of the various ML and benchmark models we considered for nowcasting GDP.

2.1 Models

i. Autoregressive Model (AR)

As our simple benchmark, we use a univariate AR model of order 1 for quarterly GDP growth (y_t):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t$$

where $\alpha_0$and $\alpha_1$are parameters and u_t is the residual term.

ii. K-Nearest Neighbour Regression (KNN)

KNN is a widely used non-parametric method that uses a distance function, in our case the Euclidean distance, to measure the similarity between the data used for training and testing purposes. The algorithm finds which k most observations in the training set have features most similar to the features of the hold out set and predict the outcome of the new data to be the outcome of the average of the k observations previously defined. The prediction is based on the mean of the k –most similar instances. We choose k using the grid search method available in Python's Scikit-Learn library.

iii. Boosted Trees (BT)

This is an ensemble method which allows converting a set of weak learners into one high-quality predictor. The general idea is to impute a sequence of simple trees, where each successive tree is built for the prediction residuals of the preceding tree. The algorithm works by dividing the predictor space into a set of possible values for $\{x_1, x_2, \dots, x_i\}$ where x_i corresponds to the feature set of observation i. Starting with the predictor whose splitting leads to the greatest reduction in the sum of squared residuals, the algorithm then selects the next feature whose split minimises the sum of squared residuals given the split of the initial feature. This process continues until the reduction in the sum of squared residuals falls below a certain threshold. We use gradient tree boosting (GBRT) which is a forward stagewise methodology to add regression tree models one after the other to increase predictive ability. The added model is chosen by minimising a given loss function. GBRT continues this additive process until the gains in predictability from adding new models drops below some threshold value.

iv. Lasso, Ridge and Elastic Net (ENET)

These three methods are very similar to ordinary least squares (OLS) but incorporate different types of shrinkage for creating parsimonious models in presence of a large number of features. Lasso performs L1 regularisation which involves adding a penalty equivalent to the absolute value of the magnitude of the coefficients and shrinking some of them to zero. Ridge is a similar method to lasso but it performs L2 regularisation where the penalty is on the squared value of the magnitude of the coefficients. Therefore the coefficients estimated by ridge are never reduced to exactly zero. The elastic net regression is a hybrid of the ridge and lasso regressions and as such, it can shrink the coefficients of the features as well as removing them completely (a coefficient of zero). The regression penalty is, therefore, a convex sum of the ridge and lasso penalties.[4]

$$\beta^{Lasso} = argmin\left[\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right) + \lambda\sum_{j=1}^{p}\left(\alpha\beta_j^2 + (1-\alpha)\left|\beta_j\right|\right)\right]$$

The equation above shows the ridge case when α = 1, the lasso case when α = 0, and the elastic net case for any values in between. α and λ are hyperparameters that determine the relative impact of the penalty terms. When λ = 0, we have the standard ordinary least squares (OLS) approach.

v. Support Vector Machine Regression (SVM)

Support vector machine regression model is a non-parametric method first proposed by Vapnik (1995). The idea of SVM is to find a linear function of the form

$$f(x) = \langle w, x\rangle + b$$

where w is the weight vector, b is the bias and x is the input or feature vector while ensuring that the function is as flat as possible meaning one seeks a small w. One way to implement this is to minimize the norm, i.e. $\|w\|^2$. This can be formulated as a convex optimization problem:

$$0.5\|w\|^2 + C\sum_{i=1}^{l}\left|(y_l - f(x_l))\right|_\epsilon$$

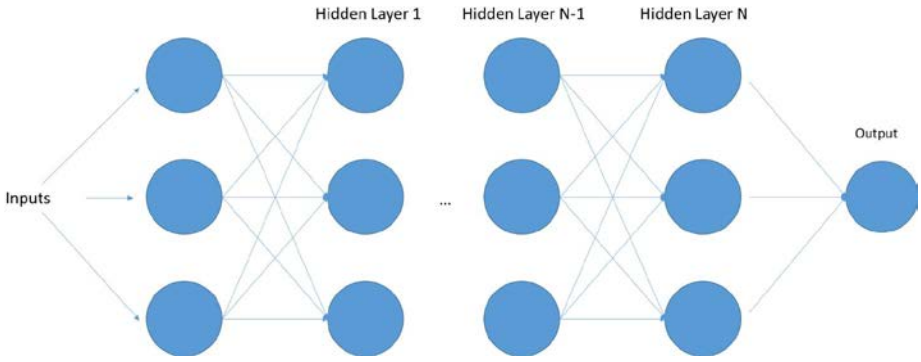
[4] For a detailed explanation of the elastic net refer to (Zou & Hastie 2005)

where C > 0 is the regularisation parameter. The first term in the error function is a penalty term that increases as the model becomes more complex. The second term is the ϵ-insensitive loss function that penalises errors that are greater than ϵ, allowing flexibility to the model. For estimating the model, we use Matlab's Statistics and Machine Learning Toolbox which implements linear epsilon-insensitive SVM (ε-SVM) regression, which is also known as L1 loss.

vi. Neural Network (NN)

Neural network model is a model that is able to capture and represent complex relationships. The model works by taking input data and using weights and an activation function to pass them through to N hidden layers of a perceptron as shown in figure 1. Each input is weighted and passed through the activation function to determine the value of a given node within the first layer of perceptrons and this is repeated for each node in the first layer. Each node of the first layer then becomes the new input variable for layer 2 and gets reweighted and passed through the activation function to determine the value of a given node in layer 3. The process is repeated until the N^{th} layer is created. The nodes in the N^{th} layer are weighted and passed through an activation function to give the output value.

Figure 1: Multi Layer Perceptron Process



The weights are initially set with random values and are updated on each iteration using an algorithm called backpropagation. With backpropagation, the input data is repeatedly presented to the neural network where the output of the neural network and the desired output is produced.[5]

[5] For more details on backpropagation see Hecht-Nielsen 1992

vii. Factor Model (FM)

Factor models use a large number of time series to produce forecasts, and therefore do not require the model builder to make strong assumptions about what particular series are important for forecasting the variable of interest. We estimate a linear factor model which assumes that the quarterly growth rate of GDP is given by

$$y_t = \alpha_0 + \alpha_1 f_1 + \cdots + \alpha_k f_k + u_t$$

where the $f_j, j = 1, \ldots, k,$ are the common factors obtained using the principal components technique. These factors are the linear combinations of all the data in the model that explain the highest proportion of the variance of the data. They can, therefore, be thought of as picking up the underlying movements in the economy that influence a large number of variables. The estimated factors are used in linear regressions of the variables of interest. We choose the optimal number of factors to incorporate in (8) using the Bai and Ng (2002) two-step procedure and use the Bayesian Information Criteria (BIC) as the benchmark for selection.

viii. Bayesian VAR (BVAR)

This model is part of the statistical models suite at the RBNZ and utilises 95 macroeconomic time series in a Bayesian VAR framework. The model produces forecasts for key macroeconomic variables such as GDP, consumption, inflation, the exchange rate and interest rates. We use the quarterly GDP growth rate forecasts from this model as the basis for our comparisons.

ix. RBNZ Statistical Suite (RBNZ SS)

The suite contains a range of different models that vary across size and complexity (Bloor,2009). The models in the suite are particularly designed to forecast medium-term movements in the economy and are used as a cross check for the central forecasts produced by the main policy model. The suite also includes several models that are aimed towards picking up shorter-run fluctuations.

2.2 Forecast evaluation methodology

We evaluate the performance of the models by means of an out-of-sample forecast exercise. We train each algorithm over an expanding window thereby replicating an actual forecasting situation starting from 2009 Q1 and moving forward a quarter at a time through to 2018 Q1. For example, for the first vintage of the data, the models

are estimated over the period 1995Q1 to 2008Q4 using real-time data for both the predictor and response variables. The resultant fitted models are used to nowcast the 2009 Q1 growth rate of real GDP. Overall, we generate 37 real-time nowcasts of quarterly GDP growth. We choose the default parameter settings for each algorithm. Next, we measure the forecast accuracy of each model by calculating the Root Mean Square Error (RMSE) and the Mean Absolute Deviation (MAD) defined as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y_t})^2}{T}}$$

$$MAD = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y_t}|$$

where y_t and $\hat{y_t}$ are the actual and forecast values of GDP growth and T is the total number of forecasts. The forecasts of a univariate (i.e. AR(1)) model provide the main benchmark for our comparisons. We use the Diebold-Mariano (Diebold & Mariano 1995) test to determine whether the forecasts obtained from each ML model are significantly different than those from the AR model.

2.3 Data

The data consist of a number of real-time vintages of a range of macroeconomic and financial market statistics. These include: New Zealand business surveys; consumer and producer prices; general domestic activity indicators (e.g. concrete production, milk-solids production, spending on electronic cards etc.); domestic trade statistics; international macroeconomic variables and international and domestic financial market variables. The data range from daily to quarterly - and are aggregated to quarterly data for model estimation. The series are seasonally adjusted. Each series is individually assessed, and either left in a level form or transformed to a stationary form depending on which form is likely to be more predictive of GDP growth. We also remove the series that contain missing values.

The storage of historical model runs has allowed us to create 37 real-time vintages of these data. Every 3 months, the RBNZ publishes a *Monetary Policy Statement*. In working towards this publication, the RBNZ's staff put together an initial set of macroeconomic projections. The data banks containing the data described above were saved down along with these projections each quarter. This process, therefore, gives us a quarterly real-time snapshot of a range of macroeconomic and financial market series. Conveniently, these snapshots were generally taken about four weeks before the release of the preceding quarters GDP estimate. For example, the initial projections for the March 2015 *Monetary Policy Statement* would have been finalized on about the 20th of February. At this point, almost all of the key macroeconomic and financial market indicators for the December 2014 quarter would have been released - and it is at this point the snapshot of these macroeconomic statistics has been saved. The December quarter GDP estimate was then released 19th

March. New Zealand does not produce flash estimates of GDP, and so there is a significant lag between the end of the quarter and the publication of the GDP estimate. This process gives us a set of data vintages from 2009 Q1 to 2018 Q1.

From this point, the data storage methodology was changed. The 'global database' containing 668 series (the detail of which I described above) was routinely saved in estimating the Bank's suite of statistical models. A version of this data set is saved at the end of the month following each *Monetary Policy Statement*. For example, for the May 2018 *Monetary Policy Statement*, a version of the data set was saved on the last working day of May. This data set contains most of the indicators up to the end of 2018 Q1. The 2018 Q1 GDP figures were then released on the 21 June 2018 - around 3 weeks after the data snapshot was taken. This process gives us a set of data vintages from 2015 Q3 to 2018 Q1.

The data available with each vintage differ somewhat, as data were added and removed from the RBNZ's data banks through time. After making the modifications described above, from a candidate of 668 series, we are left with between 532-634 series at each vintage.

In total, we have a 37 real-time vintages of this dataset, covering the period 2009 Q1 to 2018 Q1. The data in each vintage begin in 1995 Q1. These data vintages enable us to test how the forecast performance of these models compares under the conditions which the practitioner would use them - capturing the revision properties of the predictors.

3. Empirical Results

In this section, we describe the main results of our analysis. Table 1 presents data on the nowcast performances of the models for the sample period 2009 Q1-2018 Q1. In addition to the models outlined in Section 2.1, we also present the results obtained by combining the forecasts from all the ML models using alternative weighting schemes.

The results indicate that the large majority of the ML models produce forecasts that have RMSEs and MADs lower than the AR benchmark. The top three models are the SVM, Lasso and NN models which are able to reduce the average forecast errors by approximately 16-18 per cent relative to the AR benchmark. The relative success of the neural network and support vector machine models are not surprising and is in line with previous findings in the literature (Teräsvirta et al. 2005, Ahmed et al. 2010). The majority of the ML models are also able to produce RMSEs and MADs lower than the BVAR, factor model and the combination forecasts obtained from RBNZ's statistical suite. However, the DM test results based on the RMSE loss function indicate that in most cases, the null hypothesis that the forecast errors are equal cannot be rejected. The null hypothesis of equal forecast errors based on the mean absolute deviations, on the other hand, is rejected for the case of SVM, NN and the Lasso at the conventional levels of statistical significance. The DM test results, however, should be treated with caution given our small sample size of 37 observations.

It is important to note that some of these models come with the added costs of increased computational time, and a lack of tractability when it comes to the drivers of certain results. This could be significant in practice in two regards. First, it may limit the practical use of such models in situations that require a quick turnaround. Second, if the forecast accuracy of a model started to deteriorate, it may be difficult to pick apart the factors leading to such a deterioration. The most computationally intensive model is the NN model, taking approximately 31 minutes to solve. All the other models are relatively less intensive taking only several seconds to solve.[6]

Figure 2 presents the quarterly GDP growth and its nowcasts obtained from each model over the 2009 Q1-2018 Q1 period. It can be seen that all ML models have successfully predicted the sharp downturn in activity occurred in the first quarter of 2009 and also predicted the other major upturns and downturns in the GDP data successfully.

Table 1: Real-time nowcast performances of models (RMSE), 2009Q1-2018Q1

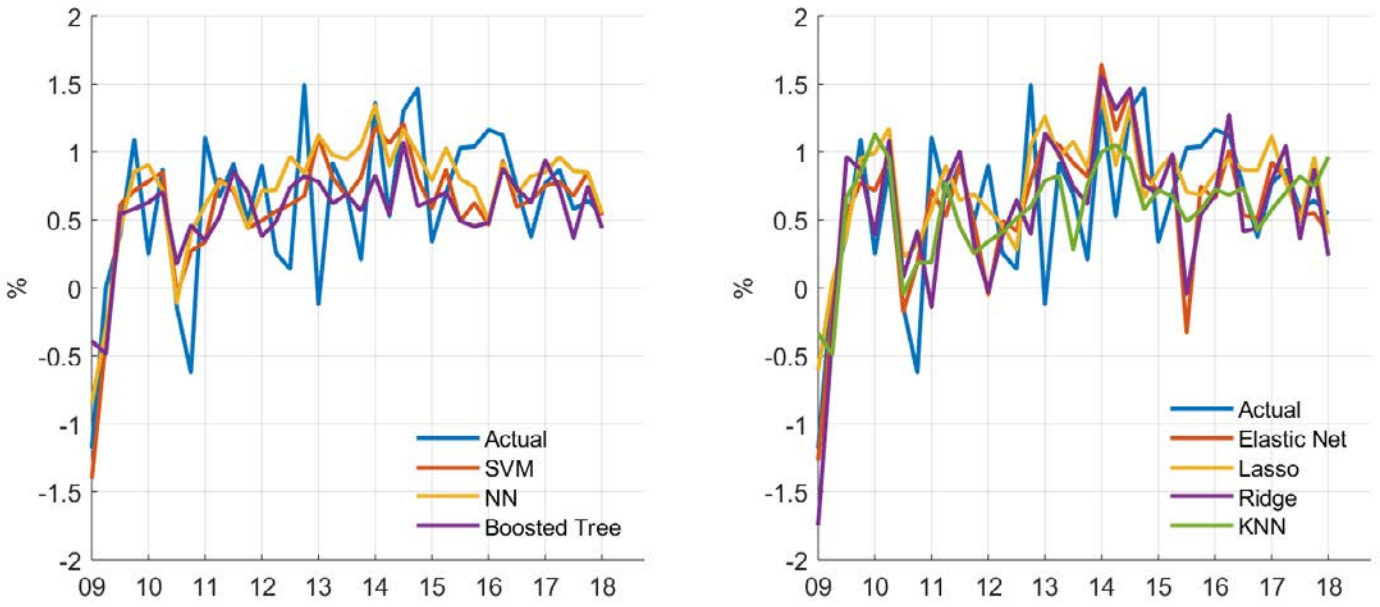
Models	RMSE	RMSE (Rel. to AR)	p-value	MAD	MAD (Rel. to AR)	p-value
SVM	0.445	0.820	0.166	0.338	0.770	0.037
NN	0.446	0.821	0.182	0.336	0.766	0.041
Lasso	0.454	0.836	0.206	0.348	0.793	0.061
BT	0.469	0.865	0.153	0.386	0.880	0.107
ENET	0.488	0.899	0.526	0.349	0.796	0.140
KNN	0.491	0.905	0.349	0.405	0.923	0.402
Ridge	0.565	1.040	0.770	0.427	0.972	0.823
FM	0.517	0.951	0.757	0.379	0.864	0.357
BVAR	0.608	1.119	0.286	0.450	1.025	0.961
RBNZ SS	0.471	0.867	0.262	0.384	0.874	0.160
AR	0.543	-	-	0.439	-	-

Notes: The first column refers to the models used to nowcast GDP. SVM: support vector machine; NN: neural network; ENET: elastic net; Lasso: lasso regression; BT: boosted tree; FM: factor model; Ridge: ridge regression; AR: autoregressive model; BVAR: Bayesian VAR; KNN: k-nearest neighbours; RBNZ SS: the statistical models suite used by the RBNZ. The second column refers to the entries for the out-of-sample RMSEs obtained from each

[6] These models were estimated using a system with 64 GB of RAM, a Xeon E5-1660, 8 core, 3.20 GHz CPU, with a Windows 10 operating system. It is important to note that increasing the number of training iterations for optimising the hyperparameters would increase these computation times significantly. The computation time of these models can be further addressed by making use of cloud computing services.

model using the methodology outlined in subsection 2.2. The third column refers to the RMSEs relative to the AR model. The fourth column refers to the p-values obtained from the Diebold-Mariano test for testing the significance of the forecast accuracy of each method versus that of the AR model. Columns 5-7 refer to the corresponding values in columns 2-4 when the loss function is the Mean Absolute Deviation (MAD).

Figure 2: Real-time nowcasts of quarterly GDP growth



3.1 Forecast combination

In the previous section, we compared the forecasts from individual models by ranking them individually according to their forecast accuracy. From a practical point of view, however, we may prefer to pick the "best" model amongst them to use for nowcasting. Therefore, an alternative approach is to combine forecasts from the set of all models under consideration to produce a single summary forecast. Forecast combinations have frequently been found in the literature to produce better forecasts than individual models. In this section, we implement this strategy using all the ML models we've considered. More specifically, we use four types of forecast combination strategies for combining point forecasts: equal weighting, Least-squares weighting, inverted mean squared error (MSE) and MSE ranks weighting. Equal weighting is a particularly simple method which works by assigning equal weights to the forecasts from all individual models at each date in the forecast sample. The Least-squares weighting strategy, on the other hand, is implemented by regressing all the forecasts against the actual values and then using the coefficients from the resultant regression as weights. The final two strategies both small use the inverted MSEs computed over the forecast horizon for weighting the forecasts where the latter is based on the inverted MSE ranks rather than the actual MSE values.

The results summarised in Table 2 suggest that there are gains in combining forecasts. It can be seen that the combined forecasts produce lower RMSEs and MADs compared to the individual model results presented in Table 1. Amongst the different weighting schemes we've considered, the Least-squares weighting scheme generates the best gains in predictive accuracy.

Table 2: Forecast combination results

	RMSE	RMSE (Rel. to AR)	MAD	MAD (Rel. to AR)
Equal weighting	0.437	0.805	0.324	0.738*
Least-squares weighting	0.427	0.786	0.317	0.722*
Inverted MSE weighting	0.434	0.799	0.324	0.738*
Inverted MSE ranks weighting	0.429	0.790	0.323	0.736*

Notes: The first column refers to the weighting methods used to combine the individual ML forecasts. Equal weighting: assigning equally weighted forecasts; Least-squares weighting: weights are assigned by regressing all the forecasts against the actual values and then using the coefficients from the resultant regression as weights neural network; Inverted MSE weighting: weights are assigned using the inverted MSEs as weights; Inverted MSE ranks: weights are assigned using the inverted MSE ranks as weights. See Table 2for column definitions. The * indicates statistical significance of the Diebold-Mariano test at the 5% significance level.

Furthermore, we investigate whether the optimal combination of the ML model nowcasts (i.e. the Least-squares weighting) adds value to the nowcasts generated by the combination of models in the RBNZ's statistical model suite.

To test this formally, we follow the approach by Romer and Romer (2008) and estimate the following regression equation:

$$y_t = c + \alpha_1 FC + \alpha_2 SS + e_t$$

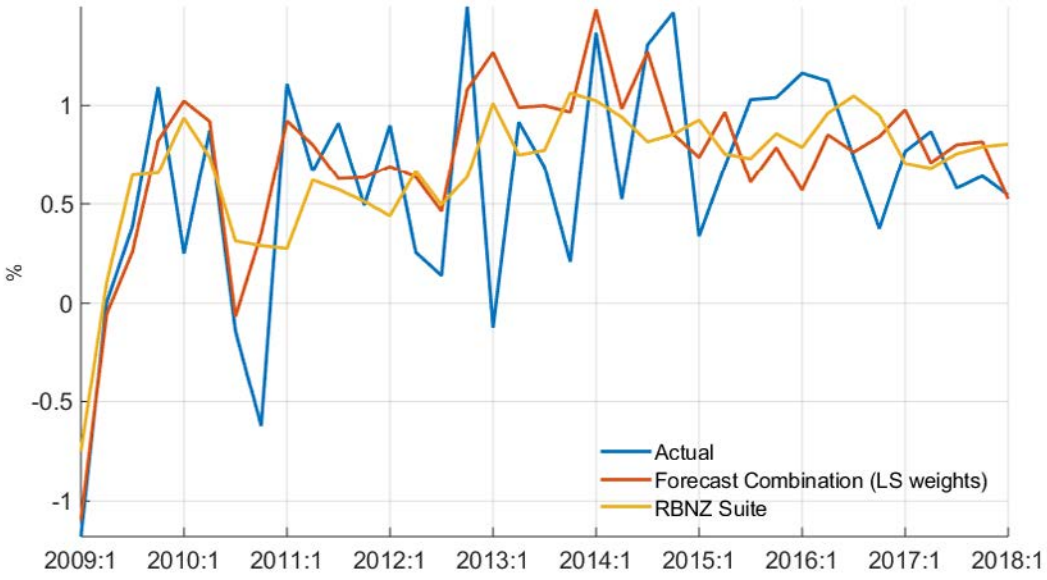
where y_t is the real-time GDP nowcasts, FC is the combined forecasts using the least-squares weights method outlined above, SS is the real-time GDP nowcasts obtained from the combination of models in the RBNZ statistical suite and e_t is the residual term. The results presented in Table 3 suggest that forecast combination adds significant value to the combined statistical-suite nowcasts as implied by the large, positive and statistically significant α_1 coefficient. Figure 3 presents the quarterly GDP growth together with these nowcasts over the period 2009Q1-2018Q1. These results

suggest that ML models would be a useful addition to the RBNZ's current suite of forecasting models.

Table 3: Does forecast combination add value to the RBNZ's statistical-suite nowcasts?

Coefficients	Estimated value	t-Statistic
α_1	0.94	4.11
α_2	-0.11	-0.31
C	0.02	0.13
R^2	0.46	

Figure 3: Quarterly GDP growth and its nowcasts (2009Q1-2018Q1)



4. Conclusion

In this paper, we evaluate the real-time performance of popular ML algorithms in obtaining accurate nowcasts of the real gross domestic product (GDP) growth for New Zealand. We estimate several ML models over the 2009-2018 period using multiple vintages of historical GDP data and multiple vintages of a large features set comprising approximately 550 domestic and international variables. We then compare the forecasts obtained from these models with the forecasting accuracy of a naive autoregressive benchmark as well as other data-rich methods such as a factor model, a large Bayesian VAR and the combined GDP nowcasts obtained from the suite of statistical models used at the RBNZ. We find that the majority of the ML models are able to produce more accurate forecasts than those of the AR and other statistical benchmarks. The results also suggest that there are some gains in combining various ML forecasts. Our results thus recommend the use of ML algorithms as a useful addition to a forecaster's suite of GDP nowcasting models.

References

Ahmed, N. K., Atiya, A. F., Gayar, N. E. & El-Shishiny, H. (2010), 'An empirical comparison of machine learning models for time series forecasting', Econometric Reviews 29(5-6), 594–621.

Banbura, M., Giannone, D., Modugno, M. & Reichlin, L. (2013), 'Nowcasting and the real-time data flow', Handbook of economic forecasting 2(Part A), 195–237.

Bloor, C. (2009), 'The use of statistical forecasting models at the Reserve Bank of New Zealand', Reserve Bank of New Zealand Bulletin 72, 21–26.

Chakraborty, C. & Joseph, A. (2017), Machine learning at central banks, Bank of England working papers 674, Bank of England.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing Predictive Accuracy', Journal of Business & Economic Statistics 13(3), 253–263.

Eickmeier, Sandra & Ng, Tim. (2011), 'Forecasting national activity using lots of international predictors: An application to New Zealand,' International Journal of Forecasting, Elsevier, vol. 27(2), pages 496-511.

Giannone, D., Reichlin, L. & Small, D. (2008), 'Nowcasting: The real-time informational content of macroeconomic data', Journal of Monetary Economics 55(4), 665–676.

Hecht-Nielsen, R. (1992), Theory of the backpropagation neural network, in 'Neural networks for perception', Elsevier, pp. 65–93.

Jansen, W. J., Jin, X. & de Winter, J. M. (2016), 'Forecasting and nowcasting real gdp: Comparing statistical models and subjective forecasts', International Journal of Forecasting 32(2), 411 – 436.

Kim, K. (2003), 'Financial time series forecasting using support vector machines', Neurocomputing 55(1), 307 – 319. Support Vector Machines.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2018), 'Statistical and machine learning forecasting methods: Concerns and ways forward', PloS one 13(3).

Matheson, T. D. (2006), 'Factor Model Forecasts for New Zealand', International Journal of Central Banking, International Journal of Central Banking, vol. 2(2), May.

Teräsvirta, T., van Dijk, D. & Medeiros, M. C. (2005), 'Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination', International Journal of Forecasting 21(4), 755 – 774. Nonlinearities, Business Cycles and Forecasting.

Vapnik, V. N. (1995), The Nature of Statistical Learning Theory, Springer-Verlag, Berlin, Heidelberg.

Zou, H. & Hastie, T. (2005), 'Regularization and variable selection via the elastic net', Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301–320.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Nowcasting New Zealand GDP using machine learning algorithms[1]

Adam Richardson, Thomas van Florenstein Mulder, Tugrul Vehbi,

Reserve Bank of New Zealand

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Nowcasting New Zealand GDP using Machine Learning algorithms

Adam Richardson, Thomas Mulder, ***Tugrul Vehbi***

RESERVE BANK
OF NEW ZEALAND
TE PŪTEA MATUA

Motivation

- Applications using machine learning are transforming our daily life
- ML methods have recently been proposed as alternatives to time-series regression models typically used by central banks
- We want to know how useful are these models for nowcasting New Zealand GDP?

The evidence is mixed…

- Makridakis et al. (2018)
  - out-of-sample forecasting accuracy of ML models is lower than that of more traditional statistical methods
- Chakraborty and Joseph (2017)
  - ML models generally outperform traditional modelling approaches in prediction tasks
- Matheson (2006)
  - Factor Model Forecasts for New Zealand.
- Eickmeier, & Ng (2011)
  - Forecasting national activity using lots of international predictors: An application to New Zealand.

Our approach

- We train a range of popular machine learning algorithms over an expanding window

- Using real-time data, we replicate an actual nowcasting situation starting from 2009 Q1 through to 2018 Q1

- We compare the predictive accuracy of these nowcasts with other benchmarks
  - AR(1), a factor model, a large Bayesian VAR and a suite of statistical models used at the Reserve Bank of New Zealand.

The models

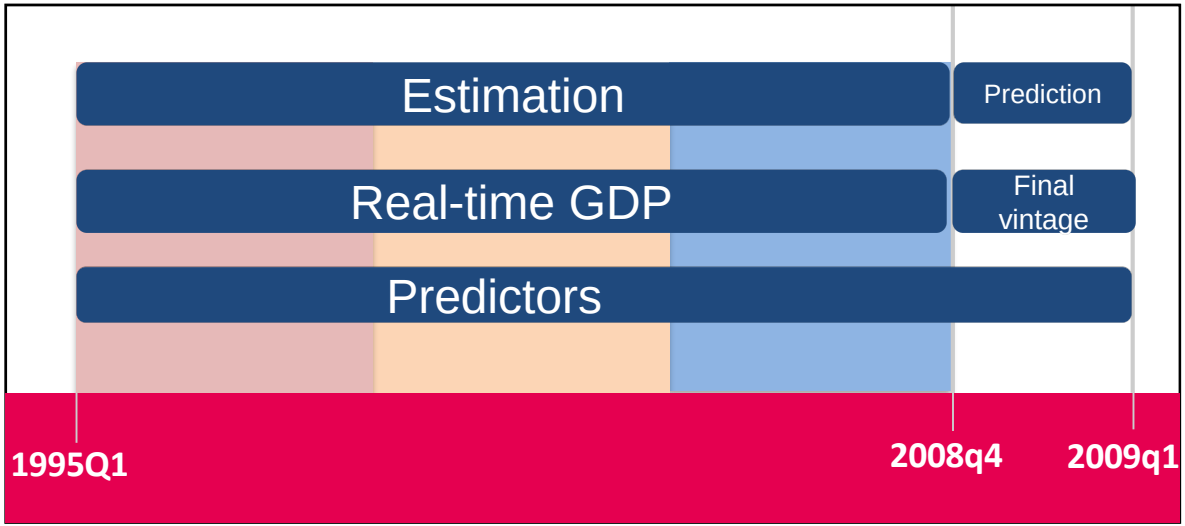
- **Machine Learning models:**
  - K-Nearest Neighbour (KNN)
  - Boosted Trees (BT)
  - Lasso, Ridge and Elastic Net (ENET)
  - Support Vector Machines (SVM)
  - Neural Network (NN)

- **Statistical benchmarks**
  - AR model
  - Factor model (FM)
  - Large BVAR
  - RBNZ statistical suite

The data

- Real-time data
- 2009q1 to 2018q1 (37 quarters)
- 1995q1 – current
- Quarterly, stationary, non-NaN
- About 550 macro indicators

Estimation example (2009q2)



Forecast evaluation

Root Mean Squared Error

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T}}$$

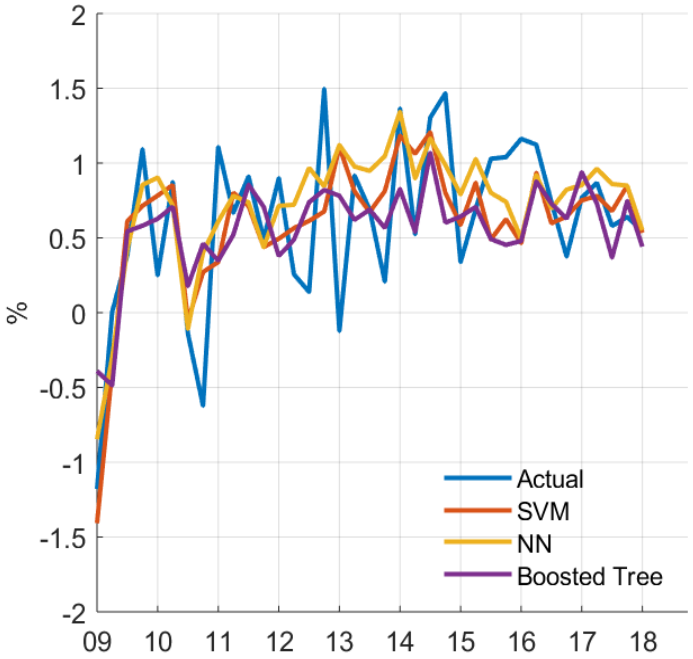
Mean Absolute Deviation

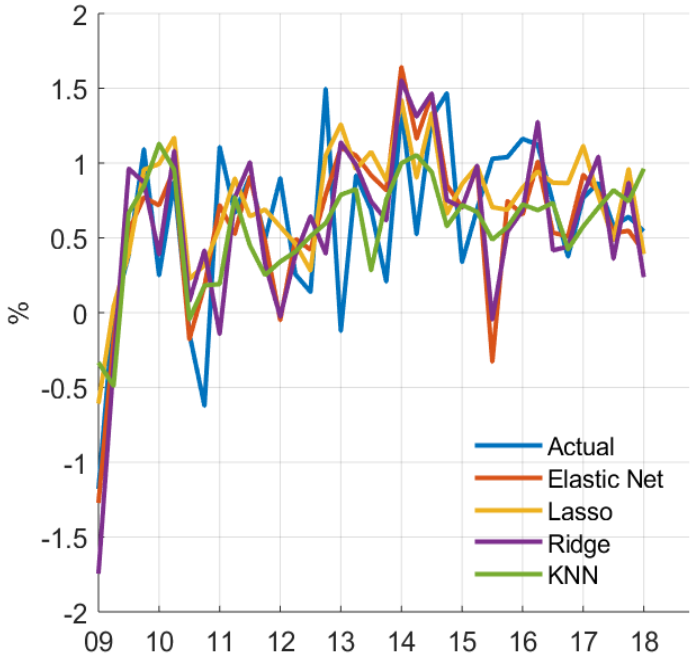
$$MAD = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t|$$

Results

	RMSE	RMSE (rel to AR)	MAD	MAD (rel to AR)
SVM	0.445	0.820	0.338	0.770*
NN	0.446	0.821	0.336	0.766*
Lasso	0.454	0.836	0.348	0.793*
BT	0.469	0.865	0.386	0.880*
ENET	0.488	0.899	0.349	0.796
KNN	0.491	0.905	0.405	0.923
Ridge	0.565	1.040	0.427	0.972
FM	0.517	0.951	0.379	0.864
BVAR	0.608	1.119	0.450	1.025
RBNZ SS	0.471	0.867	0.384	0.874
AR	0.543	-	0.439	-

Real-time nowcasts of quarterly GDP growth (2009-2018)





Forecast combination

	RMSE	RMSE (rel. to AR)	MAD	MAD (rel. to AR)
Equal weighting	0.437	0.805	0.324	0.738*
Least-squares weighting	0.427	0.786	0.317	0.722*
Inverted MSE weighting	0.434	0.799	0.324	0.738*
Inverted MSE ranks weighting	0.429	0.790	0.323	0.736*

Does forecast combination add value to the RBNZ's statistical-suite nowcasts?

We follow the approach by Romer and Romer (2008) and estimate the following regression equation:

$$y_t = c + \alpha_1 FC + \alpha_2 SS + e_t$$

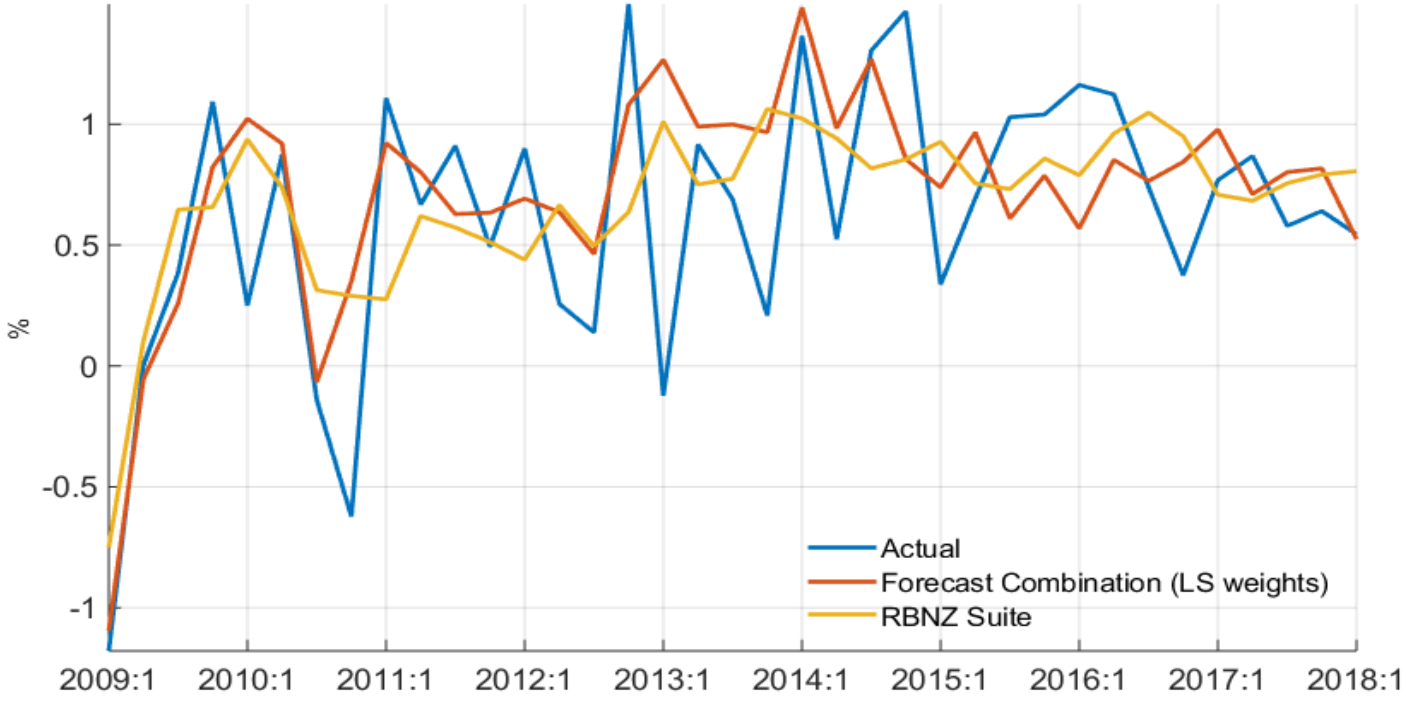
FC: combined nowcast using LS weights

SS: RBNZ statistical suite nowcast

ML forecast combination adds significant value

Coefficients	Estimate	t-Statistic
α_1	0.94	4.11
α_2	-0.11	-0.31
C	0.02	0.13
R^2	0.46	

Forecast combination vs RBNZ suite



Conclusion

- Machine learning algorithms outperform the statistical benchmarks
- Combining the nowcasts of the ML models leads to further improvements in performance.
- ML models would be a useful addition to the RBNZ's current suite of forecasting models.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Google econometrics: Nowcasting euro area car sales and big data quality requirements[1]

Per Nymand-Andersen,

European Central Bank

The paper has been published as an ECB statistics working paper

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANK INDONESIA
BANK SENTRAL REPUBLIK INDONESIA

EUROPEAN CENTRAL BANK
EUROSYSTEM

Irving Fisher Committee on
Central Bank Statistics
BANK FOR INTERNATIONAL SETTLEMENTS

Per Nymand-Andersen
Adviser, European Central Bank

Google econometrics: Nowcasting euro area car sales and quality requirements

Picture: INDOLINK Consulting's blog

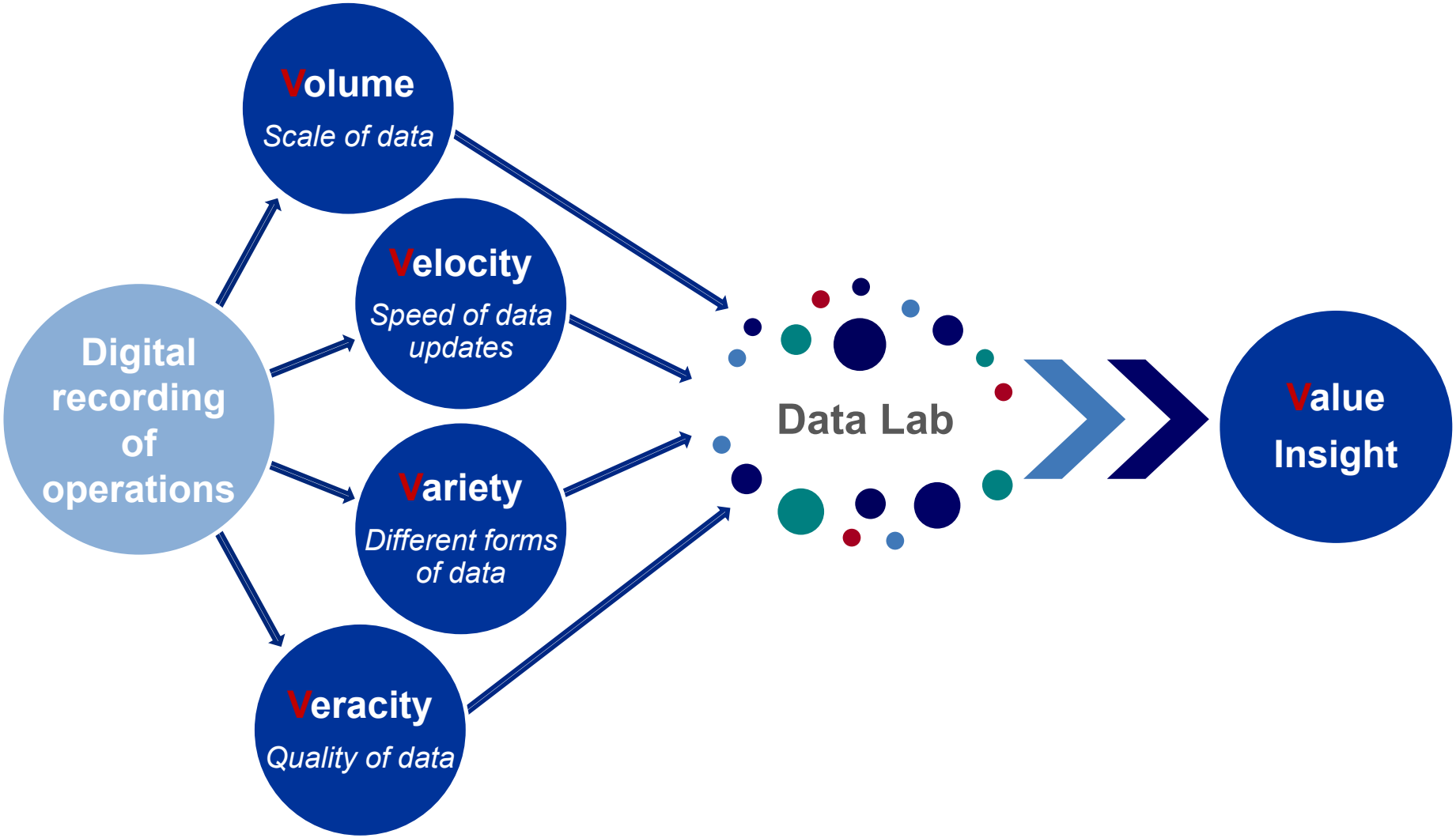
Bank Indonesia - IFC/BIS Workshop on big data for central bank policies, 23-25 July 2018

Overview

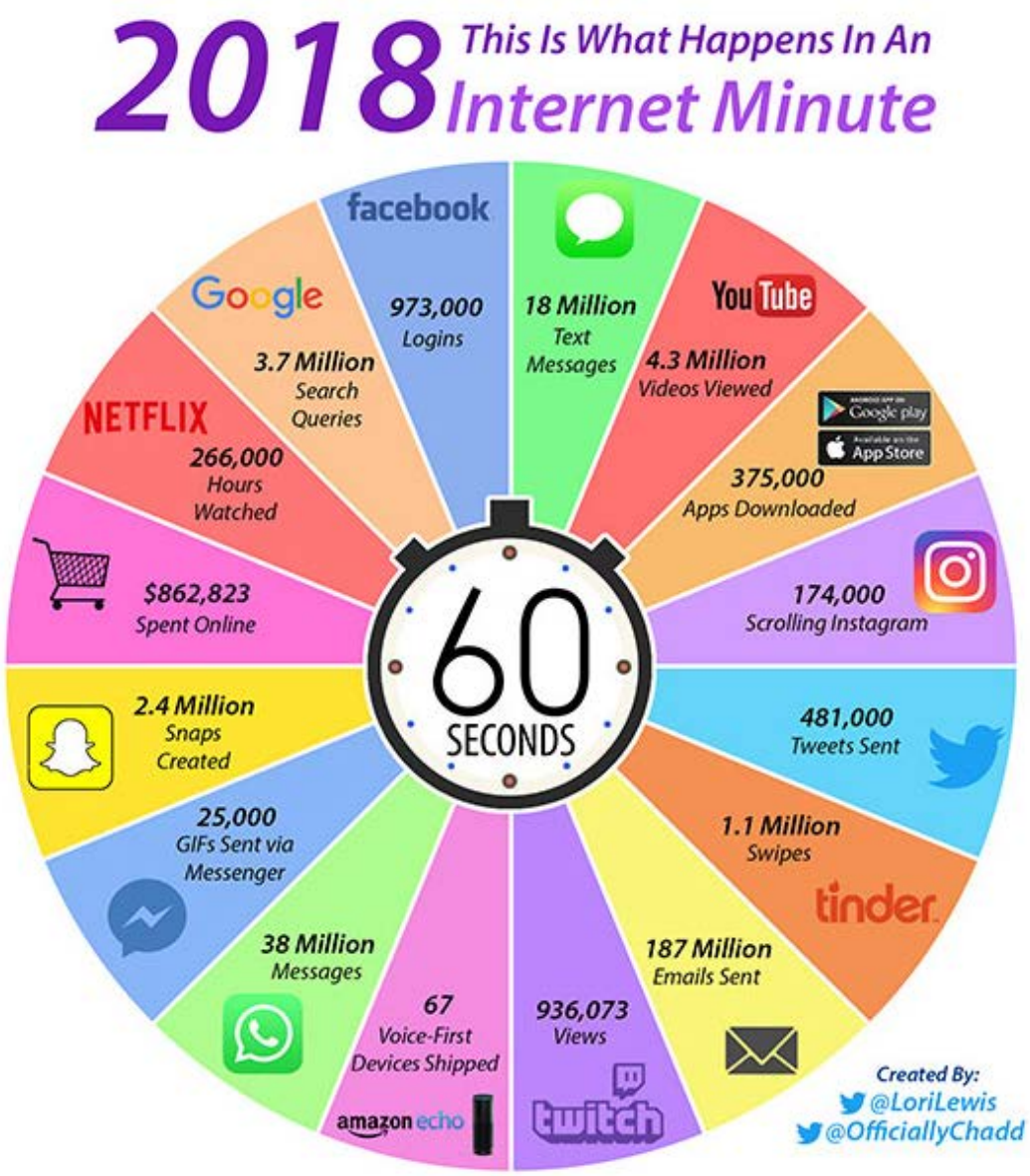
1	Pretty Big Data	- Data never sleeps	
2	Google econometrics	- Data mania versus phobia	
3	Discovery	- Big Data analytics	

Disclaimer: The opinions expressed in this presentation are not necessarily those of the European Central Bank (ECB) or the European System of Central Banks (ESCB)

Pretty Big Data – 5 Vs – Adding Value to data



Data never sleeps



Digital exploration

- Accessing
- Structuring datasets
- Linking data sets
- Slice & dice across time and datasets
- Standards (concepts, semantics)
- Transparency in methods
- Representatively
- Robustness of findings

Which preparations are needed today to have the capacity and functionality required in 3 years time?

- From experimenting to central banking tool kits?
- Linking current and past data
- Querying variety of formats
- Analytical techniques & tools
- Technical independent
- Skill sets

Source: 2018 Update: What Happens In An Internet Minute, Lori Lewis

Data mania versus phobia – a paradigm of borderless records

Digital transformation in finance and economics

E- trade

Settlement systems

Credit cards

Mobile transaction

Lending & financing

Big data

Today's talk

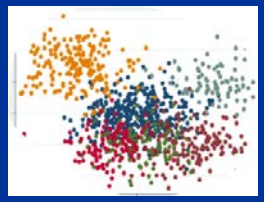
S-media

Price scans

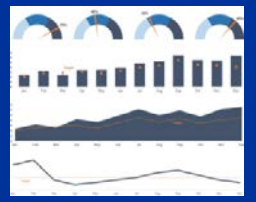
- Fintech
- Crypto assets
- DLT
- S-contracts

Data lab

Systematic acquire, Structure, Process,



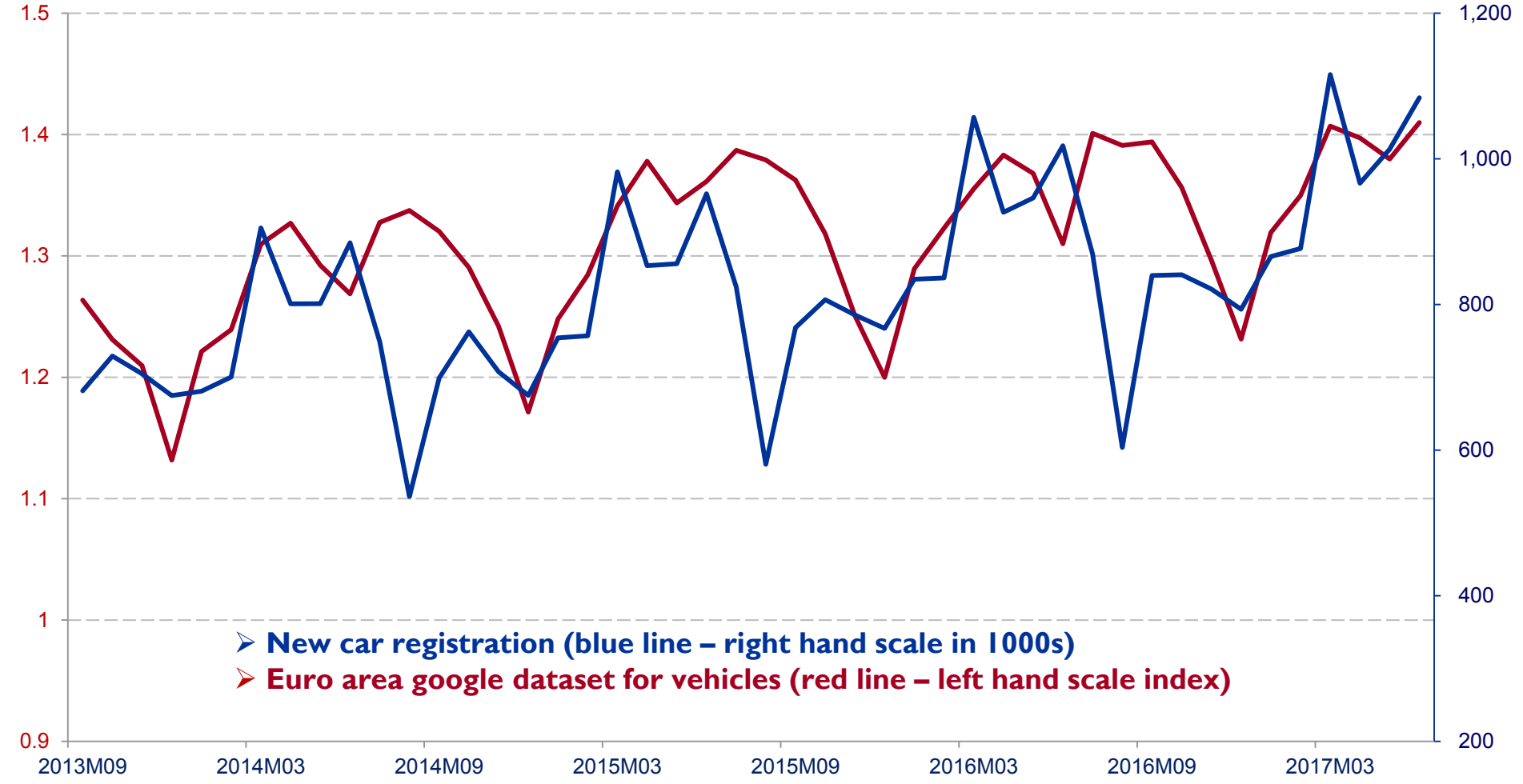
Statistical algorithm and data explorations



Packaging data for Insights & business

Google econometrics – Nowcasting car sales

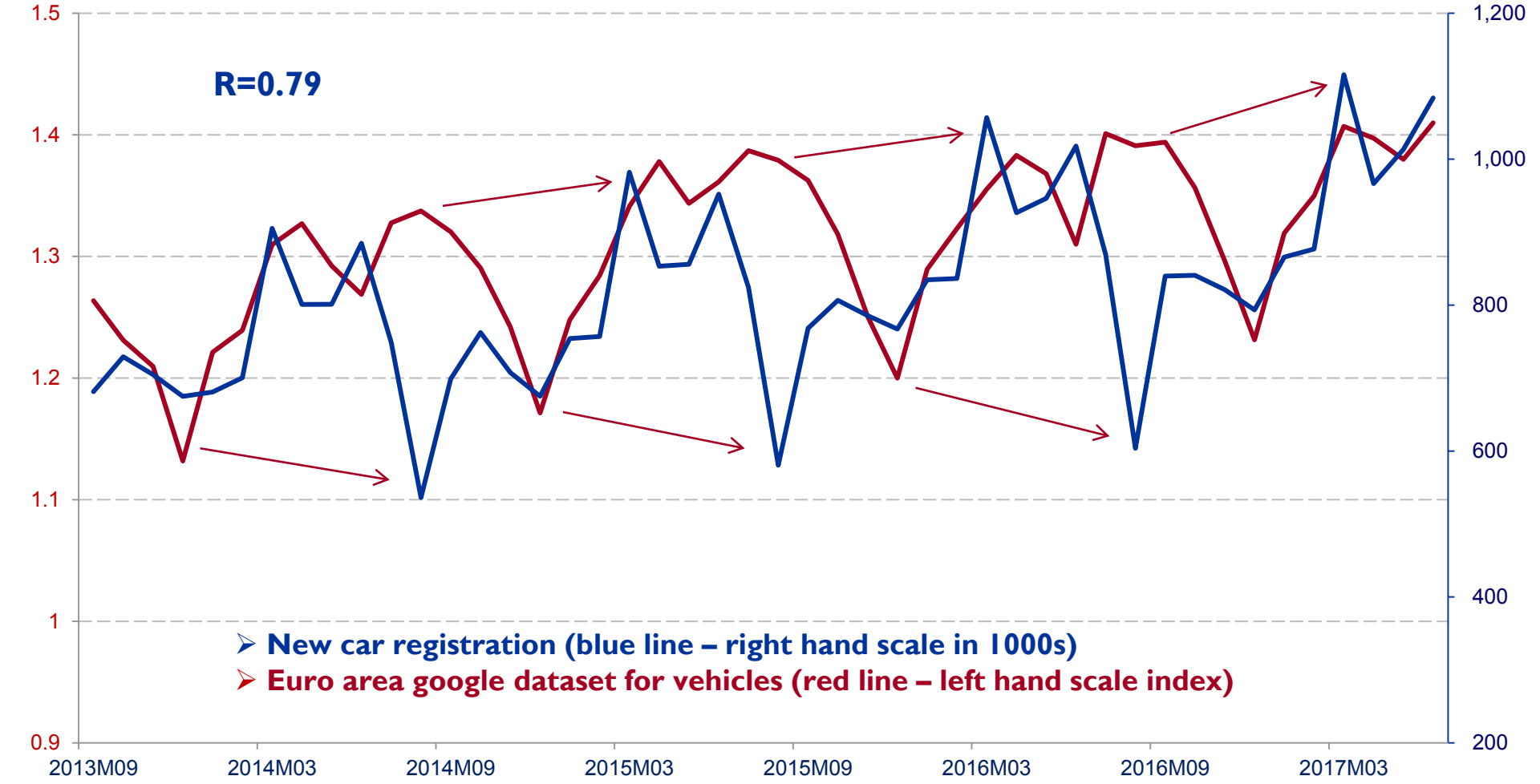
- Google data: "Autos & Vehicles",
- 10 euro area countries, weighted by national car registrations, monthly series, indexed, Represents 96% of euro area car registrations



Google econometrics: nowcasting euro area car sales and big data quality requirements, Nymand-Andersen, P., Pantelidis, E.,ECB SPS no 30

Google econometrics – Nowcasting car sales

1. There is a short term and long term dynamics between the two series
2. Two-way causal relationship and five lag lengths
3. Google data accounts for ≈ 22% of the variance in the car sales (2nd months & onwards)



Google econometrics – Nowcasting car sales

- ***7 seasonal autoregressive models*** *are used for nowcasting car sales with and without Google data and compared against a* ***seasonal autoregressive model*** (baseline)

Baseline model: $\mathbf{B_t} = \alpha + \beta_1 \log(y_{t-1}) + \beta_2 \log(y_{t-12}) + \varepsilon_t$

with Google data: $Y_t = \alpha + \beta_1 \log(y_{t-1}) + \beta_2 \log(y_{t-12}) + G_t + \varepsilon_t$, **Where**

$$\mathbf{G_t} = \beta_3 \Delta(G_t) + \beta_4 \Delta(G_{t-1}) + \beta_5 \Delta(G_{t-2}) + \beta_6 \Delta(G_{t-3}) + \beta_7 \Delta(G_{t-4}) + \beta_8 \Delta(G_{t-5})$$

Other indicators: $Y_t = B_t + [(I_i); \log(I_i)]$ V $Y_t = B_t + [(I_i); \log(I_i)] + G_t$

$I_{i=1,2,3}$ **Three indicators without and with Google data**

- **"Harmonised Index of Consumer Prices" for "Motor cars" with *one lag*.**
- **"Industrial Confidence Indicator"**
- **"Disposable Income of Households" log with *five lags*.**

The quarterly income data is converted into monthly (linear interpolation method)

Google econometrics – Nowcasting car sales

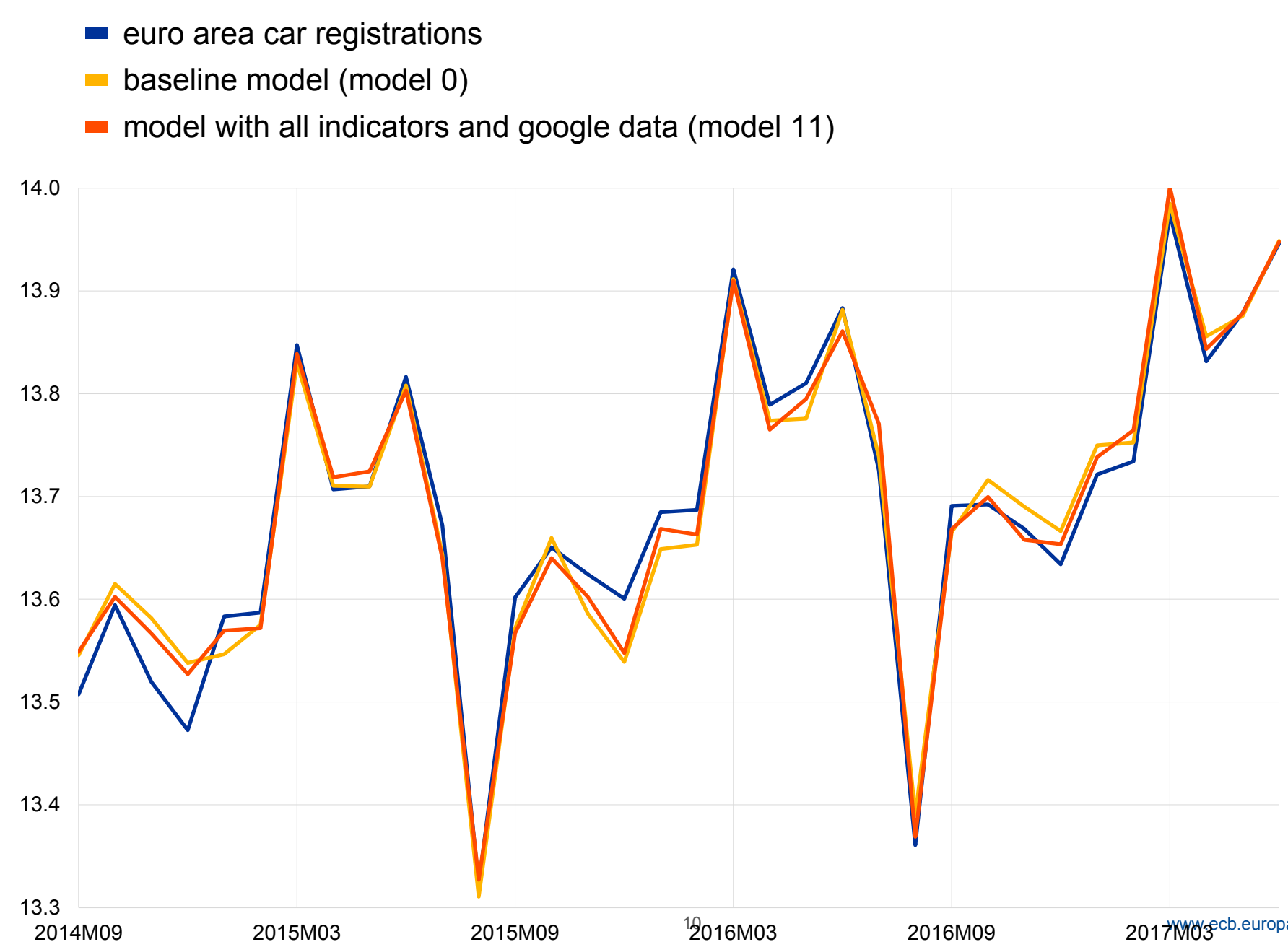
Overview of model performance and nowcasting accuracy

7 take aways

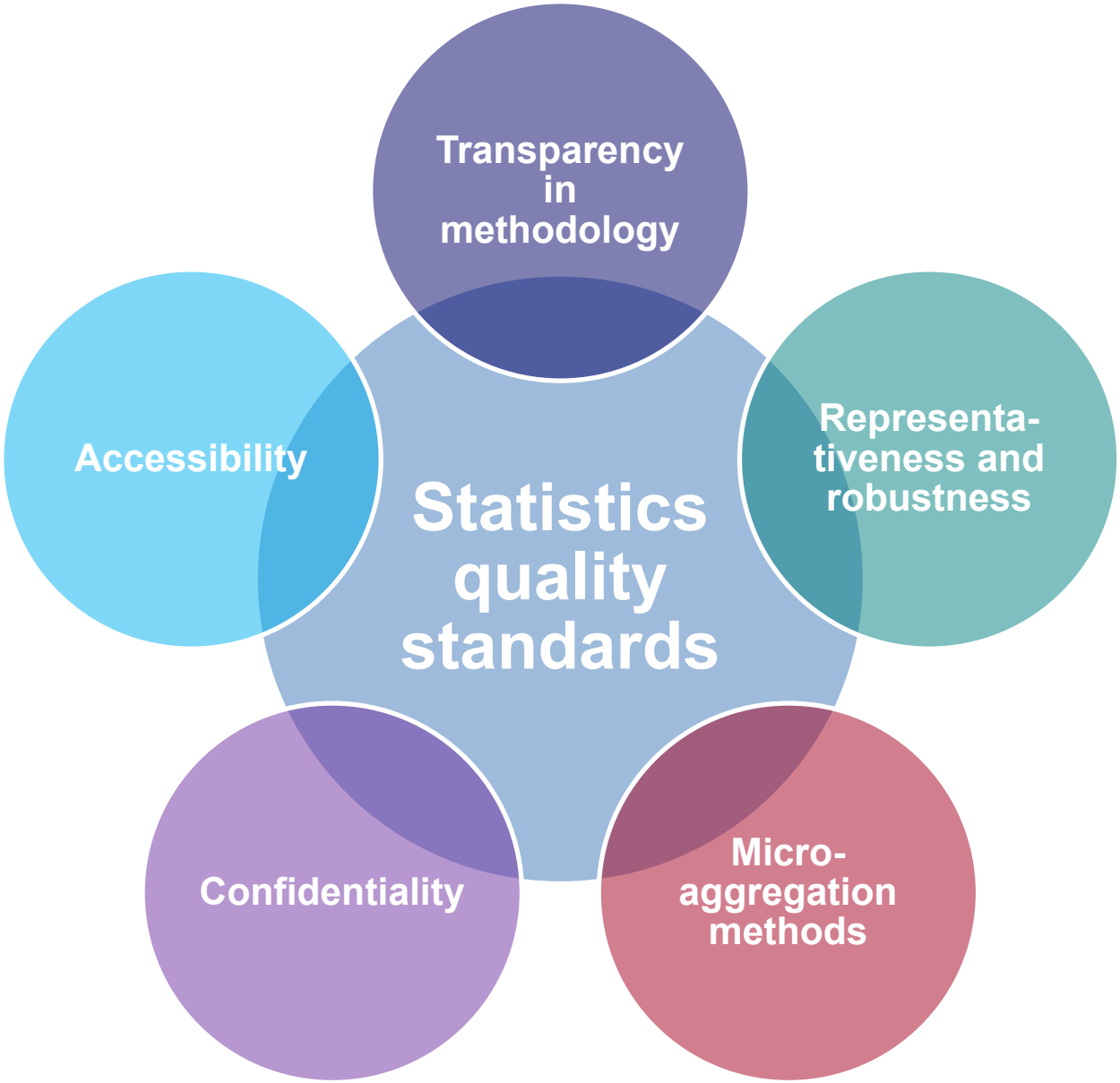
	Model/criteria	RMSE	MAE	MAPE	Improvement	DM# Base	DM# Pair (6)
0	Baseline	0.029019	0.023711	0.174652	(1)	(4)	0.0331
1	Baseline & Google (2)	0.026812	0.020303	0.149357	15.7%	0.0331	
2	With inflation rate	0.027585	0.023092	0.169944	3.4%	0.3357	0.0204
3	With inflation rate & Google	0.024885	0.019445	0.14288	21.5%	0.0383	
4	With confidence indicator	0.028942	0.023738	0.174715	4.6%	(5) 0.5289	0.0307
5	With confidence indicator & Google	0.026747	0.020334	0.149567	15.6%	0.0347	
6	With income	0.027437	0.02231	0.164112	5.8%	0.1791	0.0453
7	With income & Google	0.02296	0.018139	0.133189	(3) 30.5%	0.0195	
8	With household savings	0.027669	0.02122	0.15623	10.9%	0.0383	(7) 0.1965
9	With household savings & Google	0.025713	0.019338	0.142174	21.4%	0.0218	
	Including all explanatory variables						
10	All indicators	0.022811	0.017993	0.13253	(3) 31.2%	0.0086	0.0019
11	All indicators & Google	0.012257	0.010332	0.075806	131%	0.00004	

DM = Diebold and Mariano's test - Exploring the usefulness of internet search data on future car sales

Google econometrics – Nowcasting car sales



Big Data analytics – Quality assessment



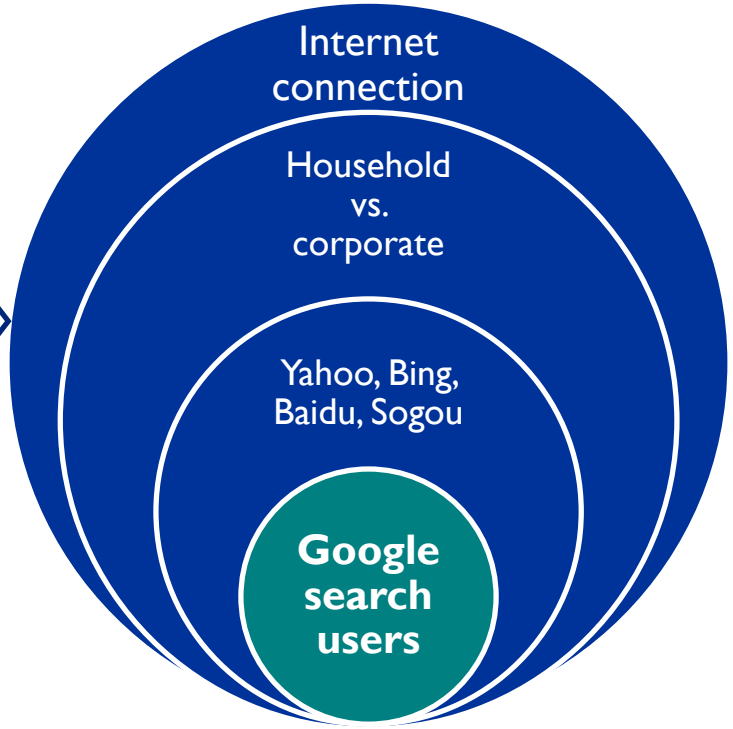
Big Data analytics – Quality assessment

One **misperception** of big data is that we **do not need** to worry about **sample bias and representativeness**, as large volumes of information supersede standard sampling theory, since big data provide census-type information

1

Digital recording of operations

Census?



2 **Unit Measurement** → **Not people - but searches**

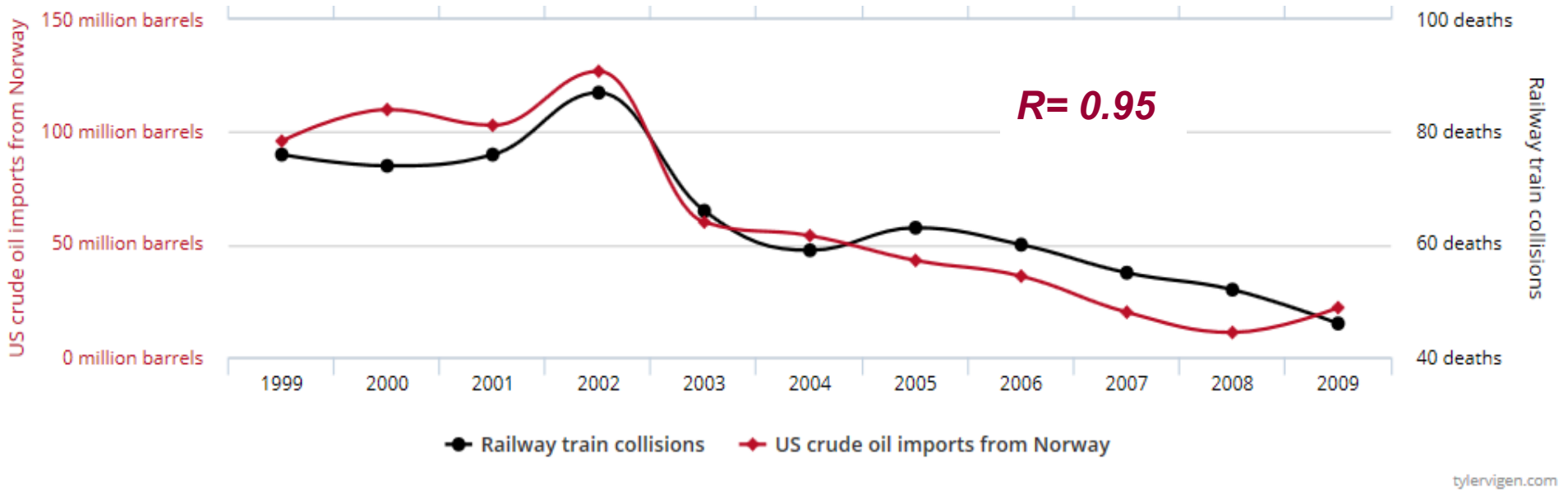
3 **Event driven** → **Volume changes may not refer to changes in demand (e.g. emission checks)**

Google logo is downloaded from *Wikipedia*.

Big Data analytics – Quality assessment

Correlation is *not (necessary)* causation

US crude oil imports from Norway correlates with drivers killed in collision with railway train



No conclusion can be drawn simply on the basis of correlations between two variables. *The similarity is a coincidence*. We *should* say that there is no causation

"The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning"
Stephen Jay Gould, American evolutionary biologist and author, 1981

Three takeaways

Progress lies in experimenting

Amara's law

Potential valuable source for nowcasting economic activities

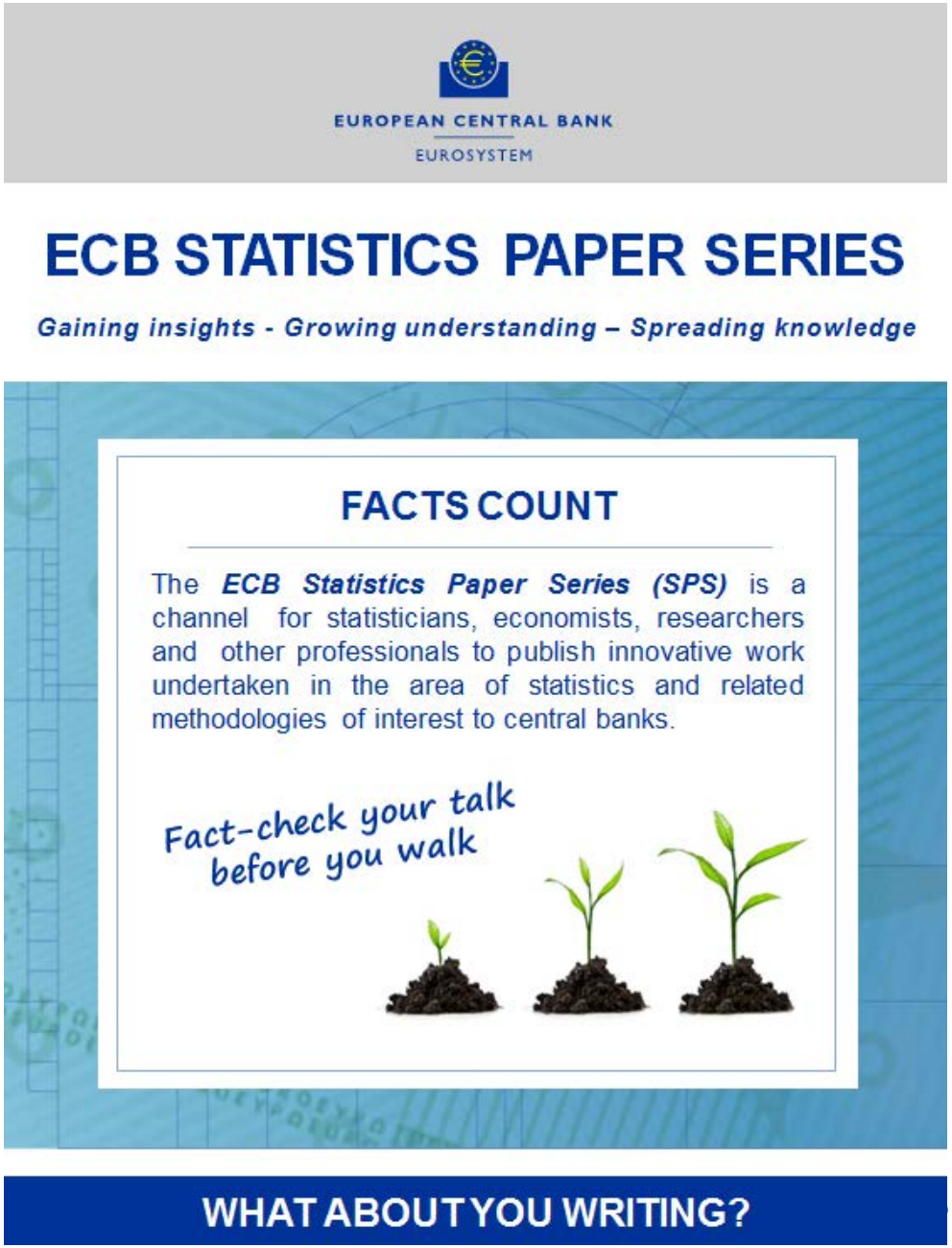
Foster transparency in data quality standards

Moving from experimenting to central banking tool kits

Collaborations - build partnerships for excellence

Google econometrics

Any question?



Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data[1]

María Gil, Javier J. Pérez and Alberto Urtasun,

Bank of Spain

A. Jesus Sánchez,

Complutense University of Madrid

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data[1]

María Gil, Javier J. Pérez, A. Jesús Sánchez and Alberto Urtasun

The focus of this paper is on nowcasting and forecasting quarterly private consumption. The selection of real-time, monthly indicators focuses on standard ("hard" / "soft" indicators) and less-standard variables. Among the latter group we analyze: i) proxy indicators of economic and policy uncertainty; ii) payment cards' transactions, as measured at "Point-ofsale" (POS) and ATM withdrawals; iii) indicators based on consumption-related search queries retrieved by means of the Google Trends application. We estimate a suite of mixed-frequency, time series models at the monthly frequency, on a real-time database with Spanish data, and conduct out-of-sample forecasting exercises to assess the relevant merits of the different groups of indicators. Some results stand out: i) "hard" and payments cards indicators are the best performers when taken individually, and more so when combined; ii) nonetheless, "soft" indicators are helpful to detect qualitative signals in the nowcasting horizon; iii) Google-based and uncertainty indicators add value when combined with traditional indicators, most notably at estimation horizons beyond the nowcasting one, what would be consistent with capturing information about future consumption decisions; iv) the combinations of models that include the best performing indicators tend to beat broader-based combinations.

[1] This article summarises the main ideas and results reported in Gil, M.; Pérez, J.J.; Sánchez, A.J. y Urtasun, A (2018): "Nowcasting private consumption: traditional indicators uncertainty measures, credit cards and some internet data". Working Paper 1842, Banco de España.

[https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/18/Files/dt1842e.pdf]

Introduction

Benchmark data to approximate private households' spending decisions are normally provided by the national accounts, and are available at the quarterly frequency. Nevertheless, usually, there exists a significant publication lag, typically of 90 days after the quarter of reference ended. More timely data is usually available in the form of economic indicators. In our paper, from a forecasting point of view, we analyze the information content of new sources of information, but in a context in which we ascertain their value in conjunction with traditional, more proven, sources of short-term information, such as the "hard" and "soft" ones mentioned above. In particular, among these new data sources we look, first, at data collected from automated teller machines (ATMs), encompassing cash withdrawals at ATM terminals, and points-of-sale (POS) payments with debit and credit cards, given the increasing and widespread use of electronic payment systems by economic agents. Secondly, in line with a recent and very active branch of the literature, we construct indicators of consumption behavior on the basis of internet search patterns as provided by Google Trends. Finally, we use measures of economic and policy uncertainty, in line with another recent strand of the literature that has highlighted the relevance of the level of uncertainty prevailing in the economy for private agents' decision-making.

To exploit the data in an efficient and effective manner, we build models that relate data at the quarterly and monthly frequencies. We follow the modeling approach of Harvey and Chung (2000).3 The mixture of frequencies, and the estimation of models at the monthly frequency, implies combining variables that at the monthly frequency can be considered as stocks with those being pure flows. The quarterly private consumption series cast into the monthly frequency is a set of missing observations for the first months of the quarter (January and February, in the case of Q1) and the observed value assigned to the last month of each quarter (say, March). Theoretically, the quarterly National Accounts series would be obtained from monthly National Accounts series by aggregation of the three months of a quarter (January to March) had them been available. We estimate such mixed-frequency models on a mixed real-time and pseudo-real-time database, for the period that starts in the early 2000s and runs through to 2017Q4, and conduct out-of-sample forecasting exercises to assess the relevant merits of different groups of indicators.

The empirical exercise

We build up a real-time database for the target variable, quarterly private consumption as measured by the National Accounts, for the period 1995Q1-2017Q4. The size of the sample for our empirical exercises, though, is restricted by the availability of some of the monthly indicators, in particular as regards Google Trends, the EPU index, and the Services Sector Activity Indicator, available for the sample starting in January 2004, January 2001, and January 2002, respectively. As regards the indicator variables we could not replicate a truly real-time dataset, so we proceeded to built up a pseudo real-time one, namely we adjusted for each point in time (month) the information set that had been available given the timing rules that we describe in the next paragraph. It is worth mentioning that the indicators that we use are not revised, which means that the pseudo-real-time approximation should be a fair representation of data available in real-time. The only discrepancy may arise from

seasonal-adjustment. While we seasonally-adjust the series that are published on nonseasonally-adjusted terms following our pseudo-real-time approach, we take official series that are published in a seasonally-adjusted form as the latest available vintage of official data.

In order to test the relevant merits of each group of indicators, as mentioned above, we consider several models that differ in the set of indicators included in each one. We estimate models that include indicators from each group at a time, several groups at a time, and different combinations of individual models.

As a mechanical benchmark we use a random walk model, whereby we repeat in future quarters the latest quarterly growth rate observed for private consumption.

We focus on the forecast performance at the nowcasting horizon (current quarter), but also explore forecasts at 1 to 4 quarters-ahead from each one of the current quarter forecast origins (first month of the quarter, second and third).

Results

First, from table1 1 (relative RMSEs) the following results can be highlighted. As regards models that use only indicators from each group (first panel of the table), the ones that use quantitative indicators and payment cards (amounts) tend to perform best than the others at the nowcasting and, somewhat less so, forecasting (1-quarter- and 4-quarters-ahead) horizons. Relative RMSE are in almost all cases below one, even though from a statistical point of view they are only different from quarterly random walk nowcasts and forecasts in a few instances. In general, the other models do not beat systematically the quarterly random walk alternative. The two main exceptions are the model with qualitative indicators for the nowcasting horizons, and the Google-Trends based ones for the longer-horizon forecasts. The latter results might be consistent with the prior that Google-based indicators deliver today information on steps to prepare purchases in the future. Lastly, it is worth mentioning that nowcast/forecast accuracy does not always improve monotonically as the information set expands, i.e. as we move from nowcast/forecast origins m1 (first month) to m3 (third month). This is explained by the real-time nature of the information set used in each case. Following the standard publication calendar, at m2-time the quarterly figure of private consumption corresponding to the previous quarter is published. This has two implications. On the one hand, the quarterly random walk alternative moves from a situation in which the reference was the $t-2$ figure to one in which the $t-1$ quarter is used. On the other hand, quarterly data corresponding to previous quarters tend to be revised at m2, which may affect the estimation of models in real-time, and eventually the accuracy of the generated nowcasts/forecasts, or at least the comparability of estimations based on different information sets. The revision of past, quarterly national accounts figures is quite apparent when going through the different panels of figures 3 and 4 in a chronological order.

Second, in the middle panel of Table 1 we show the results of the estimation of models that include quantitative indicators while adding, in turn, variables from the other groups (qualitative, payment cards - amounts, uncertainty, Google-Trends-based). The improvement in nowcast accuracy is not generalized when adding more indicators, with the exception of the "soft" ones (m1 and m3 origins). Nonetheless,

there is a significant improvement for longer forecast horizons of expanding the baseline model. In particular, for the 4-quarters-ahead one, uncertainty and Google-based indicators add significant value to the core "hard"-only based model.

Finally, as regards the third panel of results of Table 1, it seems clear that the combination (average) of models with individual groups of indicators improves the forecasting performance in all cases and at all horizons. Most notably, the combination of the forecasts of models including quantitative indicators with those with payment cards (amounts), delivers, in general, the best nowcasting/forecasting performance at all horizons. At the same time, adding the "soft" forecasts seems to add value in the nowcasting phase, when more information for the current quarter is available (m2 and m3 origins). In turn, the combination of a broad set of models (first line of the panel) produces the lowest RMSE relative to the quarterly random walk at the four quarters ahead forecast horizon. Nevertheless, the bilateral DM-test results with respect to combinations of simpler models do not tend to be, in general, significantly different from zero from a statistical point of view. In addition, according to this metric, also the models with only quantitative, qualitative and payment cards indicators individually, beat the combination of the broad set of models at the nowcasting horizons.

Conclusions

We estimate a suite of mixed-frequency models on an (almost) real-time database for the period January 2001 - December 2017, and conduct out-of-sample forecasting exercises to assess the relevant merits of different groups of indicators. The selection of indicators is guided by the standard practice ("hard" and "soft" indicators), but also expand this practice by looking at non-standard variables, namely: (i) a suite of proxy indicators of uncertainty, calculated at the monthly frequency; (ii) two additional sets of variables that are sampled at a much lower frequency: payment card transactions and indicators based on search query time series provided by Google Trends. The latter set of indicators is based on factors extracted from consumption-related search categories of the Google Trends application.

Our study shows that, even though traditional indicators make a good job at nowcasting and forecasting private consumption in real-time, novel data sources add value, most notably those based on payment cards-related, but also, to a lesser extent, Google-based and uncertainty indicators when combined with other sources.

Table 1: Relative RMSE statistics: ratio of each modelo to the quarterly random walk.[a]

Models including inidcators of only one group

	Nowcast			1-q-ahead			4-q-ahead		
	m1	m2	m3	m1	m2	m3	m1	m2	m3
Quantitative ("hard") indicators[b]	0.84	0.75*	0.79	0.75**	0.81	0.80	0.98	0.97	1.00
Qualitative ("soft") indicators[c]	1.01	0.85	0.85	1.11	1.05	1.05	1.09	1.10	1.29*
Payment cards (amounts, am)[d]	0.79	0.82	0.88	0.65***	0.84	0.69**	0.74**	0.84	0.83
Payment cards (numbers)[d]	1.05	1.15	1.13	0.90	1.10	0.98	0.75**	0.81	0.79
Uncertainty indicators[e]	1.06	0.97	0.99	1.00	1.05	1.06	0.94	1.00	1.02
Google: aggregate of all indicators	1.04	1.06	1.06	0.85	1.03	1.03	0.71**	0.79	0.79
Google: durable goods (lagged)	1.04	0.97	0.98	0.96	1.04	1.04	0.85*	0.93	0.93

Models including indicators from different groups

	Nowcast			1-q-ahead			4-q-ahead		
	m1	m2	m3	m1	m2	m3	m1	m2	m3
Quantitative & Qualitative	0.69**	0.78	0.77	0.67***	0.76*	0.72*	0.79*	0.82*	0.80*
Quantitative & Payment cards (am)d	0.90	0.82	0.91	0.67***	0.79	0.78	0.86	0.89	0.91
Quantitative & Uncertainty	0.88	0.86	0.75	0.74**	0.91	0.93	0.69**	0.76	0.76
Quantitative & Google (aggregate)	0.85	0.76	0.77	0.81*	0.94	0.89	0.77**	0.81*	0.82
Quantitative & Google (durables)	0.91	0.95	0.87	0.69**	0.83	0.88	0.72**	0.76*	0.77*

Combination of models

	Nowcast			1-q-ahead			4-q-ahead		
	m1	m2	m3	m1	m2	m3	m1	m2	m3
All models[f]	0.66**	0.71**	0.69**	0.68**	0.77*	0.68**	0.73**	0.78*	0.78*
Hard & Payment cards (am)d	0.62**	0.69**	0.71**	0.53**	0.69**	0.52***	0.79*	0.86	0.84
Hard, Payment cards (am)d & Soft	0.65**	0.67**	0.67**	0.68**	0.74**	0.59***	0.83*	0.89	0.92
Hard & Soft	0.68**	0.66**	0.66**	0.77**	0.75**	0.69**	0.91	0.94	1.02
Hard & Google (durables)	0.77**	0.78**	0.76**	0.74**	0.83	0.78*	0.85	0.91	0.9

Notes:

The asterisks denote the Diebold Mariano test results for the null hypothesis of equal forecast accuracy of two forecast methods. A squared loss function is used. The number in each cell represents the loss differential of the method in its horizontal line as compared to the quarterly random walk alternative. A single (double) [triple] asterisk denotes rejection of the null hypothesis at the 10% (5%) [1%] level of significance. of the null hypothesis at the 10% (5%) [1%] level of significance.

[a] Nowcast/forecast errors computed as the difference to the first released vintage of private consumption data. Forecasts generated recursively over the moving window 2008Q1 (m1) to 2017Q4 (m3)

[b] Social Security Registrations; Retail Trade Index; Activity Services Index.

[c] PMI Services; Consumer Confidence Index.

[d] Aggregate of payment cards via POS and ATMs.

[e] Stock Market Volatility (IBEX); Economic Policy Uncertainty Index (EPU).

[f] Combination of the results of 30 models, that include models in which the indicators of each block are included separately, models that include the quantitative block and each other block, and version of all the previous models but including lags of the variables.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data[1]

María Gil, Javier J. Pérez and Alberto Urtasun,
Bank of Spain

A. Jesus Sánchez,
Complutense University of Madrid

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data

María Gil, Javier J. Pérez (*), and Alberto Urtasun
Bank of Spain, Eurosystem

A. Jesus Sánchez
Univ. Complutense of Madrid, Spain

Bank Indonesia / IFC "International Workshop on Big Data for Central Bank Policies"
Bali, Indonesia, 25 July 2018

Directorate General Economics and Statistics

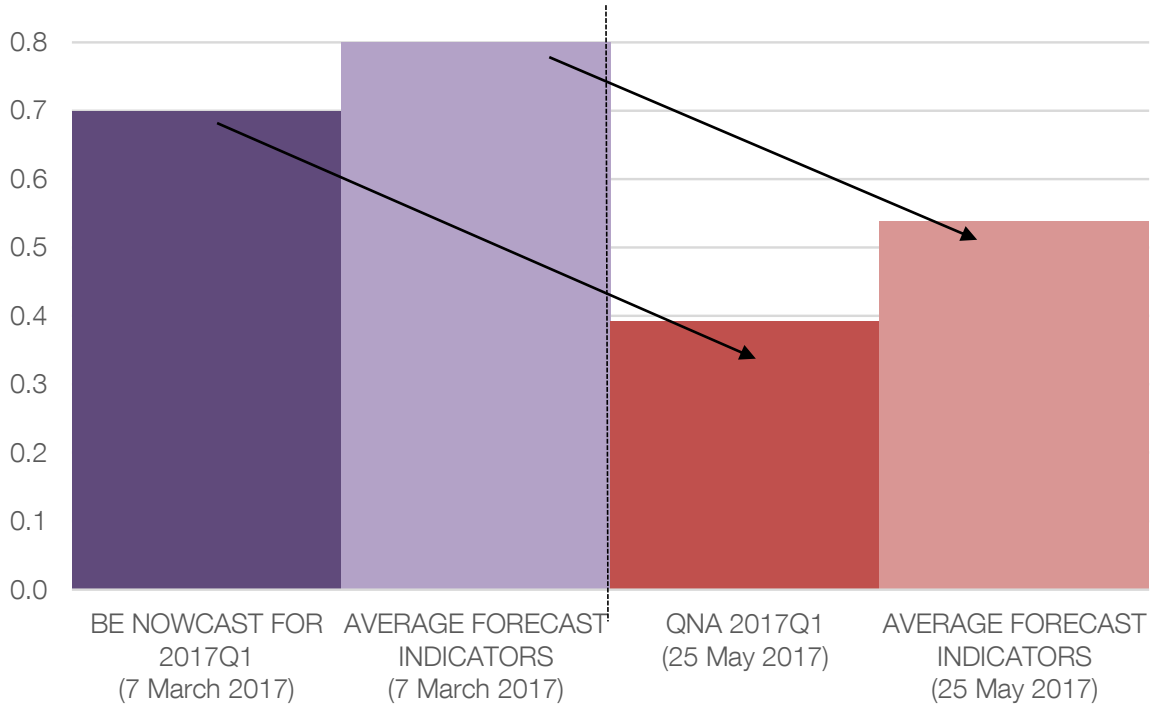
Outline

1. **Motivation**
2. Literature review
3. The data
4. Modeling approach
5. The empirical exercise
6. Selection of results and conclusions

1. Motivation

- Relevance of private consumption: some 60% of GDP
- Key variable from both a forecasting and a policy perspective
- Real-time assessment limited by availability of statistical information (QNA and monthly, traditional indicators, subject to standard publication lags)

Example: Nowcasting of q-q growth rates of QNA private consumption in the first quarter of 2017 (Spain)

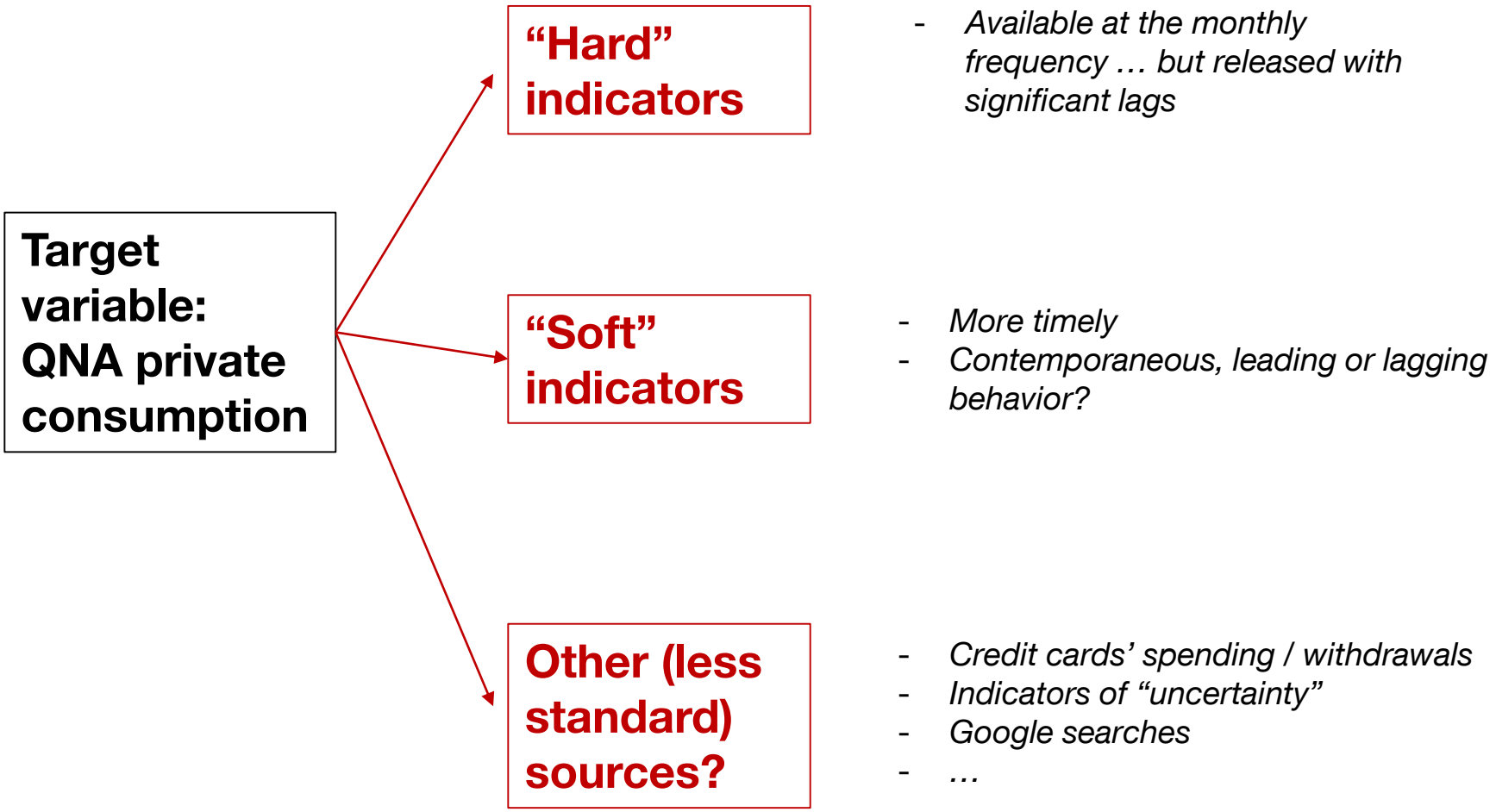


Principal component of available indicators

Principal component of available indicators

1. Motivation: traditional vs. "new" indicators

- Technological developments allow the use of new data sources



AIM OF THE PAPER:

Explore the relative merits of…

… hard vs. soft

… traditional vs. "new" indicators

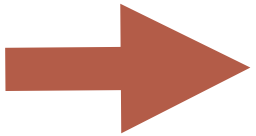
to "nowcast" Spanish real private consumption

Outline

1. Motivation
2. **Literature review**
3. The data
4. Modeling approach
5. The empirical exercise
6. Selection of results and conclusions

2. Literature review

- Traditionally, the literature on “nowcasting” has been quite focused on GDP
- Few exceptions of papers in which GDP is modelled together with its demand and/or supply components
- More recently, the literature has started to explore “new” sources
  - ✓ Not so much for the case of private consumption
  - ✓ **GOOGLE SEARCHES**
  - ✓ **ATM/Point Of Sale DATA**
  - ✓ **UNCERTAINTY MEASURES**



Outline

1. Motivation
2. Literature review
3. **The data**
4. Modeling approach
5. The empirical exercise
6. Selection of results and conclusions

3. The data

- Quarterly private consumption,

  monthly indicators

  (lower frequencies not exploited yet)

Quarterly private consumption

Social sec. registrations
Retail trade index
Services activity index
PMI – services
Consumers confidence index

"Traditional" indicators

Credit/debit cards' payments
Amounts
Number

ATM/POS

Disagreement consumption
Disagreement about inflation
Uncertainty on unemployment
Economic Policy Uncertainty
IBEX-35 volatility index

Uncertainty

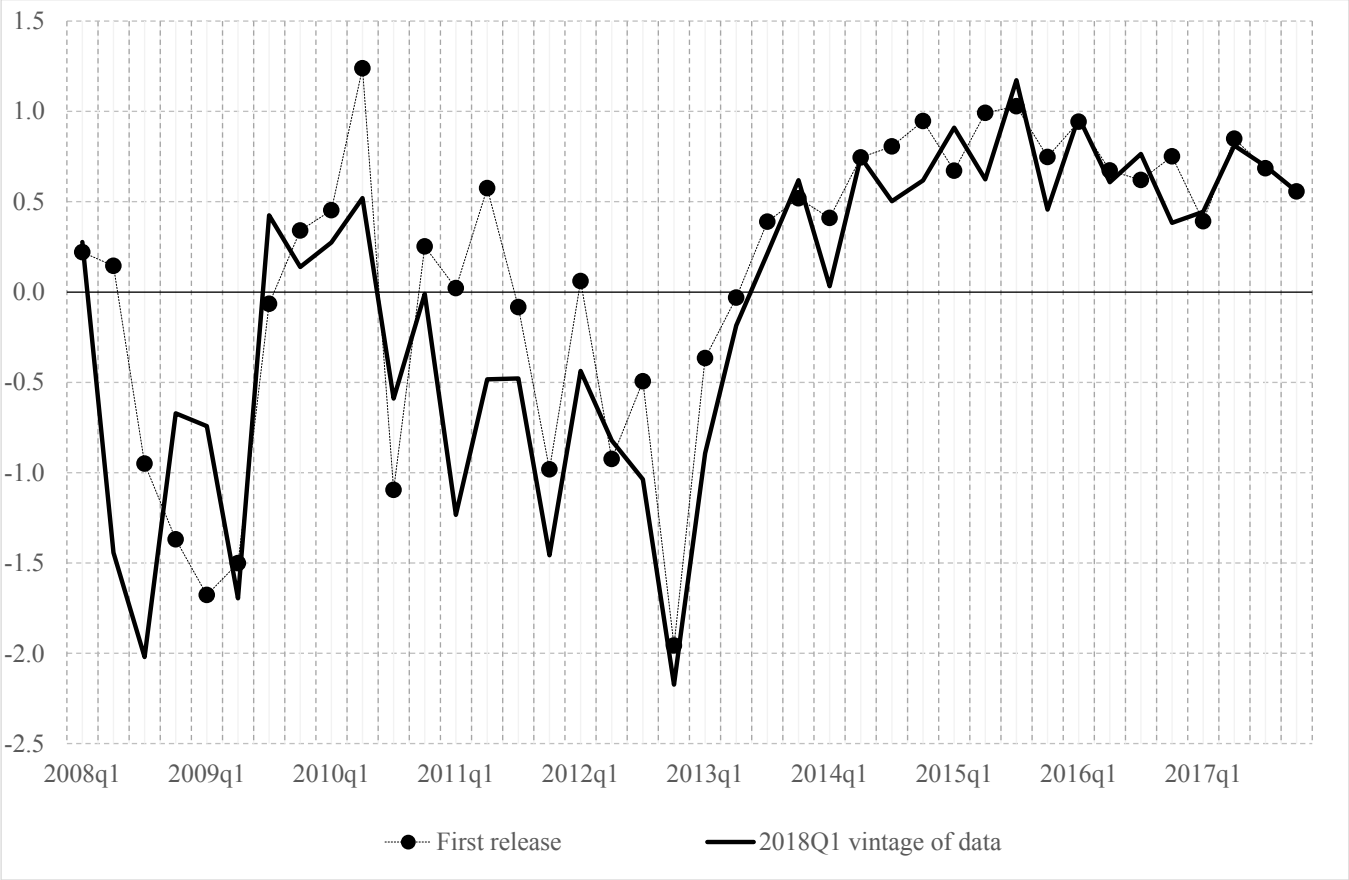
Google Trends
- Durables
- Non durables

Google search queries

3. The data

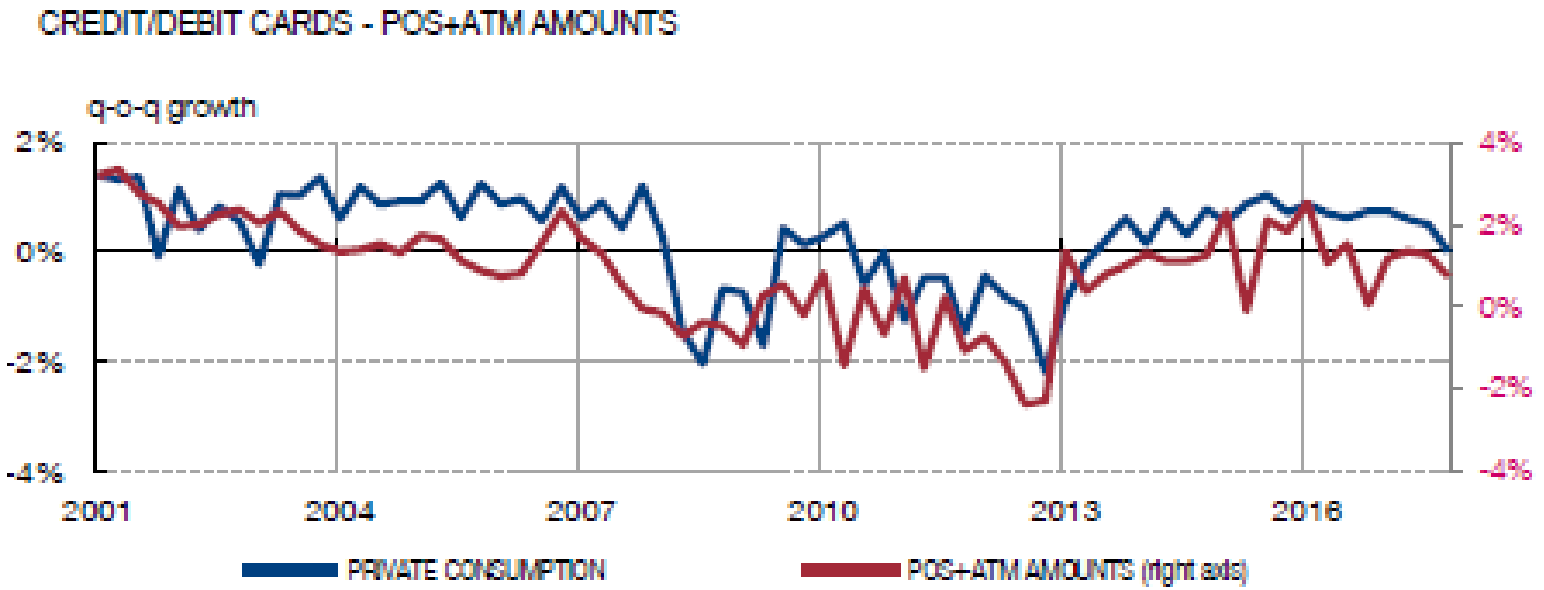
- Quarterly private consumption,

  Data revisions



3. The data: ATM/POS

- Widespread use of electronic payment systems
- Timeliness [daily/weekly frequency, in theory]
- Credit cards:
  - POS: payments by means of credit/debit cards in points of sales - amounts [seasonally adjusted, deflated by national CPI] and number of operations
  - ATMs: cash withdrawals and number of withdrawls



3. The data: Google Trends

- Households use internet to buy goods and services
  - ✓ Willingness to buy
  - ✓ Info easily available.
- Evidence of usefulness in the literature: robustness?
- "Google Trends" provides an index of the relative volume of search queries conducted through Google (daily/weekly)
  - ✓ It provides aggregated indexes of search queries which are classified into categories and sub-categories using an automated classification engine
  - ✓ We select consumption-relevant categories (~60) that match the product categories of personal consumption expenditures of the BEA's national income and product accounts

3. The data: Google Trends

- Example

Classification by national product and incc	Google categories
Durable goods	
Motor vehicles and parts	Automotive, auto financing, automotive parts, auto insurance, Seat, Mercedes Benz, Mercedes offer, second hand car, car, to buy a car
Furnishing and durable household equipment	Electrical appliance, home insurance, home remodel, home furnishing, interior decoration, interior design
Recreational goods and vehicles	Online movie, to buy a movie, watch online movie, video games
Other durable goods	Telecommunications, router wifi, mobile phone, electronic book, novel

---

- Data available since January 2004 [not seasonally adjusted → TRAMO-SEATS]
- Distinguish durable/ non durable/ services
  - ✓ "Aggregation": (i) Principal Components Analysis (literature); (ii) **NA weights**

3. The data: Uncertainty

- **Economic Policy Uncertainty Index (EPU)** (Baker, Bloom, Davis, 2016): it measures the frequency of news related to economic policy uncertainty in two of the most popular Spanish newspapers.
- **European Commission Business and Consumer Surveys**: "unemployment expectations for the next 12 months", indicator computed as

$$\sqrt{Frac_t^{+} + Frac_t^{-} - (Frac_t^{+} - Frac_t^{-})^2}$$

where $Fract^{+/-}$ is the weighted fraction of consumers in the cross section with increase/decrease responses at time t.

- **Indicators of disagreement about consumption and inflation forecasts**, calculated using the information provided by a private institute (FUNCAS) that published every two months a panel of forecasters. At each point in time, this measure is computed as the standard deviation of such cross-section of forecasters

$$\frac{1}{n}\sum_{1}^{n}(\hat{C}_i - \hat{C}_A)^2$$

3. The data

- Preliminary exploration: regressions of Δlog C on determinants and indicators
  - ✓ Indicators by blocks add value

$$\Delta \log(C_t) = \alpha_1 + \alpha_2 \Delta \log(C_{t-1}) + \alpha_3 X_t + \epsilon_t$$

p-values Sample: 2001Q1-2016Q4	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Constant	0.060 *	0.052 **	0.000 ***	0.047 **	0.288	0.280	0.001 ***
Interest rate: Euribor 3-months[a]	0.635	0.674	0.829	0.513	0.370	0.523	0.782
Households' disposable income[b]	0.952	0.564	0.267	0.338	0.265	0.761	0.487
Lagged $\Delta \log(C_t)$	0.772	0.100 *	0.002 ***	0.000 ***	0.000 ***	0.660	0.2969
Short-term Indicators:							
"Hard": Social Security Registrations[b]	0.000 ***					0.010 **	
"Hard": Retail Trade Index[b]	0.020 **						
"Soft": PMI-Services[c]		0.007 ***				0.258	0.000 ***
"Soft": Consumers' Confidence Index[c]		0.162					
Credit cards: POS amounts (real)[b]			0.003 ***			0.007 ***	0.000 ***
Credit cards: POS number of transactions[b]			0.252				
Uncertainty: Stock market volatility[c]				0.0951 *		0.872	0.970
Uncertainty: Economic Policy[d]				0.0000 ***			
Google Trends: Durable Goods[b]					0.037 **	0.192	0.086 *
Google Trends: Non-durable Goods[b]					0.207		
R-squared statistic	0.72	0.62	0.60	0.51	0.49	0.74	0.71

Notes:
a. Deviation from trend (HP-filter).
b. $\Delta \log(\bullet)$.
c. Variable in levels.
d. $\Delta(\bullet)$.

Outline

1. Motivation
2. Literature review
3. The data
4. **Modeling approach**
5. The empirical exercise
6. Selection of results and conclusions

4. Modelling approach

- Different sampling frequency: monthly (indicators), quarterly (consumption)

- Publication delays cause missing values for some of the variables at the end of the sample ("ragged-end" problem)

4. Modelling approach

- Mixed-frequencies models, in the vein of Harvey and Chun (2000)

  - ✓ Multivariate setup: Unobserved Components Model
    - *Flexibility: aggregation and modelling using the State Space representation*
    - *Models in levels: no need to worry about ex ante stationarity or cointegration*
  - ✓ Seemingly Unrelated Structural Time Series Models (SUTSE)
  - ✓ Different sampling intervals: cumulator variable
  - ✓ Optimal interpolation using the Kalman Filter and the Smoothing Algorithm

4. Modelling approach

The basic model is of the Unobserved Component Model class known as the Basic Structural Model (Harvey 1989), that decomposes a set of time series in unobserved though meaningful components from an economic point of view (mainly trend, seasonal and irregular). The model is multivariate, and may be written as

$$\begin{bmatrix} \mathbf{z}_t \\ \mathbf{u}_t \end{bmatrix} = \mathbf{T}_t + \mathbf{e}_t.$$

State-Space representation

Trend: **SRW** (no seasonal)

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{E}\mathbf{w}_t$$

$$\begin{bmatrix} \mathbf{z}_t \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}^u \end{bmatrix} \mathbf{x}_t + \begin{bmatrix} \epsilon_t \\ \mathbf{v}_t \end{bmatrix}$$

where

$$\mathbf{w}_t \sim N(0, \Sigma_{\mathbf{w}_t}), \qquad \epsilon_t \sim N(0, \Sigma_\epsilon),$$
$$\mathbf{v}_t \sim N(0, \Sigma_{\mathbf{v}_t}).$$

Aggregation

Cumulator variable

$$C_t = \begin{cases} 0, & t = \text{every January (monthly data)/ first quarter (quarterly data)} \\ 1, & \text{otherwise.} \end{cases}$$

Thus, the model turns out to be:

$$\begin{bmatrix} \mathbf{z}_t \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} C_t \otimes \mathbf{I} & \mathbf{H}\Phi \\ \mathbf{0} & \Phi \end{bmatrix} \begin{bmatrix} \mathbf{z}_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & \mathbf{H} \\ \mathbf{0} & \mathbf{E} \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \mathbf{w}_t \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{z}_t \\ \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{z}_t \\ \end{bmatrix} + \begin{bmatrix} \mathbf{0} \end{bmatrix} \mathbf{v}_t$$

$\mathbf{z}_t$: quarterly private consumption

$\mathbf{u}_t$: set of monthly indicators

Outline

1. Motivation
2. Literature review
3. The data
4. Modeling approach
5. **The empirical exercise**
6. Selection of results and conclusions

5. The empirical exercise

- Real-time database
  - ✓ Different forecast origins (information sets) within each quarter: m1, m2, m3
- Full sample: (1995Q1) 2001Q1-2017Q4
- Out-of-sample evaluation over 2008Q1-2017Q4 [40 obs. per forecast origin x 3]
  - ✓ Quantitative: RMSEs
  - ✓ Quantitative: Diebold-Mariano
  - ✓ Qualitative: sign anticipation (Pesaran-Timmermann)
- First released QNA figure as reference [results qualitatively similar for "FINAL"]
- Benchmark (QRW): repeat quarterly growth rate (economic grounds)

4. Empirical exercise: real-time database

Information available at nowcasting time…	… m1 (1st month of the quarter)						… m2 (2nd month of the quarter)						… m3 (3rd month of the quarter)					
	Previous quarter			Current quarter			Previous quarter			Current quarter			Previous quarter			Current quarter		
	1st month	2nd month	3rd month	1st month	2nd month	3rd month	1st month	2nd month	3rd month	1st month	2nd month	3rd month	1st month	2nd month	3rd month	1st month	2nd month	3rd month
Private consumption (QNA)																		
Social security registrations																		
Retail trade index																		
Services activity index																		
PMI. Services																		
Consumers confidence index																		
Credit cards - ATMs																		
Credit cards - POSs																		
Disagreement - consumption																		
Disagreement - inflation																		
Unemploymeny expectations																		
Economic policy uncertainty																		
Stock market volatility																		
Google Trends																		

Outline

1. Motivation
2. Literature review
3. The data
4. Modeling approach
5. The empirical exercise
6. **Selection of results and conclusions**

6. Selection of results and conclusions

- **Quantitative measures of forecast accuracy: RMSEs relative to QRW**

Models including indicators of only one group

	Nowcast			1-q-ahead			4-q-ahead		
	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$
Quantitative ("hard") indicators [b]	0.84	0.75 *	0.79	0.75 **	0.81	0.80	0.98	0.97	1.00
Qualitative ("soft") indicators [c]	1.01	0.85	0.85	1.11	1.05	1.05	1.09	1.10	1.29 *
Payment cards (amounts, am)[d]	0.79	0.82	0.88	0.65 ***	0.84	0.69 **	0.74 **	0.84	0.83
Payment cards (numbers)[d]	1.05	1.15	1.13	0.90	1.10	0.98	0.75 **	0.81	0.79
Uncertainty indicators [e]	1.06	0.97	0.99	1.00	1.05	1.06	0.94	1.00	1.02
Google: aggregate of all indicators	1.04	1.06	1.06	0.85	1.03	1.03	0.71 **	0.79	0.79
Google: durable goods (lagged)	1.04	0.97	0.98	0.96	1.04	1.04	0.85 *	0.93	0.93

Models including indicators from different groups

	Nowcast			1-q-ahead			4-q-ahead		
	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$
Quantitative & Qualitative	0.69 **	0.78	0.77	0.67 ***	0.76 *	0.72 *	0.79 *	0.82 *	0.80 *
Quantitative & Payment cards (am)[d]	0.90	0.82	0.91	0.67 ***	0.79	0.78	0.86	0.89	0.91
Quantitative & Uncertainty	0.88	0.86	0.75	0.74 **	0.91	0.93	0.69 **	0.76	0.76
Quantitative & Google (aggregate)	0.85	0.76	0.77	0.81 *	0.94	0.89	0.77 **	0.81 *	0.82
Quantitative & Google (durables)	0.91	0.95	0.87	0.69 **	0.83	0.88	0.72 **	0.76 *	0.77 *

Combination of models

	Nowcast			1-q-ahead			4-q-ahead		
	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$
All models [f]	0.66 **	0.71 **	0.69 **	0.68 ***	0.77 *	0.68 **	0.73 **	0.78 *	0.78 *
Hard & Payment cards (am)[d]	0.62 **	0.69 **	0.71 **	0.53 ***	0.69 **	0.52 ***	0.79 *	0.86	0.84
Hard, Payment cards (am)[d] & Soft	0.65 **	0.67 **	0.67 **	0.68 ***	0.74 **	0.59 ***	0.83 *	0.89	0.92
Hard & Soft	0.68 **	0.66 **	0.66 **	0.77 **	0.75 **	0.69 **	0.91	0.94	1.02
Hard & Google (durables)	0.77 **	0.78 **	0.76 **	0.74 ***	0.83	0.78 *	0.85	0.91	0.90

6. Selection of results and conclusions

- **Quantitative measures of forecast accuracy: pairwise relative RMSEs & DM-tests**

Nowcast origin - $m1$

	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10][g]
Q-Random Walk [1]	1.19	0.99	1.26	0.95	0.94	0.96	1.52**	1.62**	1.55**
Quantitative[a] [2]	—	0.83	1.06	0.80*	0.79*	0.81	1.27*	1.35**	1.30*
Qualitative[b] [3]	—	—	1.27*	0.96	0.95	0.97	1.53**	1.63**	1.56***
Payment cards[c] [4]	—	—	—	0.75**	0.74*	0.76	1.20	1.28*	1.23*
Uncertainty[d] [5]	—	—	—	—	0.99	1.01	1.60***	1.70***	1.63***
Google[e] [6]	—	—	—	—	—	1.03	1.62***	1.72***	1.65***
Comb: All models[f] [7]	—	—	—	—	—	—	1.58***	1.68**	1.61***
Comb: Quant.[a] & Cards[c] [8]	—	—	—	—	—	—	—	1.06	1.02
Comb: Quant.[a] & Qual.[b] [9]	—	—	—	—	—	—	—	—	0.96

Nowcast origin - $m2$

	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10][g]
Q-Random Walk [1]	1.33*	1.18	1.23	0.87	1.03	1.03	1.41**	1.45**	1.48***
Quantitative[a] [2]	—	0.89	0.92	0.65**	0.77*	0.77	1.06	1.09	1.12
Qualitative[b] [3]	—	—	1.04	0.73**	0.87	0.87	1.19	1.22	1.26**
Payment cards[c] [4]	—	—	—	0.71**	0.84	0.84	1.15	1.18	1.21*
Uncertainty[d] [5]	—	—	—	—	1.18	1.18	1.62***	1.66***	1.71***
Google[e] [6]	—	—	—	—	—	1.00	1.37**	1.41**	1.44**
Comb: All models[f] [7]	—	—	—	—	—	—	1.37**	1.41**	1.45**
Comb: Quant.[a] & Cards[c] [8]	—	—	—	—	—	—	—	1.03	1.05
Comb: Quant.[a] & Qual.[b] [9]	—	—	—	—	—	—	—	—	1.03

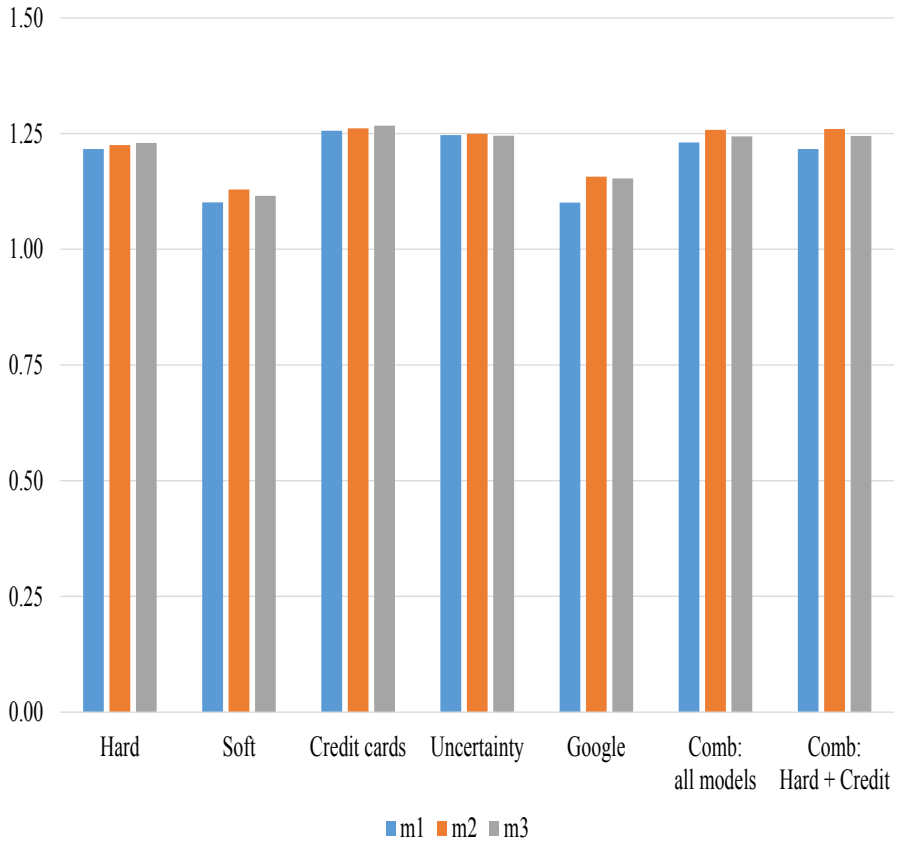
Nowcast origin- $m3$

	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10][g]
Q-Random Walk [1]	1.27	1.18	1.14	0.89	1.01	1.02	1.46**	1.42**	1.49**
Quantitative[a] [2]	—	0.93	0.89	0.70**	0.79	0.81	1.15	1.11	1.17
Qualitative[b] [3]	—	—	0.97	0.75**	0.85	0.87	1.24	1.20	1.26**
Payment cards[c] [4]	—	—	—	0.78*	0.88	0.90	1.28*	1.24**	1.30**
Uncertainty[d] [5]	—	—	—	—	1.13	1.15	1.64***	1.60***	1.67***
Google[e] [6]	—	—	—	—	—	1.02	1.45**	1.41*	1.48**
Comb: All models[f] [7]	—	—	—	—	—	—	1.42**	1.38*	1.45**
Comb: Quant.[a] & Cards[c] [8]	—	—	—	—	—	—	—	0.97	1.02
Comb: Quant.[a] & Qual.[b] [9]	—	—	—	—	—	—	—	—	1.05

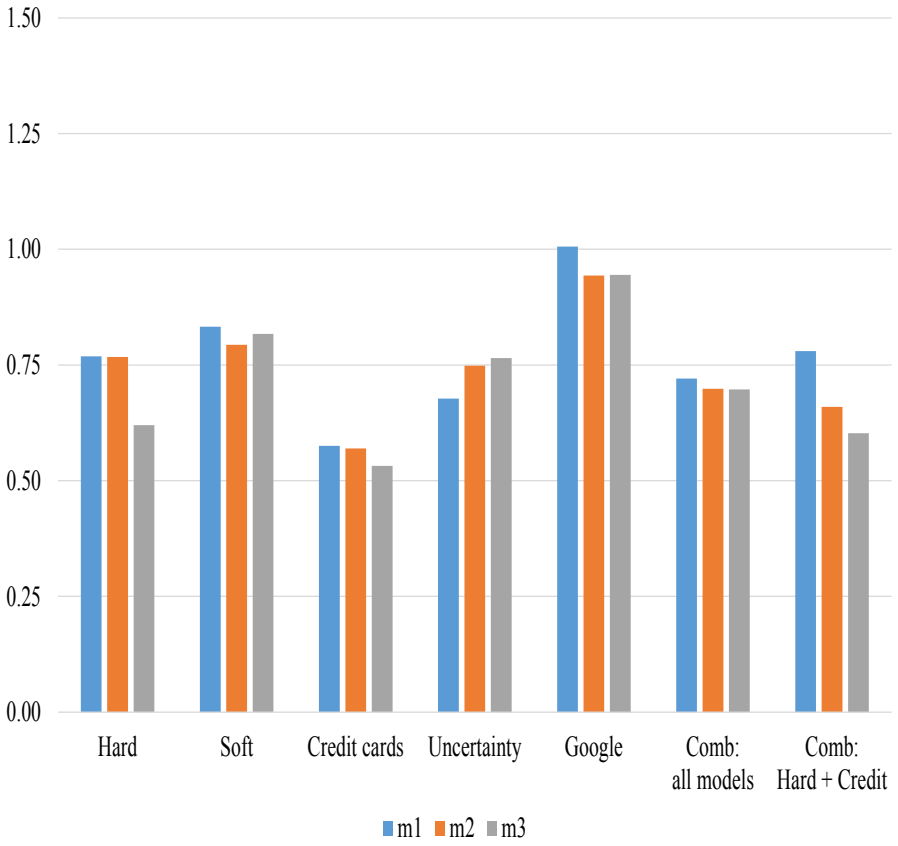
6. Selection of results and conclusions

- Better forecast performance in **“good times”** than in “bad times”
- Performance in **“good and bad times”**: change in relative behavior (cards, others)

Recession (2008-2012) vs Full sample
Relative RMSE



Expansion (2013-2016) vs Full sample
Relative RMSE



6. Selection of results and conclusions

- **Qualitative measures of forecast accuracy: Pessaran-Timmerman tests**

Models including indicators of only one group

	Nowcast			1-q-ahead			4-q-ahead		
	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$
Quarterly Random Walk	0.70**	0.80***	0.80***	0.64	0.69**	0.69**	0.50	0.56	0.56
Quantitative ("hard") indicators [b]	0.78***	0.78***	0.73***	0.77***	0.82***	0.77***	0.64***	0.69***	0.67***
Qualitative ("soft") indicators [c]	0.78***	0.85***	0.83***	0.69**	0.79***	0.77***	0.50	0.61	0.53
Payment cards (amounts,am)[d]	0.78***	0.78***	0.80***	0.74***	0.79***	0.85***	0.56	0.58	0.61
Payment cards (numbers)[d]	0.75***	0.75***	0.75***	0.72**	0.69**	0.77***	0.50	0.69**	0.67**
Uncertainty indicators [e]	0.73***	0.78***	0.78***	0.67**	0.72***	0.69**	0.50	0.56	0.56
Google: aggregate of all indicators	0.40	0.80***	0.80***	0.26	0.51	0.41	0.25	0.28	0.44
Google: durable goods (lagged)	0.73***	0.73***	0.73***	0.69**	0.72***	0.72***	0.53	0.58	0.58

Models including indicators from different groups

	Nowcast			1-q-ahead			4-q-ahead		
	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$
Quantitative & Qualitative	0.78***	0.78***	0.78***	0.77***	0.87***	0.85***	0.64***	0.67***	0.67***
Quantitative & Payment cards (am)[d]	0.78***	0.83***	0.78***	0.74***	0.77***	0.77***	0.61**	0.67***	0.61**
Quantitative & Uncertainty[e]	0.75***	0.73***	0.75***	0.79***	0.74***	0.74***	0.56	0.61	0.61
Quantitative & Google (aggregate)	0.78***	0.78***	0.70***	0.74***	0.77***	0.82***	0.58	0.58	0.64**
Quantitative & Google (durables)	0.78***	0.78***	0.78***	0.77***	0.82***	0.79***	0.64**	0.61	0.61*

Combination of models

	Nowcast			1-q-ahead			4-q-ahead		
	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$	$m1$	$m2$	$m3$
All models [f]	0.88***	0.83***	0.78***	0.74***	0.85***	0.85***	0.56	0.58	0.67***
Hard & Payment cards (am)[d]	0.83***	0.78***	0.78***	0.85***	0.87***	0.85***	0.61*	0.69***	0.69***
Hard, Payment cards (am)[d] & Soft	0.83***	0.85***	0.78***	0.74***	0.82***	0.87***	0.53	0.61*	0.67**
Hard & Soft	0.83***	0.80***	0.80***	0.74***	0.79***	0.82***	0.50	0.64**	0.58
Hard & Google (durables)	0.78***	0.78***	0.75***	0.77***	0.77***	0.79***	0.56	0.58*	0.61**

- **Summing up:**

  - ✓ **"Traditional hard" indicators tend to dominate the nowcasting race**
  - ✓ **But credit-debit cards perform similarly in many dimensions, and work quite well combined with "hard"**
  - ✓ **The other "new" sources (Google, Uncertainty) and "Soft" indicators add value…**

    … when combined

    … at longer forecast horizons (expectations, preparations)
  - ✓ **"New" indicators (Google, Credit Cards) may also be add more value if used at a higher frequency (e.g. weekly data)**

THANKS FOR YOUR ATTENTION

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Standardised approach in developing economic indicators using internet searching applications[1]

Paphatsorn Sawaengsuksant,

Bank of Thailand

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Standardized Approach in Developing Economic Indicators using Internet Searching Applications[†]

Paphatsorn Sawaengsuksant

July 2018

Abstract: This study introduces new approach of utilizing internet search data set in monitoring economic conditions. These internet search data are well-known for their ability to enhance predictive power of forecast model as well as the almost real-time availability. Yet, choosing the right searching words has always been the major hindrances for real-world application. And by raising predictive power alone is not sufficient especially for policy makers as the data could be sometimes suffered from an inconsistency as well as unintentionally sample biased from choosing the wrong words. This study introduces a more standardized approach to traverse those problematic difficulties and, at the same time, enhance reliability and economic meaningfulness of internet search data while maintaining predictive performance of indicators. In addition to the well-known Google search engine, a complementary internet application, namely the Google Correlation, is also proved to be useful in creating new economic indicators following the new introduced standardized steps. Four indicators are developed accordingly and have currently been applied in real economic condition monitoring process in the Bank of Thailand including (1) purchasing power of private household, (2) consumer confidence of private household, (3) private consumption expenditure of durable products, and (4) number of unemployed persons.

Keywords: internet search data, macroeconomic monitoring indicator, nowcast, big data

[†]Sawaengsuksant: Macroeconomic and Monetary Policy Department, Bank of Thailand (BOT), Email PaphatsS@bot.or.th. I am grateful to colleagues at BOT, especially Napat Phongluangtham, Pranee Sutthasri, Jirawat Poongam, Teerapap Pangsapa, and Nutchaphol Jaroonpipatkul, for their suggestions and encouragement.

Disclaimer: The opinions expressed in this paper are those of the author and should not be attributed to BOT. All errors are my own.

1 Introduction

Google Trend is currently one of the most common analytic tools noted by various studies and applying by policy maker units. There are many reasons behind such as timeliness, broad range of applicable study fields, friendly user interface and free data access. Choi and Varian (2009a, 2009b and 2011) shows that Google Trend Index is useful for nowcasting several economic variables, for instances, retail sales, vehicle sales, home sales, travelling demand, as well as unemployment initial claims. Vosen and Schmidt (2011) introduce new indicator for private consumption using Google Trend.

Google Correlate is also useful for analytic tasks. The Google Correlate has been launched since 2011. This interface provides a word list whose searching frequency is matched with time series inserted into the interface. Instead of trial and errors, this program learns from historical searching pattern and automatically deliveries words with high correlation to the interested series. Location of search, like countries, is also available. But it is not very common in academic literature since, in most of the time, those words fail to provide meaningful economic sense. Understanding pros and cons of these data is important to prevent data misinterpretation. Advantages and disadvantages of these data are discussed as followed;

Advantages These internet search data are well-known for their ability to enhance predictive power as well as almost real-time availability. This searching program acts like internet-based survey asking people what topics they are interested in at a certain point of time. Therefore, the potential research frontier is extremely considerable as long as internet users concern and fill in the web browser. This frontier is strikingly outbound traditional surveys. And not just timeliness, these data could perform as (1) alternative indicators to previous ones as well as (2) an answer to new questions that traditional data is too expensive or too late to conduct a survey.

Disadvantages Data are unstructured. They are not originally generated for analytic purposes. In case of traditional surveys, questions are carefully designed to answer specific questions and to minimize unrelated noises as much as possible. Meanwhile, data from Google applications provide searching frequency which is actually a bundle of numerous signals: both the informative signals and unrelated noises. Sources of noises are also different from the traditional structured data, raising a challenge in further applications' validity. For examples, noises could be generated simply from technical issues of the program interface, or from human behaviors unrelated to the questions. Bortoli and Combes (2016) note that shortness of series and lack of transparency about treatments and sampling processes are weakness of Google Trend. Therefore, careful data quality assessment process is crucially needed since using data without noticing these irrelevant noises could lead to seriously inappropriate policy decisions.

Despite earlier disadvantages, this study shows that utilizing Google Trend and Google Correlate interfaces together can deliver useful words for monitoring latest economic developments. However, without proper words filtering, set of searching words might perform well in prediction but fail to provide sensible explanation and useful insights. This study suggests general criteria to filter undesirable words, in an objective that searching frequency of filtered words are suitable for macroeconomic monitoring purpose.

This study is structured as followed; the second section explains criteria of searching words to handle unfavorable and serious noises, and the third section shows study cases

applying the criteria to construct practical monitoring indicators. Limitation of data and this method is discussed in the fourth section.

2 Criteria of Searching Word

In the real-world application, information receiving from internet search data is always unstructured and untidy. Noises occur from technical issues could lead to several problems, including inconsistent patterns of series after a change in small word punctuation or change in researcher's IP address number. There are also noises from human behavior unrelated to the question but coincidently alter patterns of series. Five criteria are introduced to filter out these noises, as followed.

2.1 First criteria: searching word is not specific to certain product or brand

This criteria objective is to confirm that the chosen words are generalized enough to cover the overall economic conditions, not to certain names. This is to avoid problems that such words are unintentionally tied to unrelated events. Although some specific names of popular products and brands might sound sensible to monitor economic conditions. However, search of those names actually represents specific shocks to the brand or product. For example, searching frequency of famous brand might be affected by competition among firms, temporary promotions or news of the brand. Some brands might even be coincidently similar to name of songs, locations, movies, or related events, which are not associated to the real economic conditions.

Figure 1

This figure shows searching frequency of the recently popular e-business brand in Thailand. The series is too short to confirm that it can capture developments of e-business in Thailand.

Source: Google Trend Interface, searching "บัตรแรบบิท" in Thailand, data as of June 2017

2.2 Second criteria: searching word covers sufficient large sample size

This criteria objective is to confirm that, the real searching number of the chosen words are sufficiently large. This is to prevent significant revision of the whole series which is arisen by a simply change in timing of using the Google Trend interface or change in researchers' IP

address number. Reactions of searching series to important economic events are also observable given that the chosen series is long enough.

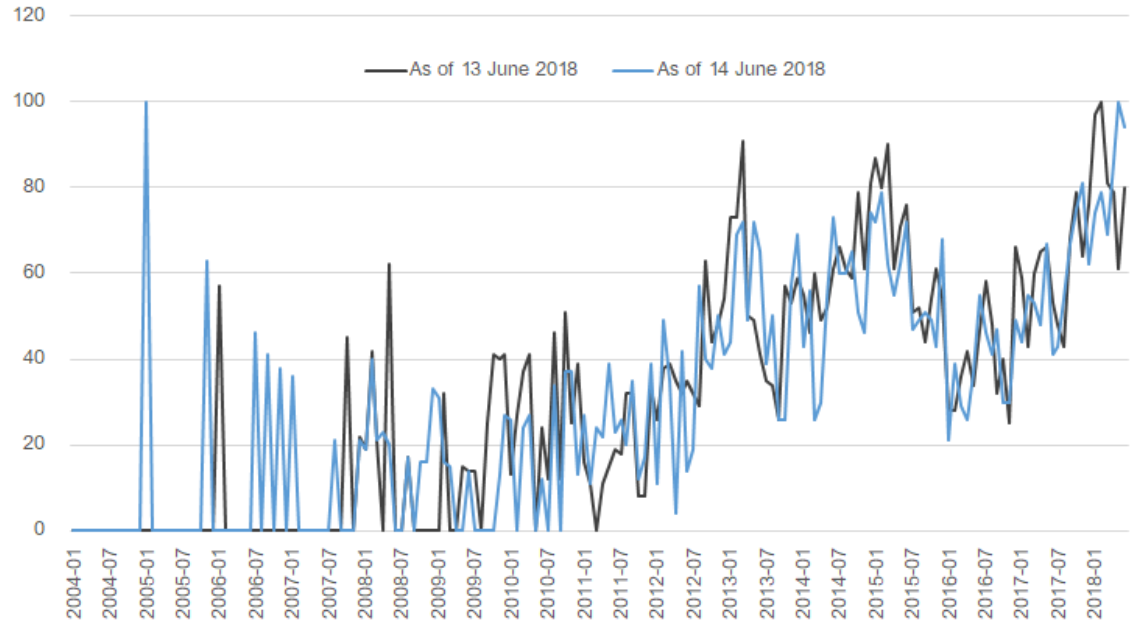
In Google Trend interface, searching frequency is normalized and ranged from zero to a hundred. Zero value means that almost no user putting this word into his web browser while a hundred value shows the most frequent search period. Some words, for examples, newly created words or those words which have becoming popular just recently, could show zero value for long time before spiking in a certain period. Performance evaluation of these kinds of words is impossible as the series is too short. This study uses only Google Trend Index which contains no zero value since 2008. The reasons are that (1) words which have been searched for at least 10 years should cover sufficiently large number of searches (2) the length of 10 years is long enough to capture major cyclical economic development, such as the Global Financial Crisis (2008 – 2010), the big flood in Thailand (2011), and the political uncertainty (2013-2014).

Moreover, the Google Trend Index is automatically calculated from sample, not the total population. Consequently, merely change in timing of using the Google Trend interface or researcher's IP address number result in "revision" of series. In case that popularity of search is not sufficiently large, a revision could significantly alter a story suggested from the series. But when total population is sufficiently large, the revision is insignificant as shown in Figure 2. Alternatively, central value, like average, median, and mode, of series can be applied to reduce this inconsistent from data revision.

Figure 2

Google Trend Index using "ซื้อกองทุนรวม", meaning "purchase mutual fund"

Every time the date or researcher's IP address number is changed, searching frequency changes significantly and returns inconsistent story.



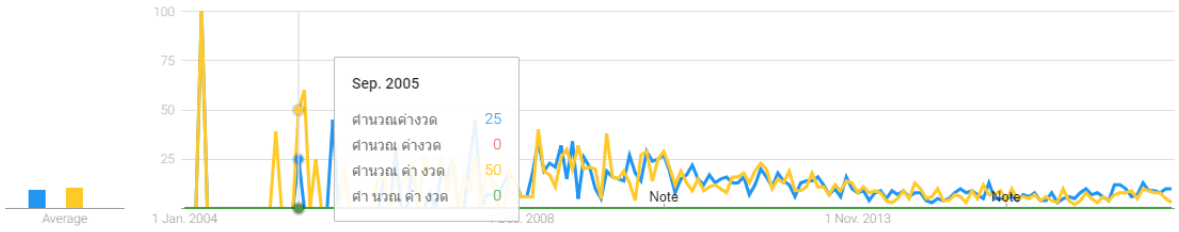
Source: Google Trend Interface, data as of June 2018

2.3 Third criteria: change in punctuation provides consistent searching series

This criteria objective is to confirm that, pattern of the whole series behave consistently after a small variation in word pattern[1]. For instances, space between two words means that searching number of those two are included, regardless of their ordering. Punctuation, however, normally appears in Thai language as separation of sentences, not words. But sometimes an insignificant punctuation cause a large change in Google Trend Index as shown in Figure 3 and 4. Different searching language may encounters this kind of noises differently.

Figure 3

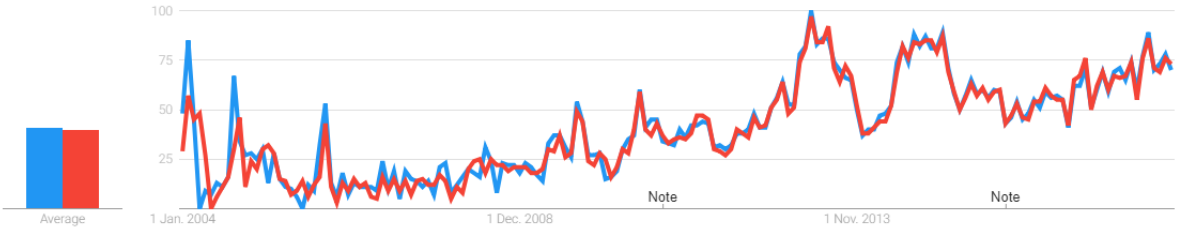
Google Trend Index searching for "คำนวณค่างวด", meaning "calculate payment", in different but insignificant punctuation, namely "คำนวณค่างวด" "คำนวณ ค่างวด" "คำนวณ ค่า งวด" and "คำ นวณ ค่า งวด". However, Google Trend Index of these words behave differently.



Source: Google Trend Interface, data as of June 2017

Figure 4

Google Trend Index searching for "ข่าวหุ้น", meaning "stock news", in different but insignificant punctuation, namely "ข่าวหุ้น" and "ข่าว หุ้น". In this case, Google Trend Index behave accordingly especially since 2008 where number of total real searches seems to be sufficiently large compared to earlier period.



Source: Google Trend Interface, data as of June 2017

2.4 Fourth criteria: searching series are statistical significant with the reference series

The objective of this criteria is to make sure that, the chosen series statistically correlated to the reference series, such as the traditional survey-based data. Simple linear

[1] For more details, https://support.google.com/trends/answer/4359582?hl=en

correlation is applied in this study. The Google Trend Index which is at least 50 percent correlated to the reference series passes this filter. Reason of the threshold 50 percent is simple, the chosen series perform better in prediction than a random guess, like using a coin toss. For example, Google Trend Index searching for "ข่าวหุ้น", meaning "stock news", results in 85 percent correlation with private consumer confident index in Thailand.

2.5 Fifth criteria: searching word provides economic meaningfulness.

The objective of this criteria is to test validity of the series, whether it provides meaningful insights related to an interested question. In other words, this criteria leaves room for economic judgment in a complementary to earlier statistical criteria. For example, according to Figure X, the word "stock news" could measure a change in household economic expectation. Private households would generally seek for investment return whenever they perceive a better economic condition or expect future favorable growth (Zatlin, 2016.)

3 Use case

Finding a set of rational words for Google Trend is one of the most challenging tasks. Practically, only weekly and monthly can be plugged into the Google Correlate interface. In this study, these five criteria are applied to create four macroeconomic indicators including (1) purchasing power of private household, (2) consumer confidence of private household, (3) private consumption expenditure of durable products, and (4) number of unemployed persons. These indicators are commonly used by policy maker in monitoring macroeconomic conditions.

3.1 Private household purchasing power

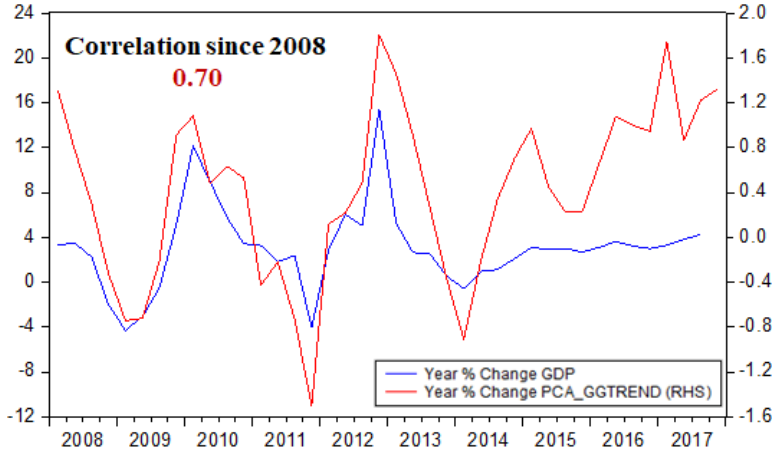
Private household purchasing power is normally correlated with Gross Domestic Product (GDP) since around 60 percent of Thailand GDP is distributed to private household. This is correspondent to the fact that, Private Consumption Expenditure (PCE) always contributes the largest share in GDP in several economies. However, both PCE and GDP are reported quarterly and usually lag for 6 - 7 weeks. In Thailand, earning of employees from Labor Force Survey (LFS) are alternatively used as private household purchasing power. These data are survey-based data collected by the National Statistical Office of Thailand (NSO). The survey is rich in both sample and data features. However, earning of employee covers only a half of total household income. The other half, such as earning of self-employed and small household businesses, are not collected.

Since GDP is available only in quarterly basis, Google Correlate is not able to suggest additional word in this case. In order to figure out potential word lists, this study utilizes basic economic framework: household income should be correlated to consumption expenditure as well as household demand to investment choices. Four words have passed all five criteria, namely, (1) "LTF + RMF" which is the most common tax-deductible and investment product for salaried employee in Thailand, (2) "ภาษี รายได้", which means "income tax", (3) "ลงทุน" which means "invest" and (4) "ซื้อ" which means "buy or purchase."

For monitoring purpose, data dimensionality reduction techniques are required to squeeze only important signals embedded in set of Google Trend Indexes. In fact, there are various tools for dimensionality reduction. This study applies Principal Component Analysis (PCA), extracting only the main common signals from all series. Correlation between annual growth of GDP and the PCA is 70 percent. (Figure 5)

Figure 5

Annual growth of GDP and PCA of Google Trend Index of four chosen words



Source : Office of the National Economic and Social Development Board, Google Trend Index using four chosen words, own calculation.

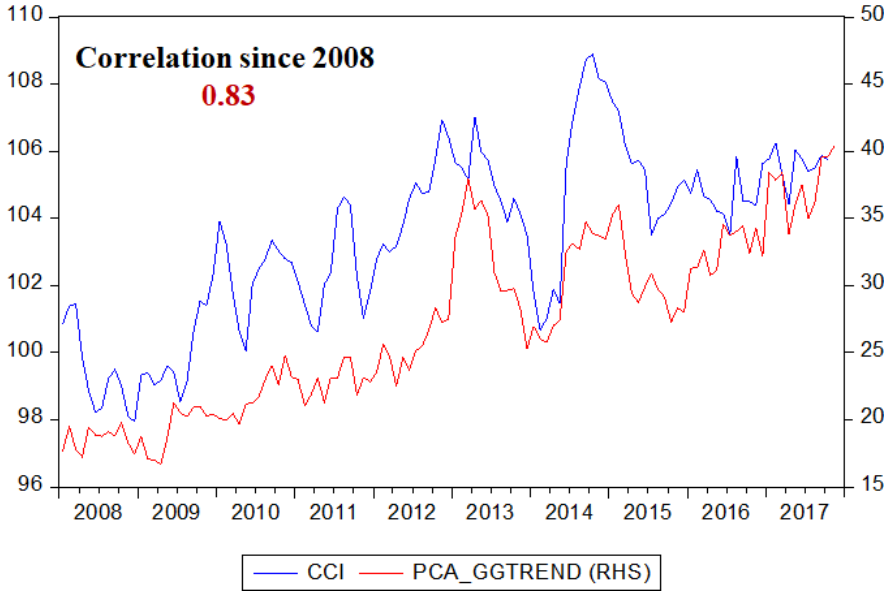
3.2 Consumer confidence of private household

In Thailand, Consumer Confidence Index (CCI) is collected from around 3,500 samples across every provinces. This survey questions private households about their perspective about current and future economic conditions, labor market conditions, as well as their current and expected future income. In addition to CCI, Indicator of Human Well Being (HHI) is surveyed by University of the Thai Chamber of Commerce asking private households about their social circumstances and change in their happiness. Both indicators are monthly reported.

According to Google Correlate, words related to investment news and stock market analysis are suggested. This is correspondent to Zatlin (2016), households' sentiment goes in line with their interest in seeking return from their assets. And nine words have passed all five criteria, namely, (1) "ข่าว หุ้น" which means "stock news", (2) "กราฟ หุ้น" which means "graph stock", (3) "กองทุน ปันผล" which means "dividend fund", (4) "หุ้น ราคา" which means "stock price", (5) "วิเคราะห์ หุ้น" which means "stock analysis", (6) "หุ้น กองทุน" which means "stock fund", (7) "ซื้อขาย หุ้น" which means "stock trade", (8) "น่า ลงทุน" which means "recommended investment", (9) "การ เล่น หุ้น" which means "invest in stock market". After dimensionality reduction, correlation between CCI and the PCA is 83 percent. (Figure 6)

Figure 6

Consumer Confidence Index (CCI) and PCA of Google Trend Index of eight chosen words



Source : Ministry of Commerce, Google Trend Index using eight chosen words, own calculation.

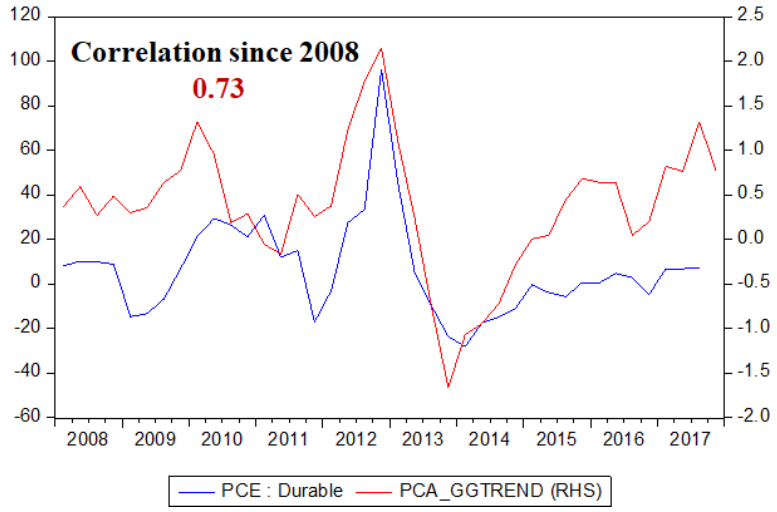
3.3 Private consumption expenditure of durable products

Share of private consumption spending for durable goods is only 10 percent of total PCE. Fast indicator of this product category, mostly vehicles, is useful because it is applicable to indicate cyclical change of macroeconomic condition (Black and Cusbert, 2010) than spending on non-durable goods and services as the latter can be more readily postponed in times of economic slowdown. Also, durable goods are expensive and their payment relies on future income flow, households tend to delay purchases of durable goods during weak business cycle.

Sales of vehicles was monthly reported by the Federation of Thai Industries (FTI) but no meaningful words are able to pass all five criteria. Therefore, this study uses two words including (1) "วัน รับ รถ" which means "the date to take a purchased car" and (2) "ทะเบียนรถ" which means "vehicle license plate". These two words are chosen because, after the purchase of vehicle, households normally wait for certain periods, around 1-3 weeks, before signing purchasing agreement and getting their vehicles. The process of registering the vehicle license plate is detailed and comprehensive and households take this time to see what they need to inspect when they receive the car. Frequency of searching these two words therefore scopes down real demand for vehicle, and filters out short-term interest of household. After dimensionality reduction, correlation of the Private spending on durable products (PCE: durable) and the PCA is 70 percent.

Figure 7

Annual growth of Private Consumption Expenditure on durable products (PCE : durable) and PCA of Google Trend Index of two chosen words



Source : Office of the National Economic and Social Development Board, Google Trend Index using two chosen words, own calculation.

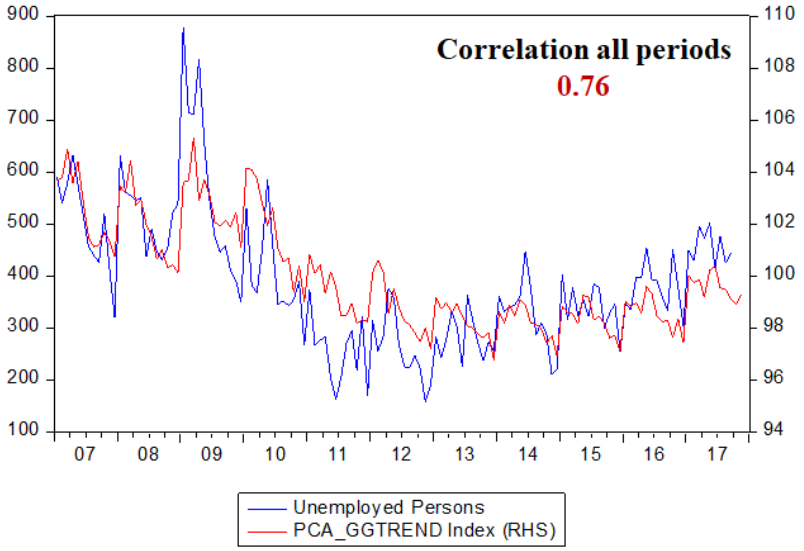
3.4 Number of unemployed persons

Number of unemployed persons is a very common macroeconomic indicators. Lekfuangfu, et al (2017) suggests three potential words that can represent unemployed persons in Thailand including (1) "หางาน" which means "finding a job", (2) "ประกันสังคม" which means "social security", and (3) "สมัครงาน" which means "register for a job". This study finds that additional five words suggested from Google Correlate Interface are also passed the five criteria including (4) "เรียน ต่อ" which means "studying higher degree", (5) "เขียน resume" which means "write resume", (6) "resume example", (7) "ปริญญาโท" which means "master degree" and (8) "ตัวอย่าง resume" which means "resume example".

These five words reflects many unemployed persons' behaviors, for example, when they are unemployed, they would search for new jobs, prepare some documents like job resume, review their social security benefit during unemployed periods, or alternatively getting higher education degree. After dimensionality reduction, number of unemployed persons and the PCA is 76 percent (Figure 8).

Figure 8

Unemployed persons and PCA of Google Trend Index of eight chosen words



Source : National Statistical Office of Thailand, Google Trend Index using eight chosen words, own calculation.

4 Limitation

The four new indicators is useful for both prediction and in-time monitoring purposes. However, some limitations are worth mentioned as followed.

4.1 These indicators informs direction, not magnitude of development. This is because the fourth criteria is based on simple correlation value, which simply measures corresponding direction of two series. Other aspects than direction would require additional statistical tools like regression model. Also, it is quite usual that magnitude of Google Trend index is higher than the real behavior suggested from the reference series.

4.2 Searching frequency of the chosen words might be affected by other sources than economic reasons. This could seriously lead to an invalid interpretation. For example, word related to flu, such as “fever” and “cough”, might not genuinely represent flu cases but news instead. In fact, the underlying algorithms may also cause a change in outturn series as well. Continuing quality assessment of word lists are required to confirm that change in searching frequency is not disturbed by other factors than the interested sources.

4.3 Frequency of search is calculated from sample not the population data. Therefore, change in date of using Google Trend interface or researcher’s IP address number would result in different Google Trend Index. However, given that searching word covers sufficiently large number of total searches, the change would simply be a small revision and implications informed by the series would be consequently consistent.

4.4 Process of developing new indicators especially choosing the right word, always takes time and requires specialized insights in certain area. Number of words suggested from the Google Correlate Interfaces are almost a hundred per single input series. And number of candidate

words could possibly reach around a thousand. Although four criteria have already straight-forward statistical tools, prudential economic judgment is still necessary to assess validity of economic reasons. Complementary information, such as traditional indicators or knowledge of related topics, is also required to ensure validity of the indicator.

5 Conclusion and Discussion

Google data is proved to be useful in many cases and in many countries. But improving prediction performance alone cannot guarantee validity of conclusion. With different types of prediction model, including an autoregressive model, adding Google Trend would not always improve model performance (Bortoli and Combes, 2016.) Ones may try to harness the richness of Google data, by bundling all words suggested from Google Correlate interface or several words categories in Google Trend interfaces, and let the data speak. This approach creates a serious risk to the policy makers as there is totally no useful insight to support decisions. Strong prediction performance can also arise simply because of an overfitting problem. Instead, the most important task is to extract new and useful insights from new data source as much as possible. Not only that the outturn series could suggest an accurate story, additional knowledge gained from the data can considerably support policy decisions especially in the case of unexpected and unprecedented shocks. This study suggests a guideline for filtering words which are generalized enough for macroeconomic monitoring purpose: total search numbers is large enough and meaning of word tends to reflect overall condition, not a specific or unrelated event. Statistical test and prudential assessment are also suggested to enhance validity and economic meaningfulness of the chosen word.

Reference

Andrew Zatlin (2016), Google Data Points to Strong Consumer Sentiment, URL http://www.moneyballeconomics.com/google-data-points-to-strong-consumer-sentiment/

Black and Cusbert (2010), Durable Goods and the Business Cycle, Reserve Bank of Australia Bulletin September, pp 11-18, URL https://www.rba.gov.au/publications/bulletin/2010/sep/pdf/bu-0910-2.pdf

Greg Tkacz (2013), Predicting Recession in Real-Time: Mining Google Trends and Electronic Payments Data for Clues, C.D. Howe Institute Commentary

Hyunyoung Choi and Hal Varian (2009a), Predicting the Present with Google Trends, Technical report, URL http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf

Hyunyoung Choi and Hal Varian (2009b). Predicting initial claims for unemployment insurance using Google Trends. Technical report, URL http://research.google.com/archive/papers/initialclaimsUS.pdf

Hyunyoung Choi and Hal Varian (2011), Predicting the Present with Google Trends, URL http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf

Iren Onder (2017), Forecasting Tourism Demand with Google Trends: Accuracy Comparison of Countries Versus Cities, International Journal of Tourism Research

Konstantin A. Kholodilin, Maximilian Podstawski, and Boriss Siliverstovs (2010), Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US, DIW Berlin Discussion Paper No.997

Lynn Wu and Erik Brynjolfsson (2015), The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales, URL http://www.nber.org/chapters/c12994

Nick McLaren and Rachana Shanbhogue (2011), Using Internet Search Data as Economic Indicators, DIW Berlin Discussion Paper No.899

Rawley Z. Heimer, Timothy Stehulak, and Daniel Kolliner (2015), Assessing Consumer Confidence with Google Search Terms, Federal Reserve Bank of Cleveland

Simeon Vosen and Torsten Schmidt (2011), Forecasting private consumption: survey-based indicators vs. Google trends, Journal of Forecasting, Volume 30, Issue 6 URL https://onlinelibrary.wiley.com/doi/abs/10.1002/for.1213

Stephanie Combes and Clemen Bortoli (2016), Nowcasting with Google Trends, the more is not always the better, First International Conference on Advanced Research Methods and Analytics, URL http://www.carmaconf.org/carma2016/wp-content/uploads/pdfs/4226.pdf

Xinyuan Li (2016), Nowcasting with Big Data: is Google useful in Presence of other Information?, URL https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IAAE2016&paper_id=215

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Standardised approach in developing economic indicators using internet searching applications[1]

Paphatsorn Sawaengsuksant,

Bank of Thailand

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Standardized Approach in Developing Economic Indicators using Internet Searching Applications

Paphatsorn Sawaengsuksant

July 2018

Disclaimer: The opinions expressed in this paper are those of the author and should not be attributed to the Bank of Thailand. All errors are my own.

Macroeconomic Indicators from Google Trend and Google Correlate

Advantages:

…. Almost Real time: monthly, weekly, and daily basis

…. Reflecting Real Activities: comparable to internet-based survey asking topics which citizens are interested in

…. Applicable for several research areas: Household and Business Sentiment, Private Household Purchasing Power, Private Consumption Expenditure, Unemployment, Demand for tourism sectors, Property Price, E-Commerce, Popularity of certain policy tools etc.

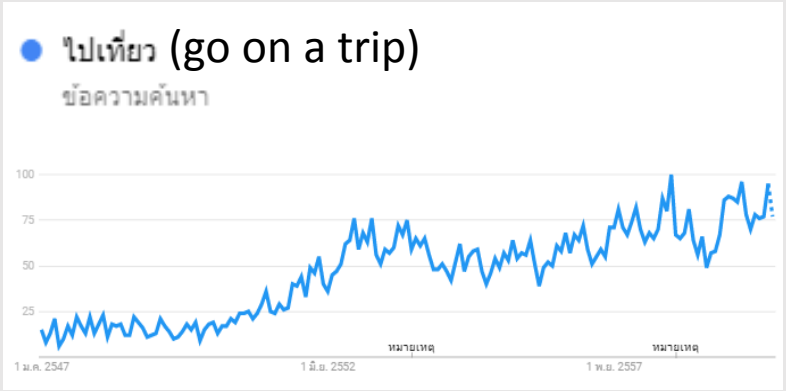
…. User friendly and free access

Google Trends and Correlate Interface

Google Trends

Input : Reference words
Output : Time Series

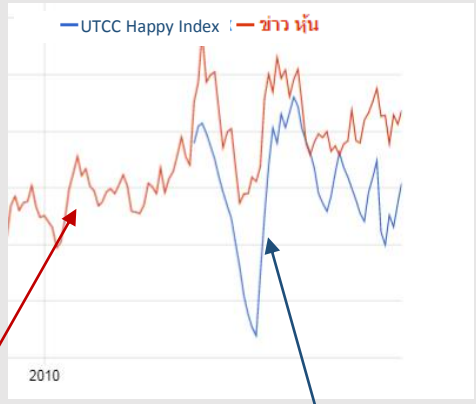
Searching Frequency Index



Google Correlate

Input : Reference Time Series
Output : Correlated Words

List of highly correlated (but mostly spurious) words



Correlated with **UTCC Happy Index**

0.7617	ข่าว หุ้น (Stock News)
0.7529	www.mybycat.com
0.7512	พร บ ตำรวจ แห่ง ชาติ (Cop Act)
0.7509	หุ้น nusa
0.7471	กองทุน หุ้น (Stock and Equities)
0.7411	gorilla glass (Screen Protector)
0.7394	5 chord (Guitar Chord)
0.7370	หัน หุ้น (Stock News)
0.7367	รูป รถยนต์ (Car Image)
0.7347	サイコパス (Anime Name)
0.7343	หุ้น cash balance

Searching Frequency of "ข่าวหุ้น" or "Stock News"

Reference Series: Household Happiness Index

Source: University of the Thai Chamber of Commerce

New Data = New Challenges

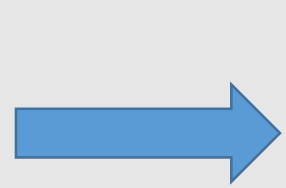
…. Data is unstructured: not initially collect for analysis purpose

…. Garbage In = Garbage Out ⇒ Policy Decision??

…. Choosing the **right searching word** is very crucial to construct **reliable economic indicators**, especially for policy maker.

(1) Words suggested from the Google Correlate are mostly spurious; not always provide <u>meaningful insights</u>.

Household Sentiment Index



(My Dog) (Screen Protector) (Stock News) (Counterfeit)
(Car Image) (Watch Movie) (Guitar Chord) (Cop Act)
(Public Officer Job Title) (My Pet) (Migration) (Anime Name)
(My Cat) (Prove of Citizen) (Music Name) (Alienate worker)

New Data = New Challenges

…. Garbage In = Garbage Out ⇒ Policy Decision??

…. Choosing the **right searching word** is very crucial to construct **reliable economic indicators**, especially for policy maker.

(2) Words inserted in the Google Trend might not be <u>sufficiently general</u>, for example, the word is too new, too specific, or covers too small number of searches.

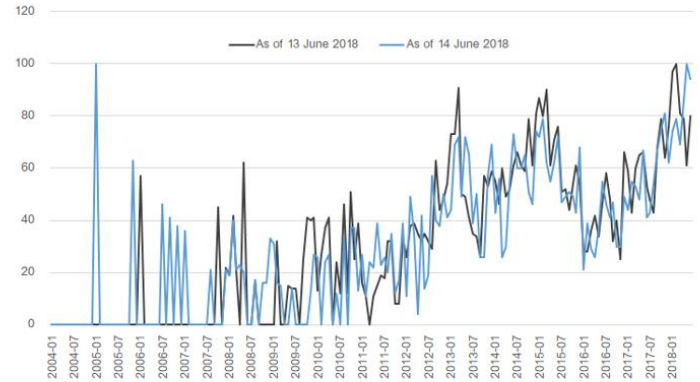
Figure 1

This figure shows searching frequency of the recently popular e-business brand in Thailand. The series is too short to confirm that it can capture developments of e-business in Thailand.

Source: Google Trend Interface, searching "บัตรแรบบิท" in Thailand, data as of June 2017

Figure 2

Google Trend Index using "ซื้อกองทุนรวม", meaning "purchase mutual fund"



Source: Google Trend Interface, data as of June 2018

Word Filtering

for Generalization of words (1st-3rd filters) and Economic Intuition (4th-5th filters).

(1) Not specific to certain brands

(2) Sufficiently popular words
google trend returns non-zero value since 2008 so that (2.1) series are not too sensitive to change in timing of search and (2.2) series' behaviors during all significant cycles can be observed

(3) Not sensitive to insignificant word change, like word ordering or innotation

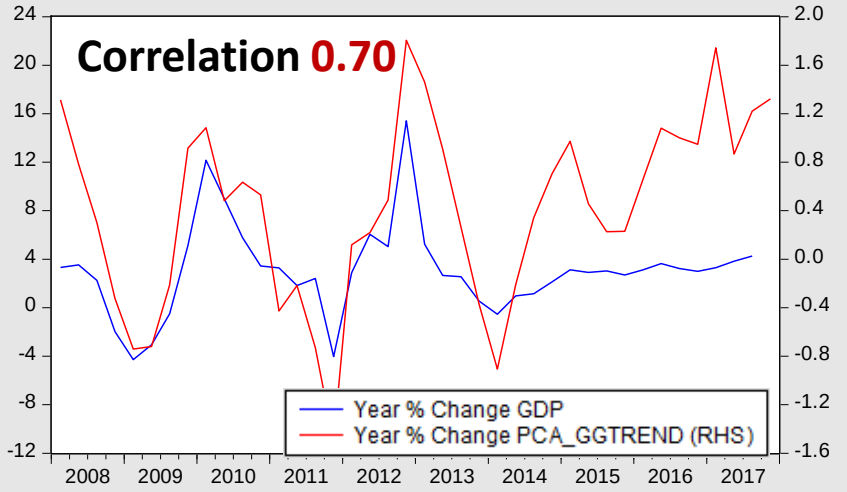
(4) Correlation with Alternative Series > 0.5

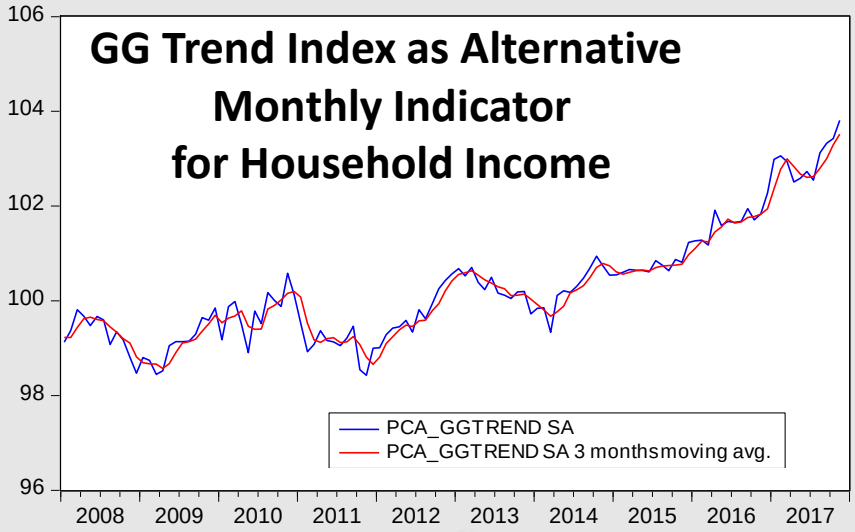
(5) Strongly Intuitive

Application :

I. Household's Income (Quarterly Report)

Reference Series	4 Word Lists	Correlation with Real GDP
-	LTF + RMF (Popular Income Tax Reduction Products in Thailand)	0.74
-	ภาษี รายได้ (Income Tax)	0.56
-	ลงทุน (Investment)	0.61
-	ซื้อ (Purchase)	0.50

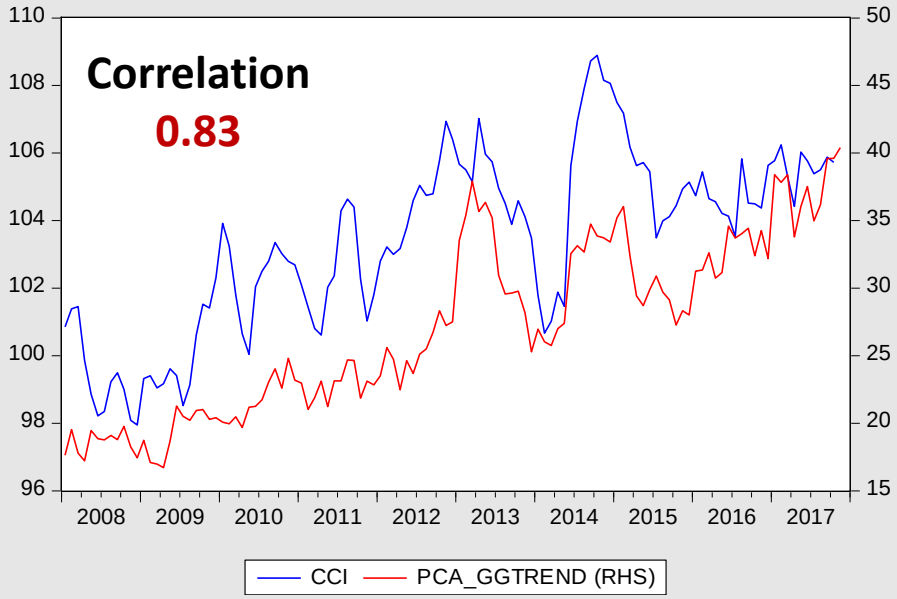




Application :

II. Household's Sentiment (Monthly Report)

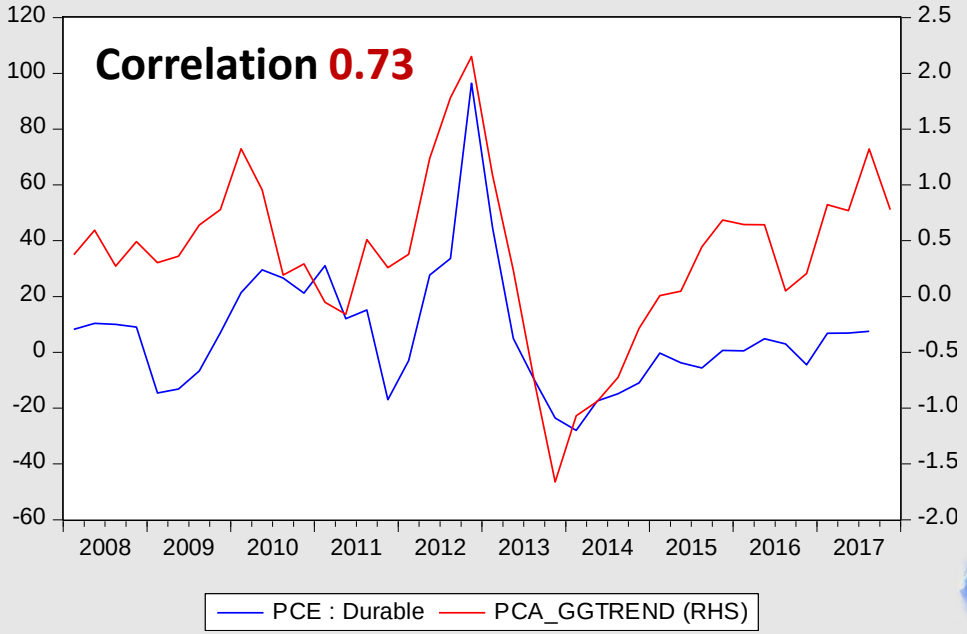
Reference Series	9 Word Lists	Correlation with CCI
Consumer Confident Index: CCI (Source: Ministry of Commerce, Thailand)	ข่าว หุ้น (Stock News)	0.85
	กราฟ หุ้น (Stock Chart)	0.83
	กองทุน ปันผล (Dividend Fund)	0.83
	หุ้น ราคา (Stock Price)	0.82
	วิเคราะห์ หุ้น (Stock Analysis)	0.81
	หุ้น กองทุน (Purchase Fund)	0.76
	ซื้อขาย หุ้น (Stock Purchase)	0.70
	น่า ลงทุน (Profitable Investment Products)	0.68
	การ เล่น หุ้น (Stock Investment)	0.55



Application :

III. Private Consumption Expenditure (PCE) : Durable Goods (Quarterly Report)

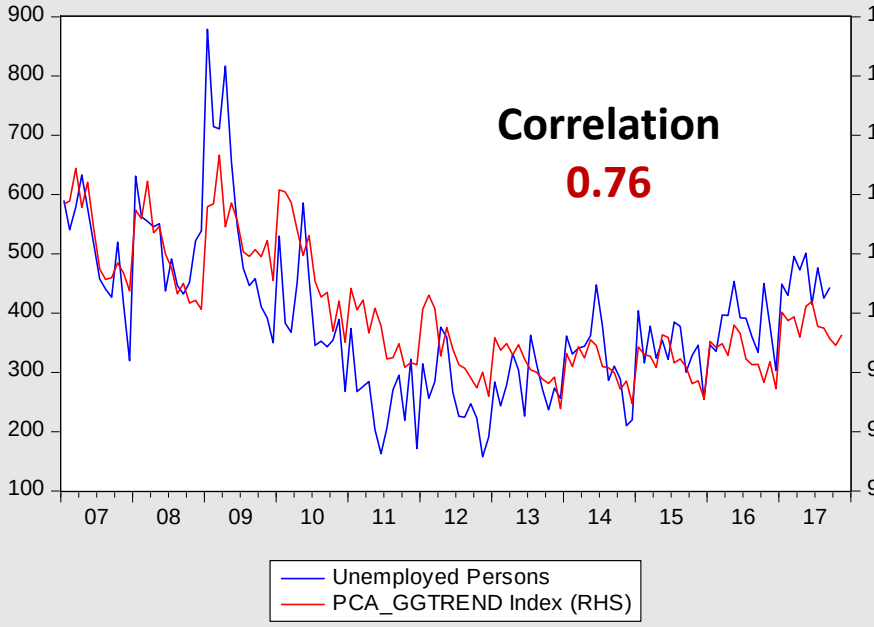
Reference Series	2 Word Lists	Correlation with PCE Durable Goods
PCE: Durable Goods	วัน รับ รถ (the date to get a purchased car)	0.63
-	ทะเบียนรถ (vehicle license plate)	0.66



Application :

IV. Unemployment Rate (Monthly Report)

Reference Series	8 Word Lists	Correlation with Unemployed Persons
Unemployed Persons	หางาน (Finding Job)	0.77
	เรียน ต่อ (Further Study)	0.66
	เขียน resume (Writing Resume)	0.64
	resume example	0.63
	ปริญญา โท (Master Degree)	0.56
	ประกันสังคม (Social Security)	0.56
	ตัวอย่าง resume (Example of Resume)	0.54
	สมัครงาน (Job Application)	0.48



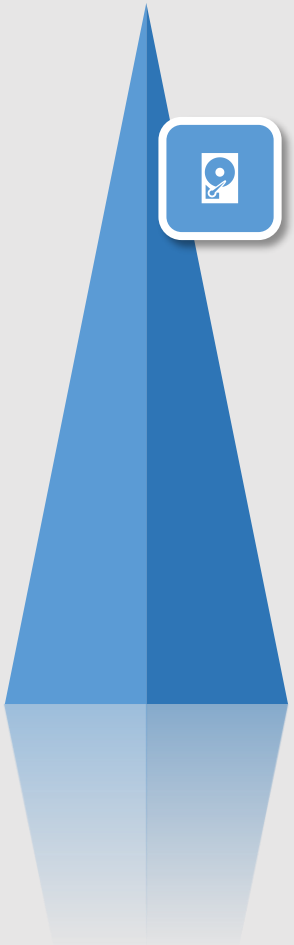
Source : Labor Force Survey, "Labor Market Insights: The Power of Internet-Based Data" *Presented at Bank of Thailand Symposium (15 Sep 2016).*

Summary :

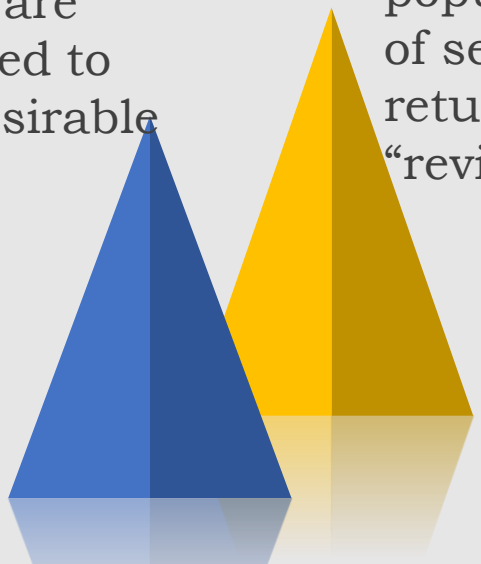
Application : **Early and Alternative Indicators**	**Alternative and Early Indicators for**	**Searching Words List**	**Performance:** **Corr. between PCA Series and Alternative Series**
HH Income	Gross Domestic Products Source: NESDB	LTF + RMF, ภาษี รายได้, ลงทุน, ซื้อ	0.70
HH Sentiment	Consumer Confidence Index Source: UTCC and MOC	ข่าวหุ้น, กราฟหุ้น, กองทุน ปันผล, หุ้น ราคา, วิเคราะห์ หุ้น, กองทุน หุ้น, ซื้อ ขาย หุ้น, น่า ลงทุน, การเล่นหุ้น	0.83
Consumer's Expenditure on Durable Goods	Private Consumption Expenditure : Durable Goods Source: NESDB	วัน รับ รถ, ทะเบียน รถ	0.73
Unemployment	Unemployment Source: Labor Force Survey	หางาน, เรียนต่อ, เขียน resume, resume example, ปริญญาโท, ประกันสังคม, ตัวอย่าง resume, สมัครงาน	0.76

Limitation and Future Development :

Google Trend Information is subjected to changes in households' searching behavior. Instead of consumption/income determinants, change in series might be contaminated by certain changes in searching behavior. Close monitoring of individual series is required.



Google Series perform well as indicators of growth momentum, not magnitude. Additional methods/filters, like bivariate model, are specifically needed to capture other desirable aspects.



Series Index is subject to timing of search. Like other data, google trend is drawn from samples, not population. Different timing of search will therefore returns slightly different or "revised" series index.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Measuring stakeholders' expectations for the central bank's policy rate[1]

Alvin Andhika Zulen and Okiriza Wibisono,

Bank Indonesia

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Measuring stakeholders' expectations for the central bank's policy rate

Alvin Andhika Zulen[1] and Okiriza Wibisono[2]

Abstract

In recent decades, the role of market expectation on central bank's policy rate has been increasingly acknowledged in monetary policy formulation. In this research, we develop a machine learning-based technique for identifying the expectation of stakeholders on Bank Indonesia's policy rate. The expectations are extracted from news, starting from 14 days before the monthly Board of Governor's meeting. We achieve an F1-score of 76.8% from out-of-sample evaluation on classification result. The resulting monthly expectation index has 78.6% correlation with the index generated from Bloomberg's monthly survey.

Keywords: policy rate expectation; text mining; machine learning; big data

JEL classification: C02, E52, E58

[1] Statistics Department – Bank Indonesia; E-mail: alvin_az@bi.go.id

[2] Statistics Department – Bank Indonesia; E-mail: okiriza_w@bi.go.id

Contents

1. Background

Expectations on future economic conditions are among the factors that greatly influence the economic actors in making decision. If consumers expect higher inflation in the future, then they increase their consumption expenditures in the present.

One of the indicators that central banks consider in formulating monetary policy is markets' expectation on policy rates. Quoting (Fischer, 2017), "... those times when financial markets and the central bank have different expectations about what a central bank decision will be. Such situations lead to surprises and often to market volatility. " The main objective in measuring expectations on central bank's policy rate is to avoid market volatility that occurs when market participants have different expectations from the monetary policy taken by central bank. Unexpected movement of Fed Fund Rate is proven in affecting yields of Treasury Bills (Kuttner, 2000) and stock prices (Bernanke & Kuttner, 2004) significantly. If central bank will take a monetary policy that is different from market expectations, a communication strategy is needed so that the volatility in financial markets can be minimized (Fischer, 2017). In addition to avoid volatility, the measurement of policy rate expectations can be an input for projection of macroeconomic indicators, such as inflation and GDP, as implemented by the Monetary Policy Committee (MPC) of the Bank of England (Joyce & Meldrum, 2008).

Because the variable is unobservable, the measurement of expectation on economic indicators, including policy rates, is a nontrivial task. There are two main methods for measuring expectations, i.e. market-based method and survey-based method.

In market-based method, expectations are estimated based on the movement of the price of certain instruments in financial markets. For example, in U.S. financial markets, there is Fed Funds Futures instrument that serves for hedging against changes in The Fed's monetary policy. The price of this instrument is linked directly with the average of overnight Fed Funds Rate. If the average is decreased then the price of Fed Funds Futures will go up, and vice versa. Thus, expectation on policy rate can be estimated from Fed Funds Futures prices, and changes in expectation are estimated from the instrument's price movement.

For countries with no interest rate hedging instruments similar to Fed Funds Futures, the measurement of expectation is based on the price of the instrument that moves along with the policy rate, e.g. Treasury Bills, unsecured interbank loan, Forward Rate Agreement (FRA), and Overnight Index Swap (OIS) (Joyce et al., 2008). Nevertheless, measurement with those instruments is more difficult because of additional factors that contribute to pricing, such as credit risk, liquidity risk, and term premium. It is necessary to apply specific calculations and assumptions to exclude these factors in order to obtain an accurate expectation on policy rates.

Survey-based method offers a simpler alternative to measure policy rate expectation. In this method, the survey institution (which can be the central bank itself) asks respondents directly about their expectation on policy rate in the future. This method is also in accordance with recommendation in Manski (2004) that the expectation level can't be inferred only from the observed choice or action (revealed preference analysis). An expectation measure should be supported by numbers that are explicitly expressed by the respondents.

In Indonesia, Bloomberg conducts a monthly survey of expectations on Bank Indonesia's policy rate (BI 7-day Reverse Repo Rate, formerly BI Rate), i.e. the Economist Estimates Survey. Respondents of the survey are mostly from banking and securities company. Approximately, two weeks before the monthly Board of Governor's Meeting, Bloomberg asked 20-30 respondents about their estimation of Bank Indonesia's policy rate that will be set in the meeting.

This research aims to develop a new measure of stakeholders' expectation on Bank Indonesia's policy rate, as a complement to the Bloomberg survey. From methodological perspective, we show how to utilize textual data to develop the new measure, by employing machine learning-based technique. Based on our observations, a fair amount of expectations on policy rate are quoted in news articles, as seen in Figure 1. Expectations quoted in the news tend to have more varied sources. In addition to market participants, governments, authorities (e.g. Financial Services Authority (OJK), Deposit Insurance Corporation (LPS), Indonesia Stock Exchange (BEI)), and real sector entrepreneurs often express their expectations on Bank Indonesia's policy rate. Hence, it has potential to be used as data source for measuring the expectations.

The paper is organized as follows. In section 2, we provide literature reviews on measuring policy rate expectation and text mining for economic news. In section 3, we discuss the data and methodology. In section 4, we provide a summary of the results and evaluation of the model. In section 5, we conclude the paper and offer some thoughts for future works

Example of Expectation on Bank Indonesia's Policy Rate in News Articles — Figure 1

2. Literature Review

2.1 Survey-Based & Market-Based Method for Measuring Expectation on Policy Rate

Questions on policy rate expectations have been included in various economic and financial surveys. For example, Christensen & Kwan (2014) used the monthly Blue Chip Financial Forecast survey and Survey of Primary Dealers to evaluate whether expectations of market participants are aligned with expectations of the Federal Open Market Committee (FOMC) expectations or not. At Bank Indonesia, the results of Bloomberg survey as described in the previous section are utilized to provide information on policy rate expectation in the Board of Governors Meeting.

Survey-based method has a major advantage over market-based method, i.e. simpler for analysis. Several studies (Christensen & Kwan, 2014; Joyce & Meldrum, 2008; Friedman, 1979) used average or median values to aggregate policy rate expectations of all respondents. For comparison, a research with market-based method (de los Rios & Reid, 2008) used three instrument prices for estimating the probability of Bank of Canada's policy rate changes.

In addition to simpler analysis, we can also calculate the distribution of respondents' expectations with survey-based method. If there are 30 respondents, for example, we can calculate the percentage of respondents who expect a policy rate cut and the percentage of respondents who expect a policy rate hike. In market-based method, the distribution of these expectations can't be provided (Christensen & Kwan, 2014).

However, survey-based method also has several disadvantages compared to market-based method. Market-based method captures the real expectation in the market, i.e. the price of the instrument will move along with market expectation because they are "risking" their money in the instrument (money on the line). Given its subjective nature, in survey-based method, it's possible that the respondents didn't respond according to their actual expectation. Another disadvantage is that the survey-based method is not practical to be done in high frequency (e.g. daily), whereas with market-based method, expectation can be calculated on a daily basis or even from hour-to-hour, if the referred instruments are widely traded.

2.2 Text Mining on Economic News

Text data have been widely used for research in economics and finance. Sahminan (2008) identified keywords that reflect a tight, neutral, or loose monetary policy inclination in the press release statement of Bank Indonesia over the period from January 2004 to December 2007. The econometric analysis shows that monetary policy statements that contain loose or neutral policy inclination tend to lower interbank interest rates, while monetary policy statements with tight policy inclination tend to have no impact on interbank interest rates (asymmetric effect). Rosa & Verga (2007) applied similar method to analyze the impact of European Central Bank (ECB) press releases.

In those studies, the identification of keywords in the press release text is done manually. Researchers read the press releases one by one and record the keywords that appear in the press releases. Nowadays, text mining algorithms are growing

rapidly along with the adoption of big data and machine learning. These algorithms can automatically "read" and "extract" relevant information from the text, such as the person's name, the organization's name, and the keywords. Compared to the manual way, text mining allows us to make use of much larger text data than press releases, including news and social media.

Bollen et al. (2011) proved that the mood expressed by Twitter users can be analyzed to improve the stock market prediction. Moods are identified using keywords, e.g. "I feel ..." and "I'm ...", and then categorized into different types of mood by using OpinionFinder and Google-Profile of Mood States (GPOMS). Similar to (Bollen et al., 2011), O'Connor et al. (2010) created a public sentiment index from positive and negative word occurrences in economic related tweets. This index correlated with the Gallup's Economic Confidence Index at 73.1% and with the Index of Consumer Sentiment (ICS) from the Reuters/University at 63.5%.

In addition to social media data, news data are also widely used to analyze economic conditions. Baker et al. (2016) developed an Economic Policy Uncertainty (EPU) index by using news articles from 10 leading U.S. newspapers. The EPU index reflects the frequency of articles that contain the following trio of terms: economic (''economic'' or ''economy''); policy (''Congress,'' ''deficit,'' ''Federal Reserve,'' ''legislation,'' ''regulation,'' or ''White House), and uncertainty (''uncertain'' or ''uncertainty''). The EPU indexes have also been constructed for 11 other countries with list of keywords that are tailored to the language and economy.

In terms of monetary policy, Nardelli et al. (2017) developed the Hawkish-Dovish (HD) index that measures media's perception of ECB communications. The HD index is computed by using two methods: semantic orientation (SO) and support vector machine (SVM). The HD index based on SO method is computed by counting the co-occurrences of strings with a fixed set or pre-determined words/expressions that are normally associated with "hawkish" and "dovish" concepts to determine the tone of the document. For the SVM method, instead of using predefined set of keywords, the algorithm automatically looks for patterns in text documents to select the words with the highest discriminative power and determines the tone of a document based on them. Similar Hawkish-Dovish research has also been done earlier by Lucca & Trebbi (2009) for the FOMC statements.

Those two studies measured media's perception after each press conference following monetary policy meetings. As far as our observation, there is no research utilizing news data to measure policy rate expectation before the monetary policy meetings.

3. Methodology

3.1 Data

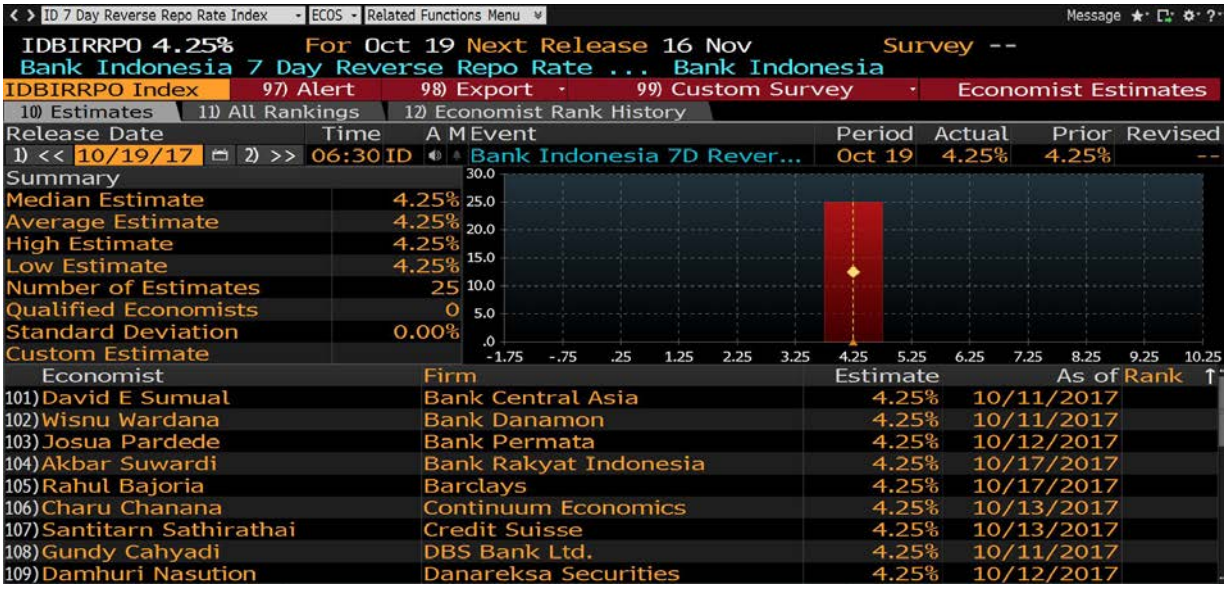
3.1.1 News Articles

The news data used in this research obtained from Bank Indonesia's Cyber Library. Cyber Library is an internal repository of news articles related to economic and financial topics. The news articles data are available on a daily basis since 1999, thus covering the whole period since Bank Indonesia set the policy rate (BI Rate) in July 2005. The data used in this research are from January 2006 to February 2018.

3.1.2 Policy Rate Expectation Survey

In order to measure the accuracy of policy rate expectation obtained from the news, a benchmark indicator is required for comparison. In this research, we use Economist Estimates Survey from Bloomberg, as described in the first chapter. Survey results are available starting from two weeks before the monthly Board of Governor's meeting, although data from several respondents are often only available close to the date of the meeting. Each respondent gives their estimation on Bank Indonesia's policy rate which they think will be set in the meeting. An example of the survey result is shown in Figure 2.

Example of Bloomberg's Economist Estimates Survey — Figure 2



3.2 Machine Learning Model

In order to extract the policy rate expectation from news articles automatically, we build a text mining model by using machine learning-based technique. This section will describe the steps taken in developing the model.

3.2.1 Data Filtering

News articles collected from Bank Indonesia's Cyber Library are not entirely relevant for measuring policy rate expectations. First of all, the news articles are filtered in following steps:

1. Publication Date Filtering

   From all the news articles available in Cyber Library, we only used news articles that are published within 14 to 1 days prior to each monthly Board of Governor's meeting.

2. Sentences Spliiting

   News articles are splitted into sentences to simplify the extraction of policy rate expectation. Text splitting is done automatically by using Natural Language Toolkit (NLTK) in Python.

3. Keywords Filtering

Sentences from the previous step are filtered again, leaving only sentences that contain keywords related to Bank Indonesia's policy rate, e.g. "BI Rate", "BI 7-days reverse repo rate", and "Bank Indonesia's policy rate".

Thus, the result from these stages is a collection of sentences containing keywords related to Bank Indonesia's policy rate and published on D-14 to D-1 prior to each monthly Board of Governor's meeting. In total, there are 5,700 news articles (2% of overall news in Cyber Library) and 16,000 sentences (0.2% of overall sentences in Cyber Library) that meet the specified criteria.

3.2.2 Annotation

Text mining that make uses of machine learning techniques require annotated datasets for training the algorithms. Annotation is the process of attaching additional information into a collection of texts. Annotation is needed to "teach" the text mining algorithm how to extract the information from the texts, so that the process can be done automatically in the future.

In this research, annotation is done on sentence-level, as the smallest data unit. We added a categorical information about policy rate expectation to each sentence, with 4 (four) possible values as follows:

1. 0: sentence with no expectation information;
2. 1: expecting no change in policy rate;
3. 2: expecting policy rate hike;
4. 3: expecting policy rate cut.

This categorical information will be used as target class in machine learning algorithms.

Each sentence is annotated by two annotators to minimize human error and subjectivity. If a sentence is annotated differently by both annotators, the sentence will be annotated by the third annotator. We also provide an annotation guidance so that the annotations can be given consistently by each annotator.

In total, we collected 4,445 sentences that have been annotated, out of 16,000 sentences generated in previous steps. Table 1 shows the proportion of sentences for each policy rate expectation category.

Annotated Sentences — Table 1

Policy Rate Expectation Category	Number of Annotated Sentences	Percentage (%)
Policy Rate Hike	355	8%
Policy Rate Cut	660	15%
Policy Rate Unchanged	490	11%
No-Expectation	2,940	66%

3.2.3 Pre-processing

After annotating the sentences, one more step is required in order to start training the classification model using machine learning algorithms. Each sentence must be

transformed into numerical vector, because machine learning algorithms can only process numerical data.

Each sentence is transformed into numerical vector that contains following information:

1. bag-of-keywords: number of keywords' occurences in the sentence;
2. number of words in the sentence;
3. number of characters in the sentences;
4. numbers and percetages quoted in the sentence;
5. word embedding vector.

All transformations are done by using Pandas and Scikit-learn libraries in Python.

3.2.4 Model Construction

Sentences that have been annotated and transformed into numerical matrix (1 line = 1 sentence) are used as input for machine learning algorithms. Machine learning algorithms will learn the patterns in input data to construct classification model with target function to classify the category of policy rate expectation.

$$\hat{f}(sentence\_vector) \in \{rate\ hike, rate\ cut, rate\ unchanged, no\ expectation\}$$

The data are splitted into 2 datasets: training dataset and test dataset. Training dataset is used to build the classification model. Test dataset is used in model evaluation to provide unbiased evaluation on the model. We split the data using approximately 80:20 ratio (training dataset: 3,645 sentences; test dataset: 800 sentences).

We use 5 (five) machine learning algorithms in this research to find the best classification model for solving the task, i.e.:

1. Logistic regression: modeled the linear relationship between independent variables and the expectation category as dependent variable;
2. Naïve bayes: modeled the probability of expectation category based on Bayes' theorem with the independence assumptions between predictors;
3. Decision tree: modeled the decision tree that predict expectation category (represented in the leaves) based on a set of decsion rules (represented in the branches);
4. Random forest: combined the predictions of multiple decision trees with bootstrapping aggregation; and
5. Xgboost: an implementation of gradient boosted tree by DMLC (http://dmlc.ml/).

3.3 Index Calculation

3.3.1 Expectation Index from News

The best classification model that has been constructed in previous step is then applied to classify the policy rate expectation category on all 16,000 sentences in the dataset. From the classification results, we calculate the monthly policy rate expectation index in following steps:

1. Each sentence with policy rate expectation is given a score: +1 for expecting policy rate hike; -1 for expecting policy rate cut; 0 for expecting no change in policy rate. Sentences with no information on policy rate expectation ware excluded from index calculation.
2. Each news article is given a score: the mean score of sentences (as calculated in 1st step) in the article.
3. The expectation index from news for month t is defined as the mean score of articles (as calculated in 2nd step) that are published in that month.

$$Expectation\ Index\ News_t = \frac{1}{|C_a|}\sum_a score(a) = \frac{1}{|C_a|}\sum_{s_a}\left(\frac{1}{|C_{s_a}|}score(s_a)\right)$$

$|C_a|$ = number of articles in month t

$score(a)$ = score of article a

$|C_{s_a}|$ = number of sentences in article a with policy rate expectation

$score(s_a)$ = score of sentence s in article a

The monthly expectation index has following characteristics:

- Range of index: [-1,+1].
- The index will be close to +1 if there are more news with expectation of policy rate hike.

  The index will be close to 0 if there are more news with expectation of unchanged policy rate.

  The index will be close to -1 if there are more news with expectation of policy rate cut.
- Positive index means more news with expectations of policy rate hike compared to policy rate cut.

  Negative index means more news with expectations of policy rate cut compared to policy rate hike.
- If $index_{t1} > index_{t2}$ then the proportion of news with expectation of policy rate hike is greater in t_1 than in t_2.

  If $index_{t1} < index_{t2}$ then the proportion of news with expectation of policy rate cut is greater in t_1 than in t_2.

3.3.2 Expectation Index from Bloomberg Survey

As described earlier in section 1, in the Economist Estimates Survey, Bloomberg asked respondents about their estimation on Bank Indonesia's policy rate that will be set in the next Board of Governors' meeting. These estimation numbers need to be converted so that they are comparable with the expectation index. The conversion is done as follows:

$$score(x)_t = \begin{cases} +1 : if\ prediction(x)_t > BI\ Rate_{t-1} \\ 0 : if\ prediction(x)_t = BI\ Rate_{t-1} \\ -1 : if\ prediction(x)_t < BI\ Rate_{t-1} \end{cases}$$

$score(x)_t$ = score of respondent x in month t

$prediction(x)_t$ = policy rate prediction respondent x in month t

$BI\ Rate_{t-1}$ = Bank Indonesia's policy rate in month $t-1$

The expectation index from Bloomberg survey for month t is defined as the mean score of all respondents in the month.

$$Expectation\ Index\ Bloomberg_t = \frac{1}{|C_x|}\sum_x score(x)$$

$|C_x|$ = number of respondents in month t

4. Result & Analysis

4.1 Classification Model Evaluation

Classification models that have been trained in the previous steps need to be evaluated in order to measure their accuracy in predicting the target class (i.e. policy rate expectation). We use F1-score as metric for evaluation, in order to get a balanced classification model with the optimal balance of recall and precision.

The result of out-of-sample evaluation for each machine learning model are given in Table 2. We can see that the logistic regression model has the best F1-score (76.8%), compared to other machine learning models. The model also has the best accuracy and precision score. Hence, the logistic regression model becomes our choice for measuring policy rate expectations in the following sections.

Classification Model Evaluation — Table 2

Classification Model	Accuracy	Recall	Precision	F1
Logistic regression	83.4%	**83.2%**	71.2%	**76.8%**
Naïve bayes	80.6%	83.2%	64.5%	72.7%
Decision tree	73.0%	65.7%	53.4%	58.9%
Random forest	78.0%	72.6%	63.3%	67.6%
XGBoost	**84.1%**	75.9%	**75.6%**	75.7%

Note: Blue-shaded cells denote the best result for each evaluation metric

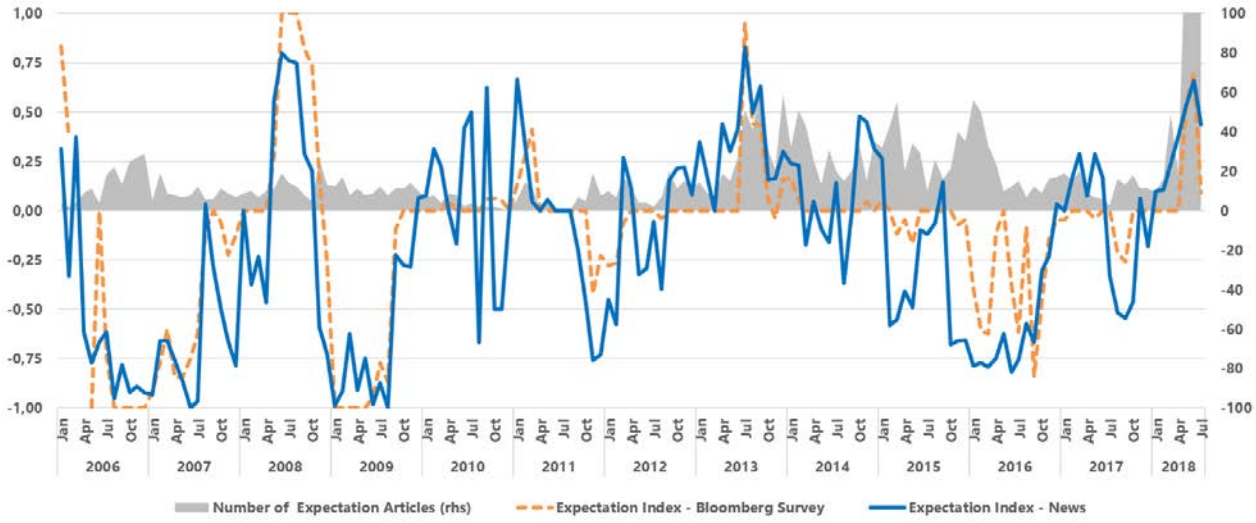
4.2 Result Evaluation

For result evaluation, we calculate the correlation between policy rate expectation index generated from news and from Bloomberg survey. Graphs of both indices from January 2012 to July 2018 are presented in Figure 3. We can see that both indices are moving in the same direction generally, with a correlation of 73% (correlation for full data period, i.e. from January 2006, is 78.6%). The correlation value indicates that the policy rate expectation index from news is potential to be used as a new measure of policy rate expectation.

The policy rate expectation index from news tends to be more volatile, e.g. in the second half of 2010. This is likely due to there are some periods (months) where number of sentences containing policy rate expectation in Cyber Library is very low. From 142 months of data, there are 49 months where the number of sentences containing policy rate expectation are less than 10.

Plot of Policy Rate Expectation Index

Figure 3



For some periods, the expectation index from news can "predict" the direction of policy rate more precisely than the expectation index from Bloomberg survey, as presented in Table 3. Overall, compared to the actual change in policy rate, the expectation index from news has a correlation of 76.6%, while the expectation index from Bloomberg survey has a correlation of 84.5%.

Comparison between Expectation Index from News and from Bloomberg Survey

Table 3

Period	Event	Expectation Index from News	Expectation Index from Bloomberg Survey
December 2007	Policy rate cut	-0.73	-0.13
February 2011	Policy rate hike	0.61	0.27
November 2011	Policy rate cut	-0.80	-0.42
February 2012	Policy rate cut	-0.63	-0.27
September 2013	Policy rate hike	0.84	0.44
November 2013	Policy rate hike	0.62	-0.04
February 2015	Policy rate cut	-0.56	0.00
June 2016	Policy rate cut	-0.78	-0.38
September 2017	Policy rate cut	-0.53	-0.26
May 17 2018	Policy rate hike	0.53	0.55
May 30 2018 (additional)	Policy rate hike	0.67	1.00
June 2018	Policy rate hike	0.66	0.69

5. Conclusion & Future Work

5.1 Conclusion

In this research, we develop a new measure of stakeholders' expectation on Bank Indonesia's policy rate. From methodological perspective, we show how to utilize news articles data to develop the new measure, by employing machine learning-based technique. The expectations are extracted from news, starting from 14 days before the monthly Board of Governor's meeting. The machine learning model is trained by using sentences that have been annotated manually.

From out-of-sample evaluation, we achieve an F1-score of 76.8% on classification accuracy by using logistic regression model. The resulting monthly expectation index has 78.6% correlation with the expectation index generated from Bloomberg's monthly survey.

5.2 Future Work

There are several improvements in the methodology that can be applied for future works.

- Opinion Holder Identification

Currently, the calculation of the expectation index of each month use the average score of the articles. This makes the index is not entirely comparable to expectation measure obtained from Bloomberg survey (news articles vs. survey respondents). We need to identify the opinion holder for each sentence that contains policy rate expectation. Once identified, opinion holders whose expectations are quoted in several articles are counted only once in index calculation.

Another benefit of opinion holder identification is for grouping expectations based on institutional group of the opinion holder, e.g. government, authorities, banking, capital market, industry, academics, and research institutes. Thus, we can further examine which institutional groups expect policy rate hike, cut, or unchanged.

- Data Source Addition

The number of news articles used in this research is not big enough, i.e. 5,700 news articles in 146 months, or about 40 news articles per month. The addition of new data sources can be done with web crawling on online news websites. In addition, we also consider to use news in English language, although additional works are needed to develop a text mining model for English language.

- Classification Model Improvement

Nowadays, artificial neural network (especially deep learning) is state-of-the-art technique for text classification, including opinion extraction task (Irsoy and Cardie, 2014). The currently used classification model, i.e. logistic regression, can be replaced with a neural network model to improve the accuracy. However, it is necessary to annotate more sentences, given the neural network model requires a large amount of training data.

- Expectation vs. Wish vs. Suggestion

Currently, annotated sentences also include phrases of wishes, hopes, and suggestions on the policy rate. We need to separate sentences that contain expectation (or prediction) with sentences that contain wish (or suggestion), so that the index only contains information related to expectations. Rule-based method (using keywords e.g. "expects" vs. "wishes") or machine learning method could be used for the task.

- Expectation Period Identification

Sometimes, sentences that contains policy rate expectations are not referring to the next Board of Governors' meeting, but rather several months or even a year later (e.g. "He predicts BI Rate to be hiked only one more time this year, at the end of 2014."). Such sentences need special handling, i.e. by classifying it as expectation of unchanged policy rate for the next meeting, and as expectation of policy rate hike for meeting at the end of 2014.

References

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics, 131*(4), 1593-1636.

Bernanke, S. B., & Kuttner, K. N. (2004). What Explains the Stock Market's Reaction to Federal Reserve Policy? *The Journal of Finance, 60*(3), 1221-1557.

Bollen, J., Mao, H., & Xiao-Jun, Z. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science, 2*(1), 1-8.

Christensen, J. H., & Kwan, S. (2014). *Assessing Expectations of Monetary Policy*. Retrieved from FRBSF Economic Letter: https://www.frbsf.org/economic-research/publications/economic-letter/2014/september/assessing-expectations-monetary-policy/

de los Rios, A. D., & Reid, C. (2008). Extracting Policy Rate Expectations in Canada. *Capital Markets: Asset Pricing & Valuation eJournal*.

Fischer, S. (2017). *Monetary Policy Expectations and Surprises*. Retrieved from Speeches of Federal Reserve Officials: https://www.federalreserve.gov/newsevents/speech/fischer20170417a.htm

Friedman, B. M. (1979). Interest Rate Expectations Versus Forward Rates: Evidence from an Expectations Survey. *The Journal of Finance, 34*(4), 965-973.

Irsoy, O., & Cardie, C. (2014). Opinion Mining with Deep Recurrent Neural Networks. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (pp. 720-728).

Joyce, M., & Meldrum, A. (2008). Market Expectations of Future Bank Rate. *Bank of England Quarterly Bulletin 2008 Q3*, pp. 274-282.

Joyce, M., Relleen, J., & Sorensen, S. (2008, December). Monetary Policy Expectations from Financial Market Instruments. *ECB Working Paper Series No.978*.

Kuttner, K. N. (2000). Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market. *Journal of Monetary Economics, 47*(3), 523-544.

Lucca, D. O., & Trebbi, F. (2009). Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements. *NBER Working Paper No. 15367*.

Manski, C. F. (2004). Measuring Expectations. *Econometrica, 72*(5), 1329-1376.

Nardelli, S., Tobback, E., & Martens, D. (2017). Between Hawks and Doves: Measuring Central Bank Communication. *ECB Working Paper Series No. 2085*.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, (pp. 122-129).

Rosa, C., & Verga, G. (2007). On the Consistency and Effectiveness of Central Bank Communication: Evidence from the ECB. *European Journal of Political Economy, 23*(1), 146-175.

Sahminan, S. (2008). Effectiveness of Monetary Policy Communication in Indonesia and Thailand. *Bank for International Settlements Working Paper No.262*.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Measuring stakeholders' expectations for the central bank's policy rate[1]

Okiriza Wibisono,

Bank Indonesia

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Measuring Stakeholders' Expectations for the Central Bank's Policy Rate

Okiriza Wibisono, Alvin Andhika Zulen, Anggraini Widjanarti, Hidayah Dhini Ari
Bank Indonesia

Disclaimer

The following opinions are those of the authors and not necessarily those of Bank Indonesia

Expectation on monetary policy

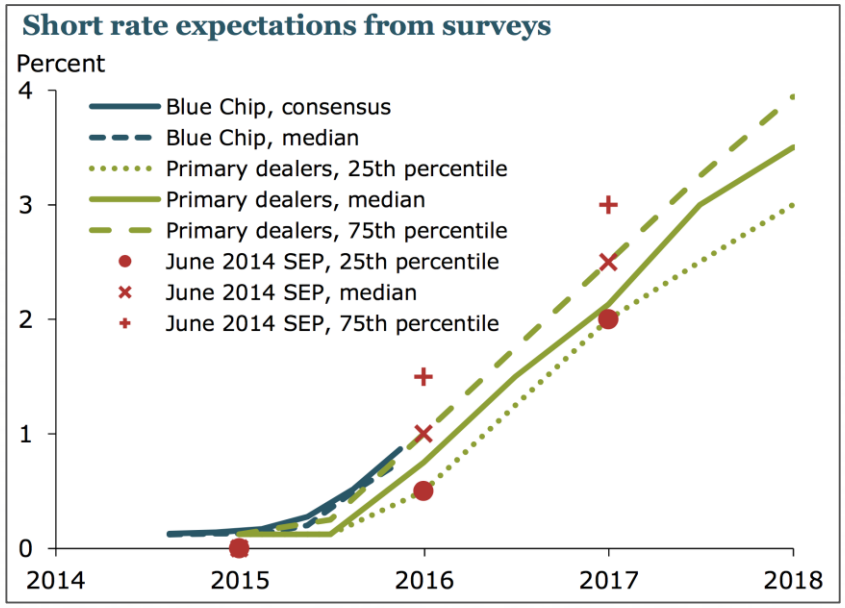
- Trend and expectation on inflation
- Trend of past policy rate
- Trend and expectation on other CB policy rates
- (Expectation on) gov't policies, e.g. gas prices
- Other info and indicators

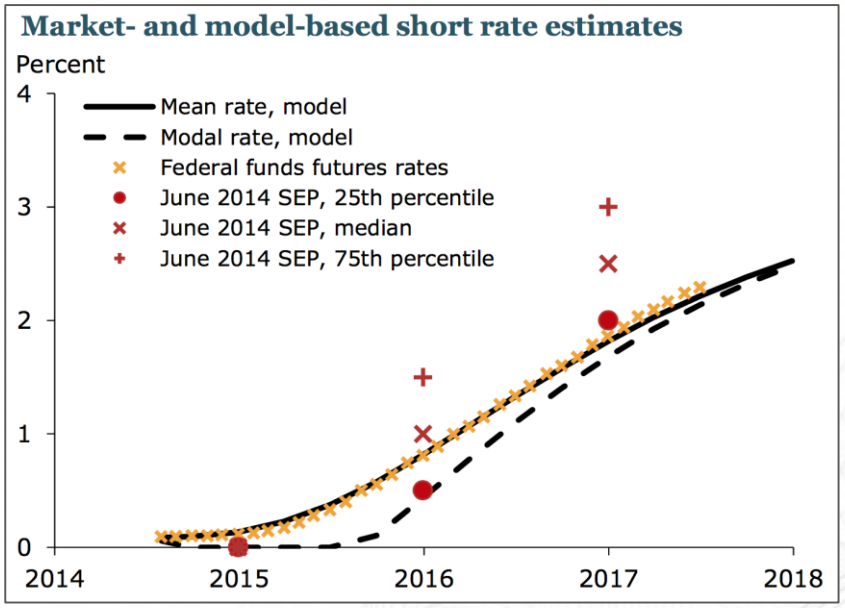
BI policy rate will ...

Financial markets, real sector, analysts, central & local gov't, general public

- Rise? → Behavior A
- Stay? → Behavior B
- Decline? → Behavior C

Measuring expectation on policy rate





Survey-based

Market-based

Charts from **Christensen and Swan (2014)**, *Assessing Expectations of Monetary Policy*

Alternative data source for measuring expectation: newspaper articles

Bank Indonesia is expected to hold its reference rate (BI Rate) at 5,75%. The central bank considers the possibility of subsidized gasoline price hike and its impact on inflation.

As monetary authority, BI is believed to not be careless in setting the interest rate. **Even if there is gasoline price hike this year, BI will probably only raise reference rate by 50 basis points at most.** "Right now, BI is more pro-growth and real sector," said Chief Economist of Danareksa Research Institute Purbaya Yudhi Sadewa to Investor Daily in Jakarta, Tuesday.

A same note is added by Director of Institute for Development of Economics and Finance (Indef) Enny Sri Hartati. Observing the trend of global oil price, Enny is certain that subsidized gasoline prices will not be increased in the near future. **"So, in the next Board of Governors Meeting, Thursday, BI will not change its policy rate.** Apart from its function, BI is also responsible to stabilize business and banking conditions", she remarked.

→ expectation of rate stay

→ expectation of rate hike

→ expectation of rate stay

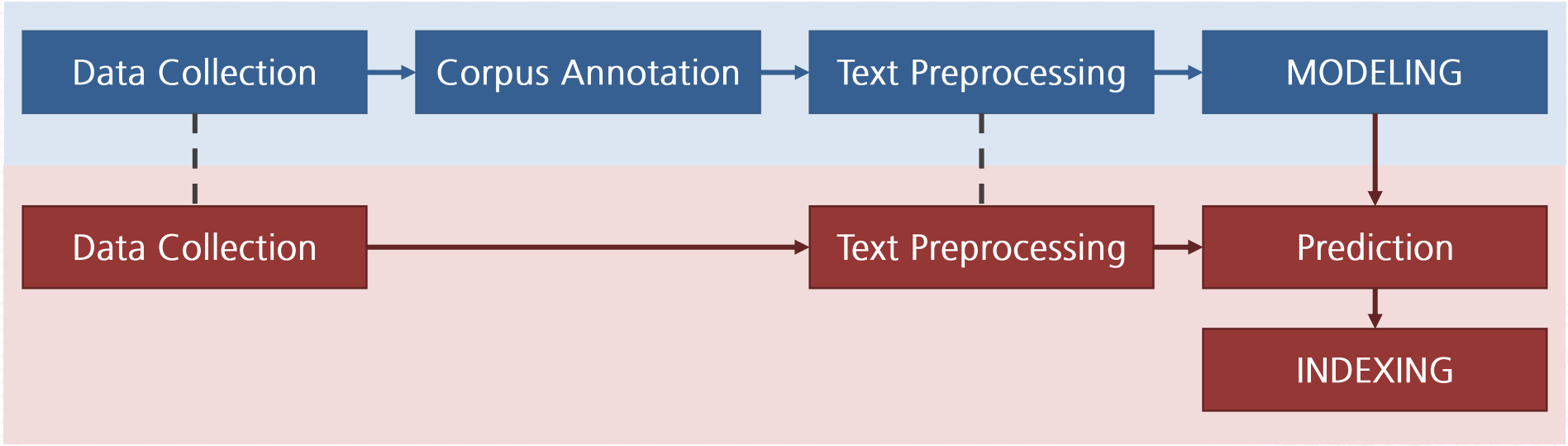
Can we develop a machine learning algorithm for automatically identifying and classifying public expectation on our policy rate from newspaper articles?

Pros-Cons of measuring expectation from news

- ✓ Available for public access
- ✓ Published in real-time
- ✓ Covers wide source of opinion:
  - ❑ Banks, financial institutions
  - ❑ Analysts & economists
  - ❑ Real sector, industries
  - ❑ Academics
  - ❑ Government
- ✓ It's what people read

- ○ May not reflect "true" expectation, no "money-in-the-game"
- ○ Likely less accurate prediction than actual survey on professionals
- ○ No control over "respondents", respondents can change every period

Methodology



MODELING

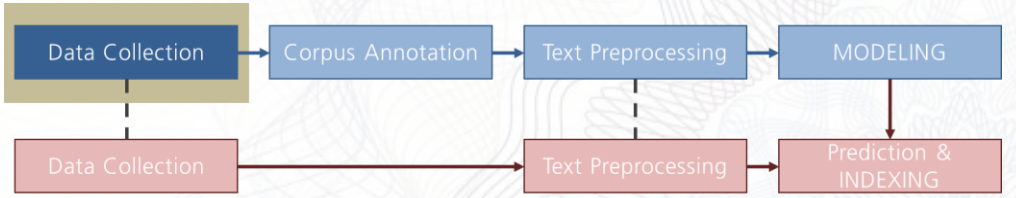
The Modeling step is done **once**
to train the best machine learning algorithm for the task

INDEXING

The Indexing step is done **every time**
Bank Indonesia will hold policy rate meeting (regularly once a month)

Modeling: Data collection

- Dataset: CyberLibrary
  - Collection of economic and financial news articles
  - Curated for internal use by Bank Indonesia
- Collection period: Jan 2006 – Feb 2018
  (BI policy rate was established in July 2005)
- Filtering
  - Kept only articles published **within 2 weeks before** scheduled policy rate meeting
  - Kept only sentences that **mention BI policy rate**
    e.g. "Bank Indonesia reference rate", "BI 7DRR"



Data collection and filtering

All articles in *CyberLibrary*

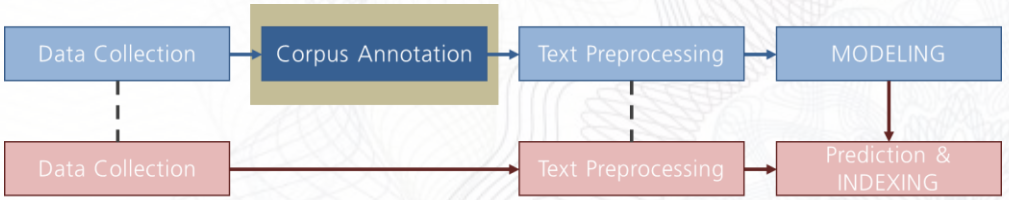
Articles published 14 to 1 day before policy rate meeting

Bank Indonesia is expected to keep **BI Rate** at 5,75% in its policy rate meeting this Thursday.

Sentences mentioning BI policy rate

Modeling: Corpus annotation

- Each filtered sentence is annotated as one of **4 classes**:
  - **0**: Non-expectation
  - **1**: Expectation of rate hike
  - **2**: Expectation of rate cut
  - **3**: Expectation of rate stay
- Each sentence is annotated by 2 persons
  - In case of differing class → 3rd annotator
- Total annotated sentences: **~4,400**
  - 66% non-expectation
  - 8% expectation of rate hike
  - 15% expectation of rate cut
  - 11% expectation of rate stay



Example annotation (in Excel spreadsheet)

date	sent	label	label
2017-08-08	Kalau memang BI ingin melakukan stimulus moneter, hal pertama adalah menurunkan 7-days repo rate dari 4,75% menjadi 4%.	2	turun
2017-08-09	Jika demikian, dia menilai kebijakan pelonggarannya tidak harus melalui pemangkasan suku bunga acuan atau BI 7-Day Repo Rate.	0	nonekspektasi
2017-08-09	Bank Indonesia juga diperkirakan mempertahankan suku bunga 7 days repo rate di level 4.75% sampai akhir 2017.	3	tetap
2017-08-09	BI diperkirakan masih mempertahankan kebijakan suku bunga 7 days repo rate di level 4,75% sampai pertengahan 2018.	3	tetap
2017-08-09	Tingginya minat investor dalam lelang SUN-, menurut Anil, terjadi karena investor mengantisipasi penurunan suku bunga acuan Bank Indonesia (BI).	2	turun
2017-08-11	Di sisi lain, untuk suku bunga acuan Bank Indonesia sejak akhir 2015 sampai saat ini telah menurunkan suku bunga acuan sebesar 150 bps menjadi 4,75%.	0	nonekspektasi
2017-08-11	Wakil Presiden Direktur Bank Central Asia Eugene Keith Galbraith mengatakan, perseroan menilai peluang penurunan bunga kredit masih terjadi karena dari sisi suku bunga acuan Bank Indonesia (BI) masih belum ada perubahan, terutama arah kenaikan.	0	nonekspektasi
2017-08-14	Doddy melanjutkan, perihal Suku Bunga simpanan ini erat kaitannya dengan Suku Bunga acuan bank Indonesia (BI 7 days reverse repo rate).	0	nonekspektasi
2017-08-14	Menurut dia, kebijakan BI yang mempertahankan BI 7 days reverse repo rate di level 4,75% masih sejalan dengan perbaikan kondisi perekonomian global dan proses pemulihan ekonomi domestik yang terus berlanjut.	0	nonekspektasi
2017-08-14	"Sepanjang 2017, BI 7 days reverse repo rate diperkirakan akan tetap flat," ujar dia.	3	tetap
2017-08-14	Chief Economist bank Mandiri Anton Gunawan mengungkapkan, secara industri perbankan, transmisi penurunan Suku Bunga acuan BI pada tahun lalu terhadap Suku Bunga kredit masih belum selesai.	0	nonekspektasi
2017-08-15	Agus di Jakarta, Jumat (4/8), men-gatakan hal tersebut, setelah sembilanbulanberturut-tu-rut BI menahan pelonggaran Suku Bunga acuan "7-Day Reverse Repo Rate" di level 4,75persen.	0	nonekspektasi

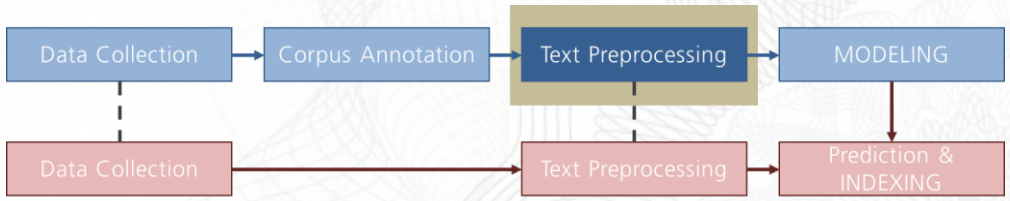
Modeling: Text preprocessing

Each sentence is **preprocessed** to make learning simpler for the algorithm

- Lowercasing
- Stop words removal
- Merging synonyms and word forms e.g.
  - predicting, predicted, expecting, expects → predict
  - BI Rate, BI policy rate, BI reference rate, BI 7DRR → BIRate

Sentence (in text format) is then **transformed** into feature vector (numeric)

- n-gram occurrences (total >5.000 1-4 grams)
- Sentence length
- Any numbers and percentages mentioned
- Sum of word embeddings (from https://fasttext.cc/)



Text preprocessing

Filtered sentences

Annotation

nonexp
cut
hike
nonexp
stay

Text + label for each sentence

Preprocessing

	var-1	var-2	…	var-d	label
sent-1	v_{11}	v_{12}	…	v_{1d}	nonexp
sent-2	v_{21}	v_{22}	…	v_{2d}	cut
sent-3	v_{31}	v_{32}	…	v_{3d}	hike
…	…	…	…	…	nonexp
sent-n	v_{n1}	v_{n2}	…	v_{nd}	stay

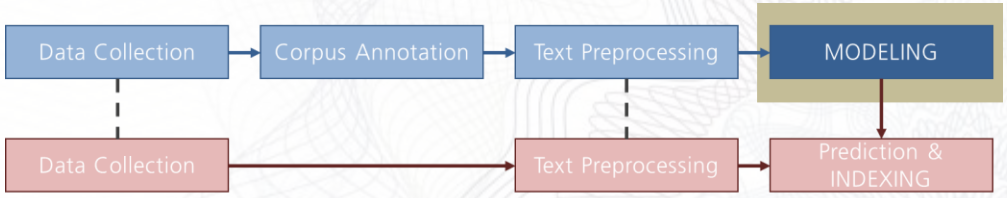
Vector + label for each sentence

Modeling: Training machine learning algorithm

Sentences as feature matrix + annotated labels is ready for machine learning training

Experiment setup:

- 80% sentences for train/validation set, 20% for test set
- Models evaluated:
  - Logistic regression
  - Naïve bayes
  - Decision tree
  - Random forest
  - XGBoost
- 25 hyper-parameter settings for each model
  - 10-fold random split for tuning
  - Each split has 20% test sentences (taken from the 80% train/validation set)



Model evaluation metric

		Predicted by model	
		Expectation (label 1, 2, 3)	Non-expectation (label 0)
Actual label	Expectation (label 1, 2, 3)	TP	FN (*Type II error*)
	Non-expectation (label 0)	FP (*Type I error*)	TN

Metric	Formula	Interpretation
Accuracy	(TP + TN) / (TP + TN + FP + FN)	% correct predictions
Precision	TP / (TP + FP) = 1 – *type I error rate*	% correct labels from all predicted as expectation
Recall	TP / (TP + FN)	% predicted expectations from all labeled as expectation
F1	2 * Precision * Recall / (Precision + Recall)	Harmonic mean of precision & recall

Evaluation results on test data

Each model with its best hyper-parameter is trained on the whole train/validation set, then evaluated on test set to measure performance

Model	Accuracy	Precision	Recall	F1
Logistic regression	83,4	71,2	83,2	76,7
Naïve bayes	80,6	64,5	83,2	72,7
Decision tree	73,0	53,4	65,7	58,9
Random forest	78,0	63,3	72,6	67,6
XGBoost	84,1	75,6	75,9	75,7

Indexing

At this point we have sufficiently accurate machine learning model for classifying expectation for (the next) policy rate meeting, given unlabeled news sentences.

In Indexing step we aim to aggregate the classifications into a single number for each policy rate meeting.

New, unlabeled sentences	Label
The market is anticipating the result of BI Board of Governors Meeting, which is predicted to keep reference rate at 4.25%.	?
"Rupiah stability in 2017 in the midst of FFR hikes, as well as relatively controlled inflation rate, may serve as the basis for BI to cut its policy rate," he said.	?
There is possibility that BI will raise its interest rate on Thursday's Board meeting.	?
…	…

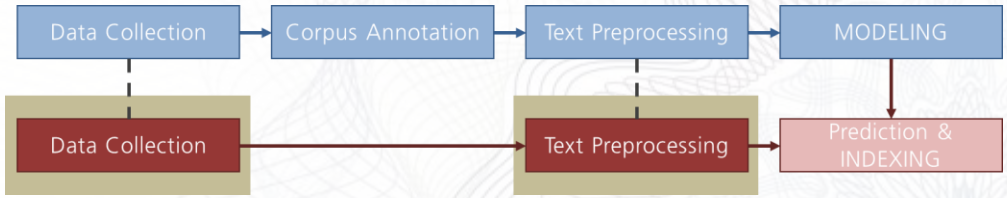
Label, predicted by model
3 (rate stay)
2 (rate cut)
1 (rate hike)
…

Indexing: Data collection and text preprocessing

Data collection and text preprocessing in Indexing is **exactly the same** as in Modeling step,

except the articles considered are grouped into corresponding policy rate meeting

e.g. articles published between 5 to 19 July 2018 will be grouped for policy rate meeting on 19 July



Indexing: Prediction and index calculation

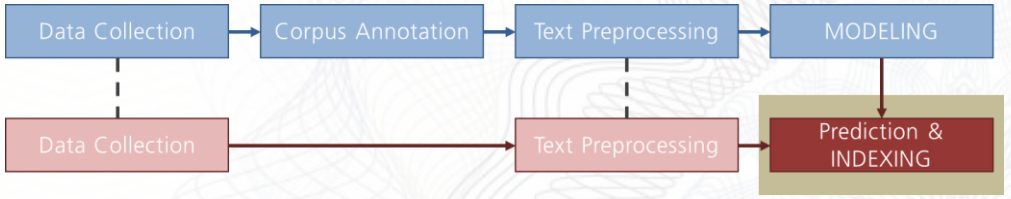
The trained model is used to classify all filtered (and unlabeled) news sentences.

Each sentence is scored based on model's classification:

- 1, if classified as expectation of rate hike
- 0, if classified as expectation of rate stay
- -1, if classified as expectation of rate cut
- Discarded, if classified as non-expectation

Each article is scored as the average of its sentences' score.

Expectation index for the next policy rate meeting is the average of article scores.



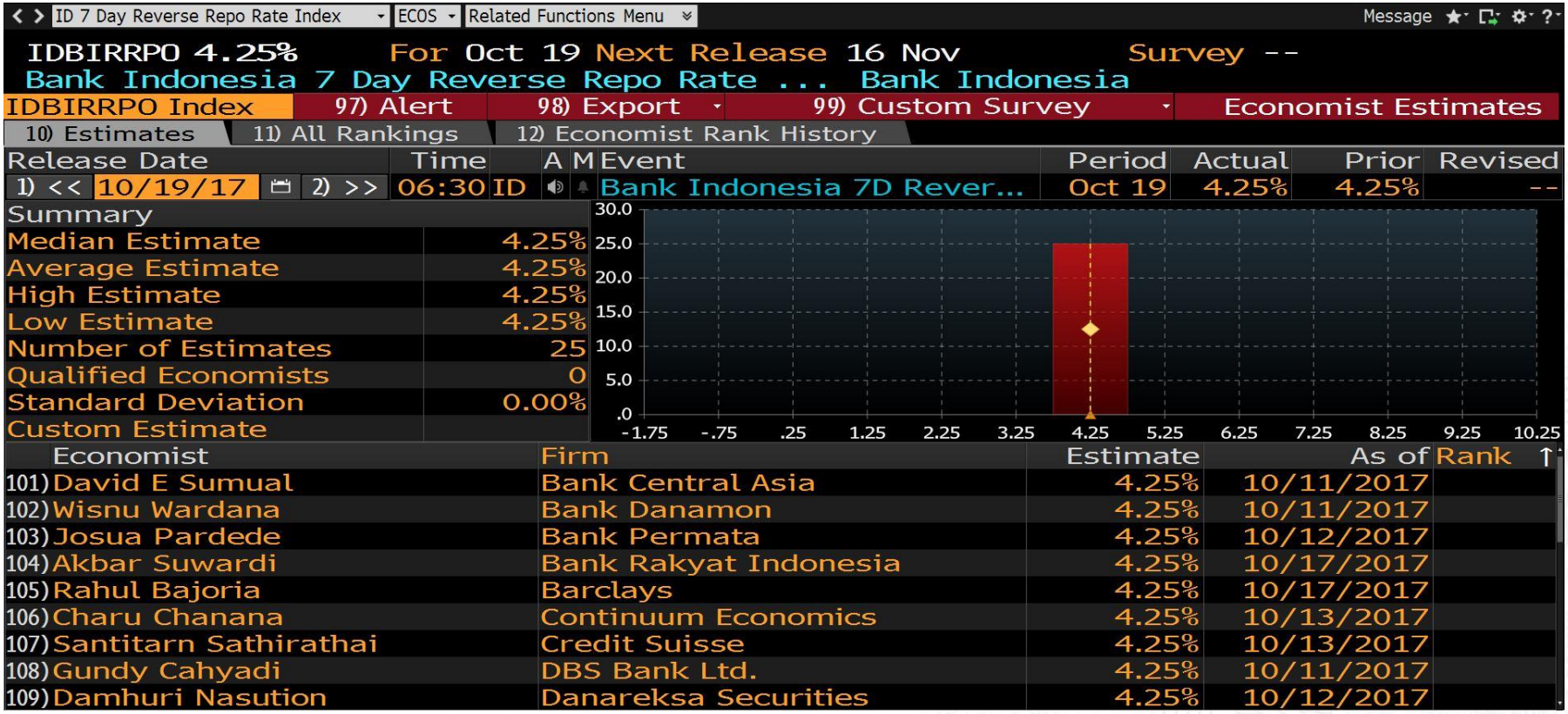
Index properties

The resulting expectation index has following properties:

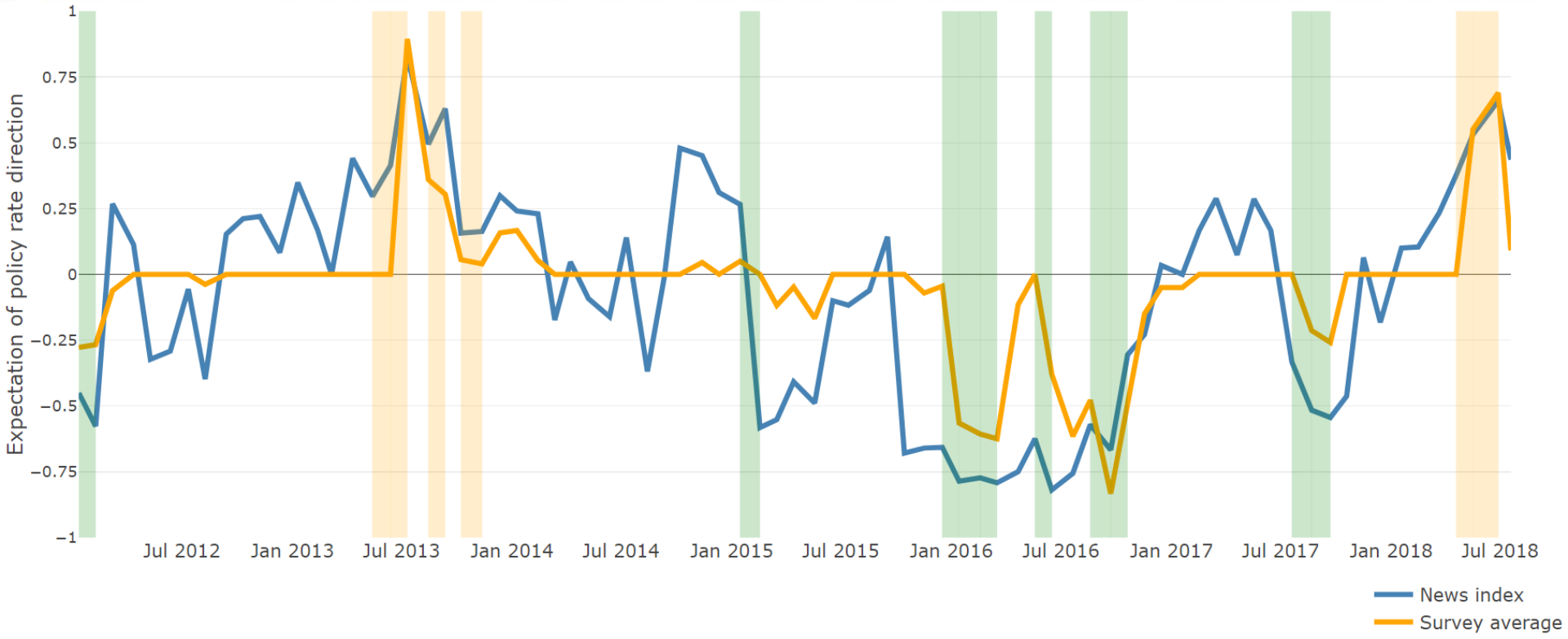
- Range from -1 to 1
- With more expectation of rate hike, index approaches **1**
  With more expectation of rate stay, index approaches **0**
  With more expectation of rate cut, index approaches **-1**
- **Positive index**: more expectation of rate hike compared to rate cut
  Negative index: more expectation of rate cut compared to rate hike
- If index at t_1 > index at t_2, then there is greater share of rate hike expectation in t_1 compared to t_2

Benchmark: Bloomberg Economist Estimates

Bloomberg surveys a number of economists for their prediction of BI policy rate that will be set in the next meeting.



Results: 2012 – July 2018

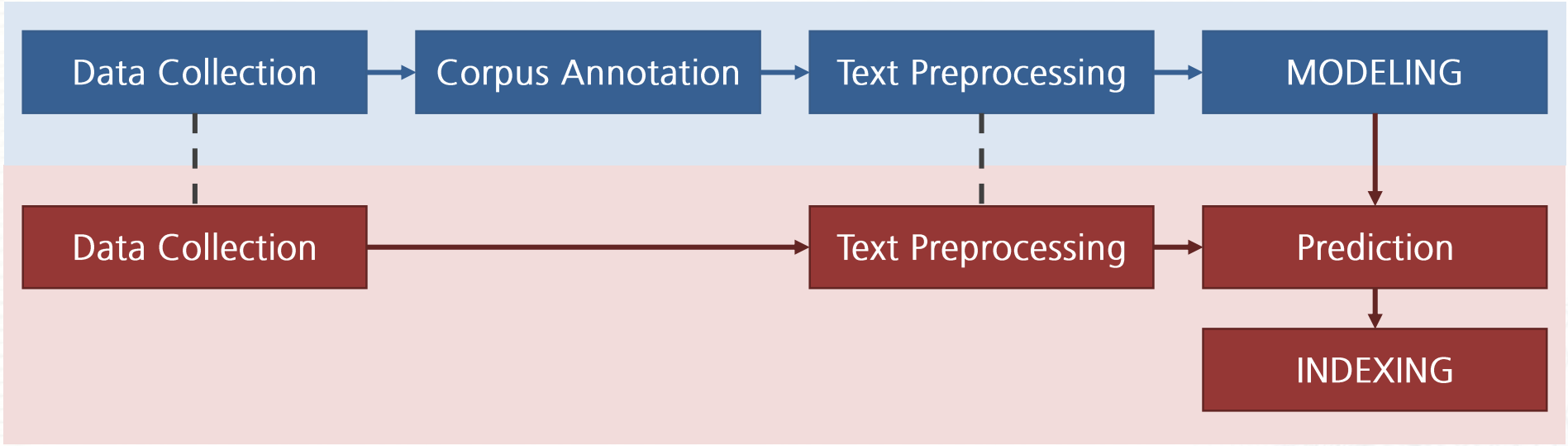


Correlation with Bloomberg survey: 73%

Actual rate hike

Actual rate cut

Methodology (review)



MODELING

The Modeling step is done **once**
to train the best machine learning algorithm for the task

INDEXING

The Indexing step is done **every time**
Bank Indonesia will hold policy rate meeting (regularly once a month)

Conclusion and future improvements

- ✓ Have shown machine learning use case for measuring public expectation on our policy rate, exploiting news articles as data source
- ✓ Overall the machine learning model has quite good accuracy, and also the resulting index tracks professional estimates quite well

<u>Future improvements</u>

- ❑ Identifying **opinion holders**: who has what expectation
- ❑ Separating **predictions vs desire**
  - ❑ "what people *think* the rate will be" vs "what they *want* the rate to be"
- ❑ More news **source**, including English ones
- ❑ More **advanced algorithm** for classification
  - ❑ Deep learning? But probably requires (far) more annotation

THANK YOU
Q&A

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Predictability in sovereign bond returns using technical trading rule: do developed and emerging markets differ?[1]

Tom Fong and Gabriel Wu,

Hong Kong Monetary Authority

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Predictability in sovereign bond returns using technical trading rules: Do developed and emerging markets differ?[1]

Tom Fong and Gabriel Wu

Hong Kong Monetary Authority

Hong Kong, China

Abstract

The study examines the predictability of 48 sovereign bond markets based on a strategy of 27,000 technical trading rules. These rules represent four popular trading rule classes, they are: moving average, filtering, support and resistance, and channel breakout rules, with numerous variants in each class. Empirical results show that (i) investing in sovereign bond markets is predictable, based on the buy-sell signals generated by trading rules, with the predictability of the emerging Asian markets being significantly higher than those of the advanced markets; (ii) the predictability is generally higher when the US tightens its monetary policies or undergoes recession; (iii) two-thirds of sovereign bond markets have a higher predictability when we use a machine learning algorithm to determine the best trading rule strategy; and (iv) the predictability of a sovereign bond market is higher when the economy has a less effective government, lower regulatory quality, narrower financial openness, higher political risk, lower income and faster real money growth. Our results suggest that shocks originating from US monetary policy or economic conditions could have a considerable spillover effect on sovereign bond markets, particularly the emerging Asian markets.

Keywords: trading rule, return predictability, monetary cycle, machine learning, business cycle, spillover, market efficiency

JEL classification: C58, D83, E32, E50, G12, G14, G17

[1] Email addresses: Fong: tpwfong@hkma.gov.hk. and Wu: gstwu@hkma.gov.hk

The authors would like to thank Lillian Cheung, Cho-hoi Hui, Giorgio Valente, participants at the Bank Indonesia – IFC workshop on "Big Data for Central Bank Policies" and an anonymous referee of the HKIMR working paper series for their valuable comments.

I. Introduction

There are extensive studies on the existence of memory in financial time series of equity and foreign exchange markets worldwide, highlighting the importance of monitoring predictability of these markets. However, only a few studies discuss sovereign bond market predictability. In fact, a predictable sovereign bond market can possibly result from an inefficient price discovery process in sovereign bonds, a substantial risk premium priced in the market prices, or a combination of both.[2] The resulting impact may have important implications for government and corporate borrowing costs and access to financing and, therefore, can weigh on economy-wide financial conditions. Therefore, predictability of sovereign bond markets merits closer scrutiny.

This paper analyses the predictability of numerous sovereign bond markets based on technical trading rule analysis. While providing an overview of market predictability, we especially assess the extent that predictability is affected by US monetary and business cycles, given that global markets are managing the transition towards US monetary policy normalisation. We examine these issues in three steps. First, we apply numerous trading rules to sovereign bond markets to assess predictability using the trading-rule strategy and predictability during different US monetary and business cycles. Second, we apply a machine learning algorithm to the trading rule strategy to determine ways to improve the return predictability. Finally, a regression analysis is conducted to uncover the relationship between return predictability and various social and economic factors. The results can shed light on several issues that are not well discussed in literature: (i) Are sovereign bond markets predictable? (ii) If yes, which markets are more predictable? Is the predictability higher during tightening of US monetary policy and US recessions? (iii) If no, can machine learning techniques increase predictability? (iv) What social and economic factors could explain predictability in sovereign bond markets?

There are four major findings in this study. First, investing in sovereign bond markets is predictable based on the buy-sell signals generated by trading rules. In particular, the predictability of emerging Asian markets is significantly higher than those of advanced markets. Second, the predictability is generally higher when the US tightens its policy or undergoes economic recession. Third, two-thirds of the sovereign bond markets have a higher predictability when we use a machine learning algorithm to determine the best trading rule strategy. Finally, the predictability of a sovereign bond market is higher when the economy has a less

[2] The profitability of technical analysis indicates market inefficiency under a strict interpretation of weak-form efficiency, which rules out return predictability based on historical information. However, such predictability may be a reflection of time-varying bond risk premia, which violates the expectation hypothesis that assumes a constant bond risk premium. In equity and currency markets, some studies find that the risk premia may not be strongly associated with returns from trading rule strategy (see Park and Irwin, 2007, and Ivanova et al., 2016).

effective government, lower regulatory quality, narrower financial openness, higher political risk, lower income and faster real money growth.

Taken together, our results contribute to the research field in two respects. First, while trading rule analysis is commonly applied to evaluate predictability of equity and exchange rate markets, this paper is one of a few to offer a comprehensive study of predictability of sovereign bond markets worldwide and of their potential responses to spillovers of US economic conditions and monetary policies. Second, to our best knowledge, this is the first paper using machine learning techniques together with popular trading rules to evaluate the predictability of sovereign bond markets.

The remainder of the paper is organised as follows. The next section reviews several major studies on technical trading rules. Section 3 discusses the methodologies, which include applications of technical trading rules, machine learning analysis and identification of determinants. Section 4 describes the data on sovereign bond returns and potential determinants of the predictability. Section 5 presents the empirical results. The last section concludes our findings and discusses its implications.

II. Major studies on sovereign bond market predictability and technical trading rules

Studies on sovereign bond market predictability and its potential determinants have grown quickly in literature of sovereign risk surveillance.[3] Focusing on the US Treasury market, Shynkevich (2016) investigates the predictability of bond returns across different market segments and varying market conditions and finds that the predictability is inversely related to interest rate risk, but positively related to default risk. Fakhry and Richter (2015) and Fakhry et al. (2017) find evidence of inefficiency in the sovereign bond markets of the US, Germany, Greece, Italy, Portugal and Spain before and during crisis, which arises from the fact that these markets are too volatile as measured by a new volatility test.[4] Zunino et al. (2012) ranked sovereign bond market efficiency for 30 sovereign bond markets estimated by the complexity-

[3] Early studies include Hall and Miles (1992), who find evidence of predictability in excess returns in the sovereign bond markets of US, Canada, UK, France, Germany and Japan. They found the slope of the yield curve helped predict subsequent bond excess return for US and Canada, while there was evidence of positive serial correlation in excess returns for four other markets.

[4] The basic argument of the test is that, in an ideal world, future cash flows should determine the behaviour of prices today; therefore, as Shiller (1992) argues, any excess volatility is evidence of inefficient markets.

entropy causality plane,[5] and found that the stage of economic development and market size of the sovereign bond market could affect efficiency levels. Charfeddine et al. (2018) study the time variation in sovereign bond market efficiency for US, UK, India and South Africa, which was found to depend on prevailing economic, political and market conditions.

Another strand of study on possible explanations for association between the predictability in international bond market returns and the monetary policies and/or economic conditions emerges as a major issue amid concerns over monetary policy normalisations in major economies. From the perspective of monetary policies, one possible explanation is that interventions by US monetary authorities may create predictable moves in international currency markets that technical traders would be able to exploit. They would subsequently create predictable changes in prices of interest rate-sensitive assets, such as sovereign bonds (Shynkevich, 2016). The associated uncertainties on monetary conditions could also contribute to varying bond risk premia, which subsequently increases return predictability (Ireland, 2015, and Jiang and Tong, 2017). Another possible explanation is that international bond markets are increasingly integrated over time, which facilitate a stronger spillover effect of the US bond risk premia on international bond markets (Dahlquist and Hasseltoft, 2013).

From the perspective of economic conditions, the predictability can be explained by the link performance of bonds and bond portfolio to business cycles, for instance the cyclical variation in bond risk premia (Ludvigson and Ng, 2009). Gargano et al. (2017) find that countercyclical bond risk premia, as driven by heightened uncertainty, contribute to higher predictability of bond market returns during recessions. They also find that disagreement spikes in bad times generate the time series momentum, leading to predictability in returns. These findings are consistent with Cujean and Hasler (2017), who attribute the higher stock market predictability during recession to the situation when economic conditions deteriorate, uncertainty rises and investors' opinions polarise, based on an equilibrium model.[6]

To evaluate market predictability, technical trading rules are regarded as one of the simplest techniques, given its objectiveness and readiness in computation. It is primarily based on the premise that past price trends predict future price movements without rigorous economic or financial theories.[7] Many financial practitioners view technical analysis as an important forecasting tool in making

[5] It measures the presence of temporal patterns in deviations from the ideal position associated to a totally random process. The distance to this random ideal location can be used to define a ranking of efficiency.

[6] For empirical studies on stock returns, Rapach et al. (2010), Henkel et al. (2011) and Dangl and Halling (2012) also found that predictability was concentrated in economic recessions and was largely absent during expansions.

[7] Depending on the class of trading rules used; some trading rules attempt to look for imminent market correction after a rapid movement in certain direction, similar to the oscillator trading rules; while some rules are trend-chasing, expecting that the current trend will continue, such as support and resistance rules.

short-term trading decisions (Menkhoff, 2010),[8] while academics mostly focus on technical trading rules[9] in investigating forecasting power of technical analysis. Sweeney (1986) and Brock et al. (1992) are pioneers in this area who find technical trading rules can generate excess returns in foreign exchange and equity markets respectively. Tian et al. (2002) expanded the set of trading rules used by Brock et al. (1992) and examined the predictability of stock price movements in markets with different efficiency level, in particular US and Chinese equity markets[10]. Hsu et al. (2016) studied the 41 currencies and the time series and cross-sectional variation in return predictability across sub-periods and geographic group. They found that emerging market currencies were more predictable with technical analysis than developed country currencies.

In literature, there are four types of theoretical models on why technical indicators could have predictive ability (Neely et al., 2014). These models include: (i) differences in the time for investors to receive information; (ii) different responses to information by heterogeneous investors; (iii) under-reaction and over-reaction to information; and (iv) effects of investors' sentiment. Among these models, the second one is useful to explain why technical indicators display enhanced predictive ability during recessions, during which the different responses are led by consumption smoothing asset sales by households that experience job losses and liquidation sales of margined assets by some investors.

III. Methodology

The section details several components that support our analysis. We first discuss several popular trading rules that are commonly used to evaluate predictability of equity and exchange rate markets. We also discuss two empirical advanced methods that are commonly employed in big-data analysis: the bootstrapping simulation and the Naive Bayes Classifier (NBC). The former tests significance of excess returns of trading rule strategy using a simulated distribution of the excess returns, while the latter determines ways to improve the predictability by learning from the historical performance of different trading rules.

[8] The study surveyed 692 fund managers in five countries, including those in the US. 87% of the respondents put at least some importance in technical analysis when making trading decisions.

[9] In a comprehensive survey by Park and Irwin (2007), 58 out of 76 modern technical analysis studies surveyed applied technical trading rules.

[10] They found that while there was no evidence to support the forecasting power of technical trading rules on US equity market after 1975, they could generate excess return in the Chinese stock markets.

3.1 Popular trading rules

In this assessment, we explore four popular classes of technical trading rules: moving average (MA), filtering (FL), support and resistance (SR), and channel breakout (CB) rules. These rules are "return-chasing" in nature and have proved useful in the literature to predict returns in equity and foreign exchange markets. Their usefulness can be explained by the existence of positive feedback traders, who buy (sell) after asset prices rise (fall), as a result of overreaction to information (Hong and Stein, 1999).

According to the MA rule, buy and sell signals are generated by two moving averages of the level of the index: a long-period average and a short-period average. Figure 1 provides a graphical illustration of how the rule generates trading signals. In its simplest form, this strategy is expressed as buying (or selling) when the short-period moving average rises above (or falls below) the long-period moving average. The rationale is that, when the short-period moving average penetrates the long-period moving average, a trend is considered initiated, so prices become predictable.

The other three trading rules could generate trading signals in a similar logic, which are illustrated in Figures 2-4. Specifically, FL rules attempt to follow trends by buying (selling) an asset whenever its price has risen (fallen) by a given percentage; SR trading rules are based on the premise that a breach of a support or resistance level (lower and upper bounds through which the price appears to have difficulty in penetrating) will trigger further rapid price movement in the same direction; and CB trading rules seek to identify time-varying support and resistance levels, or a "channel of fluctuation", on the presumption that, once breached, further rapid price movement in the same direction will ensue.

By considering several variants of each trading rule and a range of different plausible parameterisations of each variant (e.g. Sullivan et al., 1999, and White, 2000), we obtain a large number of possible trading rules. Intuitively, choosing few rules may cause bias in statistical inference due to data mining. However, choosing too many rules may reduce the power of the test due to the inclusion of many underperforming rules. We therefore find a balance and select a fairly large variety of reasonable parameters that lie in the ranges considered by Shynkevich (2016), who applies the same universe of 27,000 technical trading rules to study predictability of US Treasury markets. The detailed logic for each class of trading rule, including the specifications, is provided in the Appendix. As can be seen, the variation on holding a position of a fixed minimum of days in all four classes allows the possibility of a neutral position[11]; whereas, in the basic form, the trader would keep an open buy (sell) position until the opposite trading signal emerged. Meanwhile, the two filters of fixed percentage band and time delay applied to MA and SR rules are meant to mitigate the influence of volatility and present stronger

[11] The neutral position is triggered when the buy or sell signal is not triggered at the end of the fixed holding period.

evidence that a new trend has formed. Only one of the two filters is applied in a certain specification of trading rule.

3.2 Performance measure

For evaluation of the predictability of each sovereign bond market, we use the excess return from trading rule strategy over the buy-and-hold return, or, in short, excess return, in this study.[12] A market is considered predictable when the trading rule strategy outperforms the buy-and-hold one (i.e., the excess return of the market is greater than zero).

Apart from this setting, we impose a "double-or-out" trading strategy in calculating the excess return, as in Brock et al. (1992), Bessembinder and Chan (1998), and Shynkevich (2016).[13] Specifically, we suggest that the investor has a long position at each single trading day by default. On a certain day, if a buy signal emerges from a trading rule, the long position of the investment will be doubled at a borrowing cost for that day. In contrast, if the rule emits a sell signal, the default long position will be liquidated and the proceedings will be invested at a risk-free rate. No action will be taken if there is no signal from the trading rule. The investment will return to the default long position the next day, where the above process will be repeated.

To be specific to sovereign bond markets, the measure is slightly modified by introducing a one-day delay between the generation of trading signals (i.e., at time t) and the time when the respective trading position is taken (at time t+1) in the calculation of the excess return. The rationale behind this modification is that bonds are not as heavily traded as many of the equities or currencies, so the predictability in returns on bond portfolios can have a spurious nature due to nonsynchronous trading of the bonds.[14] Subsequently, the presence of synchronous bias inflates autocorrelations in the return series and overestimates the true predictability in returns and exaggerates the profitability of trend-chasing strategies designed to exploit the time series momentum.

[12] Another common benchmark employed in literature is the risk-free return through the "long-or-short" strategy (Sullivan et al., 1999). We do not consider this benchmark in this study as it requires taking a short position, which could be costly in the case of bond trading.

[13] The "double-or-out" strategy is a symmetric strategy where a trader will increase (decrease) the default long position by the same percentage (specifically 100%) upon a buy (sell) signal. Alternative to this strategy would be an asymmetric strategy where different reactions to buy and sell signals are assumed. However, as Bessembinder and Chan (1998) suggested, in the absence of compelling reasons, searching through the different combination of such asymmetric strategy could potentially increase the problem of data snooping bias.

[14] More specifically, the nonsynchronicity arises from the fact that components of the underlying indexes can cause spurious serial correlation in quoted index values.

Taking into account all the considerations, the net form of the excess return given a trading signal at day t over the buy-and-hold strategy, denoted by ER_t, can be expressed as: [15]

$$ER_t = [(lnS_{t+2} - lnS_{t+1}) - i_{t+1}] * \mathrm{I}_t, \quad (1)$$

where $\mathrm{I}_t = \begin{cases} 1 \text{ if buy} \\ 0 \text{ if neutral} \\ -1 \text{ if sell} \end{cases}$ and $i_t = \begin{cases} rk_t & \text{if buy} \\ 0 & \text{if neutral} \\ rf_t & \text{if sell} \end{cases}$

and S_t is the closing price of the bond index at time *t*, rk_t is the risky rate at time *t*, and rf_t is the risk-free rate at time *t*.[16]

In the empirical results, we express this excess return in a Sharpe ratio (i.e., normalising the excess returns by its standard deviation and presenting it in annualised form), so as to facilitate comparison across sovereign bond markets given that all the excess returns are expressed in standardised form.

3.3 Test for significance of trading rule returns with bootstrapping

The test aims to check whether the trading rule strategy performs no better than the benchmark buy-and-hold strategy. Specifically, the testing procedure is based on the following test statistic:

$$\overline{V_l} = \frac{1}{K}\sum_{k=1}^{K}\left(\sqrt{N} * \overline{ER_k}\right) / \sigma_k \quad (2)$$

where $\overline{ER_k} = \sum_{t=201}^{T} ER_{k,t} / N$ is the average excess return for the k-th trading rule out of *K* trading rules and *N=T-200*[17] is the sample size, and σ_k is a consistent estimator[18] for the standard deviation of $\sqrt{N} * \overline{ER_k}$.

[15] The excess return is derived as follows. Consider an investor with capital $A. In the case of a buy signal, at time t the one-day benchmark return (in amount) is $A * (lnS_{t+1} - lnS_t)$. When the buy signal emerges, investors would borrow another $A at a risky rate at time t+1 (due to the one-day delay imposed), which would earn him a total of $A * [(lnS_{t+1} - lnS_t)] + A * [(lnS_{t+2} - lnS_{t+1}) - rk_{t+1}]$. The excess return, w.r.t. initial capital $A, is then $[(lnS_{t+2} - lnS_{t+1}) - rk_{t+1}]$. In the case of a sell signal, the investor would sell at time t+1 and reinvest at a risk-free rate, which would earn him $A * rf_{t+1}$. However, at the same time, the investor would forgo $A * [(lnS_{t+2} - lnS_{t+1})]$, which would be earned if he maintained the asset at time t+1. The excess return in this case would equal $-[(lnS_{t+2} - lnS_{t+1}) - rf_{t+1}]$.

[16] As illustrated, a risky (borrowing) and risk-free (lending) rate are required for the calculation of excess return. Following Shynkevich (2016), we set the yield on a 3-month US Treasury bill as the lending rate and the 3-month US dollar LIBOR as the borrowing rate. Historical data on both interest rates are retrieved from the Federal Reserve Bank of St. Louis.

[17] We follow Shynkevich (2016) and standardize all trading rules to start generating signals only from the 201th observation because some rules require 200 days of previous data to provide a trading signal. Meanwhile, T would vary depending on sample size of individual sovereign bond index.

Since the distribution of $\overline{V_l}$ is not known, we employ the stationary bootstrap method of Politis and Romano (1994) to simulate the empirical distribution.[19] First, for each trading rule *k*, we resample the realised excess return series $ER_{k,t}$, one block of observations at a time with replacement, and denote the resulting series by $ER^*_{k,i}$. This process is repeated *B* times and for each replication *i*, we compute the sample average of the bootstrapped returns denoted by $\overline{ER^*_{k,i}}$. Finally, we construct the following bootstrap test statistics to form the distribution for $\overline{V_l}$;

$$\overline{V_{l,i}} = \frac{1}{K}\sum_{k=1}^{K}\left(\sqrt{N} * (\overline{ER^*_{k,i}} - \overline{ER_k} * I_{((\sqrt{N}*\overline{ER_k})/\,\sigma_k > -A))}\right) / \sigma_k, \qquad (3)$$

where i = 1,2,….B and I is an indicator function which equals one when the condition is satisfied and zero otherwise and A = $\sqrt{2\ln\ln N}$. The test's p-value is subsequently obtained by comparing $\overline{V_l}$ with the quantiles of $\overline{V_{l,i}}$.

This testing procedure follows the spirit of the superior predictive ability (SPA) test introduced by Hansen (2005) to address potential simulation bias, except that the SPA test compares the maximum return while our method compares the average return. Such difference is considered because we primarily want to assess the overall performance of the trading rule strategy, rather than to identify whether a few trading rules outperform. In an extreme case, if there is only one trading rule that extremely outperforms, but all the remaining rules suffer a loss, the strategy will likely be rejected, given that the average value is biased downward in magnitude (i.e., given that it takes into account those poorly performing rules as well).

3.4 Supervised machine learning algorithm using NBC

We use a machine learning technique to evaluate whether or not the returns from the trading rule strategy can be optimised through learning from the past performance of trading rules. The predictability of a sovereign bond market is higher if our machine learning algorithm can increase the returns from investing in the market with the trading rule strategy.

[18] The estimate σ_k is computed using the stationary bootstrap procedure;

$$\sigma_k^2 = \widehat{\gamma_{0,k}} + 2 * \sum_{i=1}^{N-1} k(N,i)\,\widehat{\gamma_{i,k}}, * \sum_{i=1}^{N-1} k(N,i)\,\widehat{\gamma_{i,k}},$$

where $\widehat{\gamma_{i,k}} = \sum_{i=1}^{N-1}(ER_{k,t} - \overline{ER_k})(ER_{k,t+i} - \overline{ER_k})/N$, i = 0,1…N-1, are the empirical covariances and kernel weights are given by

$$k(N,i) = \frac{N-i}{N}(1-q)^i + \frac{i}{N}(1-q)^{N-i},$$

with q being the smoothing parameter. We follow Shynkevich (2016) and set q = 0.1

[19] Politis and Romano's method resamples blocks of varying length of the original trading rule return series $ER_{k,t}$ to form a simulated return series. The block length follows a geometric distribution with expected block length equal to the inverse of a smoothing parameter.

The algorithm involves three stages. The first two stages use the data from 2000 to 2016 for in-sample estimations gauged by the NBC and model calibrations by adjusting to different market conditions respectively, while the last stage uses the 2017 data for an out-of-sample prediction. The framework is outlined in Figure 5.[20]

In the training stage, the algorithm learns the pattern of historical performances of trading rules under different market conditions using NBC. Three sample periods, including: (i) from 2000 to 2007; (ii) from 2008 to 2013; and (iii) from 2014 to 2015, are considered as reflections of tranquil, stressful and post-crisis market conditions respectively. For each of these market conditions, the algorithm is able to make a prediction for the most likely outcome (positive or negative excess return) of the rules, namely the maximum a posteriori (MAP) estimate. When new information is given, these MAP estimates are then used to formulate a strategy that is built by the portfolio of 27,000 trading rules. A higher weight is assigned to a rule that is predicted to attain a positive excess return, but zero weight to a rule that is predicted to attain a negative excess return (i.e., such rules are excluded from the strategy). In the validation stage, the algorithm determines the best strategy to maximise the excess return based on the 2016 data. In the testing stage, the algorithm uses this best strategy to predict the potential excess returns in the out-of-sample period. If the excess return of the strategy suggested by our algorithm is higher than a benchmark excess return from using all 27,000 trading rules with equal weights (i.e., without weights adjusted by our machine learning algorithm), then the algorithm is considered useful.

IV. Data

4.1 Sovereign bond market indices

This study employs 48 sovereign bond indices covering AEs and EMEs compiled by Bank of America (BofA) and Merrill Lynch (see Table 1 and Figure 6). The indices' constituents are fixed rate nominal sovereign debt with maturity over one year, weighted by market capitalisation. The indices are calculated in the form of total return price series, including those of capital gain, accrued interest and cash flow received during the month. The original data is denominated in local currency, but we convert them into US dollars so as to facilitate cross-country comparison.[21] The bond indices obtained from Bloomberg are in daily frequency with the sample period spanning from 3 January 2000, to 30 September 2017. Given this period, most of the countries (29 out of 48) have complete data for the whole sample period.

[20] The framework is primarily based on Hastie et al. (2009).

[21] Another rationale for this choice is that we can assume all trading rules are measured from a US investor's point of view.

Table 2 shows the average daily return of each sovereign bond index, its standard deviation (SD), the Sharpe ratio (i.e., the mean-SD ratio) and the sample period. Averages are reported for groups of AEs and EMEs classified according to the MSCI classification of developed and emerging markets.[22] As can be seen, there are notable differences in characteristics of sovereign bond markets among different economy groups. For example, emerging Asian sovereign bond markets have the largest daily returns on average (i.e., 5.9%) with the smallest SD (i.e., 6.9%), while other EMEs have the smallest return on average (i.e., 4.5%) with the largest SD (i.e., 14.7%). After adjusting for the risk, emerging Asian markets are found to have a higher return than other markets, with the Sharpe ratio of emerging Asian markets being the highest (i.e., 0.99 in standard score), followed by AEs (i.e., 0.56) and other EMEs (i.e., 0.34).

4.2 Relevant market characteristics to the predictability

We consider a wide range of market characteristics that are considered relevant to the predictability of sovereign bond markets. Puy (2016) considers that governance and accountability, political instability, strength of money and economic risk are potentially important for fund flows in bond markets, given that these variables can identify countries which are more sensitive to global contagion. Zunino et al. (2012) and Charfeddine et al. (2018) also indicated that the prevailing economic, political and market conditions could strongly affect the degree of return predictability of sovereign bond markets.

Table 3 describes these variables. Specifically, the variables of "government effectiveness" and "regulatory quality" measured by the World Bank reflect perceptions of public services quality and governments' ability to formulate sound policies to promote private sector development respectively. Political instability is proxied by the index of political stability and absence of violence and terrorism constructed by the World Bank, which measures perceptions of the likelihood of political instability and politically motivated violence. The financial openness is measured by Chinn and Ito (2006), which codifies the tabulation of restrictions on cross-border financial transactions reported by the IMF. The strength of money is measured by real money growth. The market depth of public bonds is measured by the size of public debt as a percentage of GDP.[23] Finally, we consider the standard deviations of GDP growth and inflation between 2000 and 2016 and the GDP per capita at Purchasing Power Parity as important economic conditions of the sovereign bond markets.

[22] Details of these groupings can be seen on the website https://www.msci.com/market-classification.

[23] Public debt refers to the cumulative total of all government borrowings less repayments that are denominated in a country's home currency. Details of the definition can be seen on the website of CIA World Factbook at https://www.cia.gov/library/publications/the-world-factbook/rankorder/2186rank. html.

Table 4 presents averages of these variables by economic group. As can be seen, there are noticeable differences in fundamental structures of economy groups that may give rise to difference in the predictability of their sovereign bond markets. For example, AEs are characterised with a deeper market for public bonds (i.e., 77.7%), a more effective government (i.e. 1.58) and a higher degree of financial openness (i.e. 0.95)[24]; emerging Asia displays stronger growth of money in real terms (i.e., 7.5%) among EMEs, while other EMEs show a stronger volatility in inflation (i.e., 2.7%) and output (i.e., 3.2%).

Tables 5 and 6 present the correlation matrix of the variables and the results of principal component analysis in descending order of proportion of total variation explained endogenously. Some highly correlated variables (such as government effectiveness, regulatory quality, political risk, real GDP per capita and financial openness) are grouped together in the first principal component, which explains 57.4% of total variation. This component can be regarded as the stage of social and economic development of a sovereign bond market given that the factor loading of these variables is notably larger than other variables with a similar magnitude (ranging from 0.36 to 0.42). The second principal component (explaining 14.3%) is economic uncertainty, given that it is composed of volatilities of output growth (0.75) and inflation (0.50). The third component (explaining 12.6%) represents the market depth of public bonds provided that the component weighs largely on the size of government debt (-0.80). The fourth component is relevant to the strength of money since it weighs heavily on real money growth (0.61). The remaining factors are considered to be unclassified since there are more contrasts between variables, which make interpretations of principal components less straight forward. That said, these components explain only 9.7% in total variations among all variables.

V. Empirical results

5.1 Are sovereign bond markets predictable?

We assess the predictability from three perspectives in this section. We first provide an overview of potential excess returns acquired from investing in each of the 48 sovereign bond indices using the 27,000 trading rules. Robustness of the predictability is tested using our bootstrapping method. Next, we examine whether the predictability of international sovereign bond markets differs during US monetary policies or business cycles. Phases of US monetary cycle (easing and tightening) are determined by whether the US Federal Fund target rate is on an increasing or decreasing trend (Figure 7), while the US business cycles (expansion and recession) are determined based on the turning points as identified by OECD

[24] Government effectiveness spans from a scale of -2.5 (least effective) to 2.5 (most effective), while financial openness measure spans from a scale of 0 (least open) to 1 (most open).

based on US real GDP (Figure 8).[25] Finally, we explore the room for improved predictability of trading rule strategy using a machine learning algorithm, attempting to provide evidence on the robustness of the trading rule strategy's predictability.

5.1.1 Which markets are more predictable?

Table 7 summarises the daily average returns from trading-rule-based investment in sovereign bond markets. As mentioned in section III, all the returns are risk adjusted, annualised and scaled up by the average annualised SD of the daily returns from applying trading rules to the 48 sovereign markets during the sample period from 2000 Q1 to 2017 Q3.[26] Markets are ranked in order of the size of the returns (column 2).

As shown in the column, most of these returns are positive, meaning that these sovereign bond markets are mostly profitable from using trading rules. The overall average is 1.1%, with the most profitable market being China (5.2%), followed by most of the EMEs, such as Philippines (4.7%), Peru (4.5%) and Greece (3.5%). In comparison, most of the AEs have much smaller returns with some being unprofitable from the trading rule strategies, including Luxembourg (-0.6%), UK (-0.7%) and Switzerland (-1.0%). Furthermore, some trading rules are unprofitable, with the percentage of unprofitable trading rules (column 4) ranging from 5.8% for China to 83.5% for Chile. When excluding these unprofitable trading rules, ranking of market returns remains largely the same (i.e., column 3), except that Philippines becomes the most profitable market (i.e., 6.2%), followed by Peru (5.7%) and China (5.6%).

Significance of these returns is reported in Table 7 and Figure 9. Column 5 of Table 7 reports the bootstrapped p-value, which checks each economy for whether or not the hypothesis of no outperformance for the overall trading strategy is rejected. Figure 9 also depicts the returns and the test of significance in one chart. The number of bootstrap resamples (i.e., B) is set to 1000 to run the test, which is considered sufficiently large to reduce the additional layer of randomness introduced by the resampling scheme.[27] As can be seen, 13 out of 48 market returns (i.e., 27%) have a smaller bootstrapped p-value than the 10% level, meaning that

[25] The OECD uses the real Gross Domestic Product (GDP) as the reference for identification of turning points in the growth cycle for the US. The turning point detection algorithm is a simplified version of the original Bry and Boschan routine, which does not include the correction for outlier, as such a correction is implanted at an earlier stage of the filtering process.

[26] The annualised daily average returns are divided by annualised standard deviation of daily returns. The average annualised SD is 8.4% based on the daily returns from applying trading rules on the 48 sovereign bond market indices. Returns are scaled up to facilitate easier comparisons with the raw returns of sovereign bond market indices.

[27] The smoothing parameter for the stationary bootstrap is set to be 0.1. Hsu and Kuan (2005) find that the smoothing parameters of 0.01, 0.1, and 0.5 in the stationary bootstrap yield similar results.

these market returns are statistically significant. Significant returns are mostly from investing in emerging Asia and other EMEs (see Figure 9 or column 6 of Table 7). Most of the AEs, however, are not significantly profitable from the investment.[28]

5.1.2 Is the predictability higher during the tightening phase of US monetary cycles and US economic recession?

Figure 10 is a scatter plot of the risk-adjusted excess returns from the trading rule strategy, conditional on US monetary regimes. Comparing returns between the US monetary tightening and easing phases, almost two-thirds of sovereign bond markets (i.e., 64.5%) scatter above the 45-degree line (i.e., the dotted line), suggesting that the trading-rule strategy could acquire a higher predictability during the tightening phase than those acquired during the easing phase. Among these markets, most of the AEs scatter closely to the 45-degree line, compared with emerging markets, which scatter widely. In particular, Philippines and Indonesia scatter noticeably above the line while China and Egypt are well below the line.[29]

Figure 11 shows a similar plot to Figure 10, but it is conditional on US business cycles. Comparing returns between the US economic recession and expansion phases, 60% of sovereign bond markets scatter above the 45-degree line, reflecting that the predictability of the trading rule strategy could be higher during the US economic recession phase than the expansion phase, particularly for China, whose return is substantially higher during US economic recessions.[30,31]

5.1.3 Can the predictability of the trading rule strategy be increased by a machine learning algorithm?

Our empirical results show that the machine learning algorithm generally improves the performance of the trading rule strategy. In particular, Emerging Asia benefits the most from the machine learning algorithm, while the additional return of AEs is, on average, lower. These can be seen in (i) Table 8, which summarises the number of sovereign bond markets that have an additional return using our algorithm and

[28] The results remain robust for the average return per transaction. Details can be available upon request.

[29] We check whether there exists any market that has a substantial deviation between different policies/cycles in predictability. Specifically, for each market, we first calculate the deviation in excess returns and then calculate the two influence statistics. A market's excess return is regarded as an outlier compared to other markets when the influence statistic is significantly large in magnitude. Three influence statistics are used, including the scaled difference, and covariance ratio. Based on two influence statistics, China and Egypt are considered outliers, while the other two are marginal.

[30] Based on the two influence statistics, China, Indonesia, Egypt and Morocco are considered outliers.

[31] This finding is also in line with the theory on different responses to information by heterogeneous investors discussed in section II.

(ii) Figure 12, which depicts the distribution of these additional returns. These additional returns are all risk adjusted, annualised and scaled up by the average SD of the excess returns.

As shown in Table 8, 31 of 48 sovereign bond markets (or 65%) have a better performance when using our algorithm. More emerging Asian markets (6 out of 8, or 75%) earn a higher return from using our algorithm, compared to AEs and other EMEs (both at 63%).

As depicted in Figure 12, emerging Asian markets have the strongest improvement when using the algorithm, with an average additional return of 1.4% and a return of 2.2% at the 75th percentile. In comparison, the improvement for AEs is smaller, as reflected in the average additional return (i.e., 0.2%). For EMEs, the additional returns lie between the other two regions (i.e., 0.6%) but have a wider distribution. Overall, the additional return is 0.5% on average, against the average return of 0.4% in the benchmark case.

These results have several implications. First, it suggests that sovereign bond markets, particularly emerging Asian economies, are significantly predictable by applying a trading rule strategy. Second, sovereign bond markets are more predictable during US monetary tightening cycles and economic recessions than during other episodes. Returns for AEs do not differ substantially during different US monetary cycles, while there is a larger dispersion among emerging Asia and other EMEs. Finally, the predictability of advanced markets remains immaterial, while the predictability of emerging markets can be improved by optimising the strategy with a machine learning algorithm. These results can be explained by the fact that US monetary shocks create predictable changes in prices of interest rate sensitive assets (Shynkevich, 2016; Ireland, 2015; and Jiang and Tong, 2017) while the spillover effect of the US bond risk premia on international bond markets have increased given heightened uncertainty in the US monetary policies (Dahlquist and Hasseltoft, 2013) and the economic conditions (Cujean and Hasler, 2017), and that emerging markets tend to have a stronger response to global liquidity conditions after the global financial crisis (Fong et al., 2018).

5.2 What are the determinants of predictability of sovereign bond markets?

In this section, we identify major factors attributed to predictability from sovereign bond markets. As discussed in earlier sections, these factors are mainly associated with four main principal components, which represent (i) the stage of social and economic development, (ii) economic uncertainty, (iii) market depth of public debt and (iv) strength of money. On the technical front, we use the conventional least square regression to link the predictability with all the principal components, in which significant components can be considered important for affecting the predictability in general. We also use a logistic regression model to relate the predictability with principal components given that returns can be categorised as

statistically significant against insignificant. In addition to the common features of linear regressions, the logistic regression offers an estimate of the odds ratio,[32] which helps identify the relative importance of different factors in the regression.

Table 9 presents the empirical results of the two specifications. Focusing on the full model of the least square regression, we find that two of the nine principal components are statistically significant at the 5% level (i.e., column 2). The first (fourth) principal component has a negative (positive) coefficient, meaning that an increase in the principal component's level will decrease (increase) the returns. These empirical findings remain consistent when insignificant variables are removed one by one based on a stepwise regression approach (i.e., column 3). Comparing the two components, the first one has a larger coefficient in magnitude than the fourth one. This suggests that, other things being equal, the predictability has a stronger association with the first component than the fourth one on average, given that all the independent variables (i.e., the principal components) are normalised to be zero mean and unity variance.

Focusing on the results of the logistic regression (i.e., columns 4 and 5), we find that results are consistent with those of the least square regression, with the first and fourth components being significant at the 10% level. After removing insignificant components, the odds ratios of the components suggest that the odds of the predictability would increase by 75.5% (127.2%) respectively when the first (fourth) components increase (decrease) by one SD.

Based on magnitude of the estimated coefficients, the empirical findings suggest that the predictability of sovereign bond markets is affected largely by i) the stage of social and economic development (i.e., the first principal component) and the strength of money (i.e., the fourth principal component). An economy with a more effective government, better regulatory quality, wider financial openness, less political risk and higher income would have a lower predictability of its sovereign bond market, while the predictability is higher when sovereign bond markets face a faster real money growth. The quality of governance and regulatory authorities, level of political risk and income level are likely to reflect the stage of social and economic development, while level of financial openness could reflect the level of openness in the sovereign bond market, which could affect the speed for the incorporation of new information on bond prices. A stronger strength of money indicates a higher likelihood of inflation booms affecting asset values (Puy, 2016). Faster real money growth may also result in higher uncertainty of expected inflation, which could increase economic value from predicting bond returns (Sarno et al., 2016).

[32] The odds of an event occurring is the probability that the event will occur divided by the probability that the event will not occur. In epidemiology, the odds ratio is a relative measure of risk, telling us how much more likely it is that someone who is exposed to a certain risk factor will develop a disease as compared to someone who is not exposed.

VI. Conclusion

By analysing the predictability of 48 sovereign bond markets using four popular classes of technical trading rules with a total of 27,000 variants and a machine learning algorithm in the sample period, we find that some sovereign bond markets, particularly emerging Asian ones, could be predictable. The predictability is also higher when the US tightens its monetary policies or undergoes a recession. In comparison, the predictability for AEs remains lower despite using a machine learning algorithm in adjusting our trading rule strategy. Finally, social and economic development and real money growth can significantly affect predictability, in which the effect of government effectiveness is the largest among other factors.

Our results suggest that some sovereign bond markets could have a higher predictability during tightening of US monetary policies. This highlights the need for policymakers in these markets to contend with potential spillovers from shifts in monetary policy expectations in the US, which are likely to lead to higher government bond interest rates and bouts of volatility. In particular, the informational efficiency of the sovereign bond markets could be a crucial factor that merits policymakers' closer attention.

References

Bessembinder, H., and Chan, K. (1998). "Market efficiency and the returns to technical analysis," Financial Management, 27(2), 5-17.

Brock, W., Lakonishok, J., and LeBaron, B. (1992). "Simple technical trading rules and the stochastic properties of stock returns," Journal of Finance, 47(5), 1731-1764.

Charfeddine, L., Khediri, K.B., Aye, G.C., and Gupta, R. (2016). "Time-varying efficiency of developed and emerging bond markets: Evidence from long-spans of historical data," Physica A: Statistical Mechanics and Its Applications, 505, 632-647.

Chinn, M. D. and Ito, H. (2006). "What Matters for Financial Development? Capital Controls, Institutions, and Interactions," Journal of Development Economics, 81(1), 163-192.

Cujean, J., and Hasler, M. (2017). "Why does Return Predictability Concentrate in Bad Times?" Journal of Finance, 72(6), 2717 – 2758.

Dahlquist, M., and Hasseltoft, H. (2013). "International Bond Risk Premia," Journal of International Economics, 90(1), 17-32.

Dangl, T., and Halling, M. (2012). "Predictive regressions with time-varying coefficients," Journal of Financial Economics, 106(1), 157-181.

Fakhry, B., and Richter, C. (2015). "Is the sovereign debt market efficient? Evidence from the US and German Sovereign Debt Markets," International Economics and Economic Policy, 12(3), 339-357.

Fakhry, B., Masood, O. and Bellalah, M. (2017). "Are the GIPS sovereign debt markets efficient during a crisis?" Journal of Risk, 19(S1), 27-39.

Fong, T.P.W., Li, K.F., Fu, J. (2018) "Accounting for sovereign tail risk in emerging economies: The role of global and domestic risk factors" Emerging Markets Review, 34, 98–110

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2017). "Bond Return Predictability: Economic Value and Links to the Macroeconomy," Management Science, Articles in Advance.

Hall, S.G., and Miles D.K. (1992). "Measuring Efficiency and Risk in the Major Bond Markets," Oxford Economic Paper, 44(4), 599-625.

Hansen, P.R. (2005). "A test for superior predictive ability," Journal of Business and Economic Statistics, 23(4), 365-380.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). "The elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer.

Henkel, S.J., Martin, J.S., and Nadari, F. (2011)."Time-varying short-horizon predictability," Journal of Financial Economics, 99, 560–580.

Hong, H., and Stein, J.C. (1999). "A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets," The Journal of Finance, 54(6), 2143-2184

Hsu, P.H., and Kuan, C.M. (2005). "Reexamining the profitability of technical analysis with data snooping checks," Journal of Financial Econometrics, 3, 606-628.

Hsu, P.H., Taylor, M.P., and Wang, Z. (2016). "Technical Trading: Is it Still Beating the Foreign Exchange Market?" Journal of International Economics, 102, 188-208.

Ivanova, Y., Neely, C.J., and Weller, P.A. (2016). "Can Risk Explain the Profitability of Technical Trading in Currency Markets?" Federal Reserve Bank of St. Louis Working Papers, 2014-33

Ireland, P.N. (2015) "Monetary policy, bond risk premia, and the economy," Journal of Monetary Economics, 76, 124-140.

Jiang, F.W., and Tong, G.S. (2016) "Monetary Policy Uncertainty and Bond Risk Premium," Available at SSRN: https://ssrn.com/abstract=2831092

Ludvigson, S.C., and Ng, S. (2009). "Macro Factors in Bond Risk Premia," Review of Financial Studies, 22(12), 5027 – 5067.

Menkhoff, L. (2010). "The use of technical analysis by fund managers: International evidence," Journal of Banking and Finance, 34(11), 2573-2586.

Menkhoff, L., and Taylor, M.P. (2007). "The Obstinate Passion of Foreign Exchange Professionals: Technical Analysis," Journal of Economic Literature, 45(4), 936-972.

Neely, C.J., Rapach, D.E., Tu, J., and Zhou, G. (2014). "Forecasting the Equity Risk Premium: The Role of Technical Indicators," Management Science, 60(7), iv-vii, 1617-1859

Park, C.H., and Irwin, S.H. (2007). "What do we know about the profitability of technical analysis?" Journal of Economic Surveys, 21, 786 – 826.

Politis, D.N., and Romano, J.P. (1994). "The stationary bootstrap," Journal of the American Statistical Association, 89, 1303 – 1313.

Puy, D. (2016). "Mutual fund flows and the geography of contagion," Journal of International Money and Finance, 60, 73-93.

Rapach, D.E., Strauss J.K., Zhou, G. (2010). "Out-of-sample equity premium prediction: Combination forecasts and links to the real economy," Review of Financial Studies, 23, 821–862.

Sarno, L., Schneider, P., and Wagner, C. (2016). "The economic value of predicting bond risk premia," Journal of Empirical Finance, 37, 247-267.

Shiller, R. J. (1992). "Market Volatility," MIT Press, Cambridge, MA.

Shynkevich, A. (2016). "Predictability in bond returns using technical trading rules," Journal of Banking and Finance, 70, 55-69.

Sullivan, R., Timmermann, A., and White, H. (1999). "Data snooping, technical trading rule performance, and the bootstrap," Journal of Finance, 1647-1691.

Sweeney, R.J. (1986). "Beating the foreign exchange market," Journal of Finance, 41(1), 163-182.

Tian, G.G., Wan, G.H., and Guo, M.Y. (2002). "Market Efficiency and the Returns to Simple Technical Trading Rules: New Evidence from U.S. Equity Market and Chinese Equity Markets," Asia-Pacific Financial Markets, 9(3-4), 241-258.

White, H. (2000) "A Reality Check for Data Snooping," Econometrica, 68, 1097-1126.

Zunino, L., Bariviera, A.F., Guercio, M.B., Martinez, L.B., and Rosso, O.A. (2012). "On the Efficiency of Sovereign Bond Market," Physica A: Statistical Mechanics and Its Applications, 391(18), 4342-4349.

Sovereign bond market indices and grouping

Table 1

Group	Country
Advanced economies	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Iceland, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, UK, US
Emerging Asia	China, India, Indonesia, Korea, Malaysia, Philippines, Taiwan, Thailand
Other emerging market economies	Brazil, Chile, Czech Republic, Egypt, Greece, Hungary, Mexico, Morocco, Nigeria, Peru, Poland, Russia, Slovakia, Slovenia, South Africa, Turkey

Summary statistics on sovereign bond indices

Table 2

Region	Country	Mean	SD	Sharpe ratio	Series start
Advanced economies	Iceland	9.14	9.10	1.00	Jan-2003
	New Zealand	7.92	13.43	0.59	Jan-2000
	Australia	6.60	12.83	0.51	Jan-2000
	Portugal	6.51	13.62	0.48	Jan-2000
	Ireland	6.47	12.48	0.52	Jan-2000
	Switzerland	6.14	11.98	0.51	Jan-2000
	Spain	5.99	12.01	0.50	Jan-2000
	Belgium	5.97	11.20	0.53	Jan-2000
	Italy	5.97	11.98	0.50	Jan-2000
	Austria	5.82	10.87	0.54	Jan-2000
	Denmark	5.74	10.52	0.55	Jan-2000
	France	5.64	10.87	0.52	Jan-2000
	Netherlands	5.54	10.63	0.52	Jan-2000
	Canada	5.53	9.44	0.59	Jan-2000
	Finland	5.39	10.45	0.52	Jan-2000
	Germany	5.36	10.56	0.51	Jan-2000
	Sweden	4.76	12.01	0.40	Jan-2000
	Norway	4.71	12.16	0.39	Jan-2000
	Singapore	4.71	6.61	0.71	Jul-2000
	US	4.59	4.59	1.00	Jan-2000
	UK	4.36	10.71	0.41	Jan-2000
	Hong Kong	3.81	2.83	1.35	Jul-2000
	Luxembourg	2.13	9.71	0.22	Jan-2009
	Japan	1.35	10.73	0.13	Jan-2000
	Group average	**5.42**	**10.47**	**0.56**	

Summary statistics on sovereign bond indices Table 2 (Cont')

Region	Country	Mean	SD	Sharpe ratio	Series start
Emerging Asia	Philippines	10.23	8.29	1.23	Jan-2005
	Indonesia	7.72	13.65	0.57	Jan-2005
	India	6.55	7.59	0.86	Jan-2000
	Thailand	5.99	7.06	0.85	Jan-2003
	China	5.61	3.36	1.67	Jan-2005
	Taiwan	4.11	4.81	0.85	Jul-2000
	Korea	4.08	2.72	1.49	Nov-2011
	Malaysia	2.97	7.47	0.40	Jan-2006
	Group average	**5.91**	**6.87**	**0.99**	
Other emerging market economies	Brazil	10.49	18.50	0.57	Jan-2006
	Poland	7.89	14.91	0.53	Jan-2000
	Hungary	7.82	16.78	0.47	Jan-2000
	Czech Republic	7.71	12.40	0.62	Jan-2000
	Slovakia	7.07	10.56	0.67	Jan-2004
	South Africa	5.51	20.98	0.26	Jan-2000
	Greece	5.38	25.49	0.21	Jan-2000
	Mexico	4.14	13.69	0.30	Jan-2002
	Peru	3.92	7.57	0.52	Jan-2012
	Turkey	3.49	16.45	0.21	Jan-2005
	Russia	3.47	10.01	0.35	Jan-2012
	Slovenia	3.45	11.61	0.30	Jan-2008
	Chile	3.02	9.93	0.30	Jan-2010
	Morocco	2.83	7.06	0.40	Jan-2011
	Nigeria	-1.86	17.88	-0.10	Jan-2012
	Egypt	-2.92	21.04	-0.14	Jan-2011
	Group average	**4.46**	**14.68**	**0.34**	

Notes:

1. "Mean" denotes annualized percentage return on respective sovereign bond index, while "SD" denotes annualized standard deviation of index's daily return.
2. Sharpe ratio is calculated as mean return divided by the standard deviation of returns.

Source: Bloomberg and author estimates.

Data source and definition of macro-economic and governance indicators

Table 3

Variable	Definition	Unit of measurement / scale	Reference time point	Source
Public debt as % of GDP	Cumulative total of all government borrowings less repayments that are denominated in a country's home currency	%	latest available (2016 or 2017)	CIA World Factbook
Real GDP per capita	GDP per capita at PPP	log US$	2016	World Bank
Government effectiveness	Perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies	-2.5 (least favorable) to 2.5 (most favorable)	2016	World Bank
Regulatory quality	Perceptions of the ability of the government to formulate and implement sound policies and regulations which permit and promote private sector development.	-2.5 (least favorable) to 2.5 (most favorable)	2016	World Bank
Financial openness	Chinn & Ito financial openness index which codifies the tabulation of restriction on cross-border financial transactions reported in the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER).	0 (least open) to 1 (most open)	2015 (latest available)	Chinn & Ito (2006)
Political risk	Political stability and absence of violence/terrorism measures perceptions of the likelihood of political stability and/or politically-motivated violence, including terrorism.	-2.5 (least favorable) to 2.5 (most favorable)	2016	World Bank
Real money growth	Year-on-year growth rate of broad money minus inflation rate	%	2016	World Bank
Inflation volatility	Standard deviation of annual inflation rate	%	2000 - 2016	World Bank
Output volatility	Standard deviation of annual real GDP growth rate	%	2000 - 2016	World Bank

Summary statistics on macro variable and governance variables

Table 4

Indicator	Advanced economies	Emerging Asia	Other emerging market economies	All
Public debt as % of GDP (%)	**77.74**	**40.51**	58.87	65.92
Real GDP per capita (In US$)	**10.77**	**9.60**	9.80	10.28
Government effectiveness	**1.58**	0.39	**0.17**	0.94
Regulatory quality	**1.57**	**0.19**	0.21	0.92
Financial Openness	**0.95**	**0.36**	0.58	0.74
Political risk	**0.86**	**-0.54**	-0.29	0.27
Real money growth (%)	**3.33**	**7.45**	6.15	4.87
Inflation volatility (%)	**1.42**	1.95	**2.68**	1.91
Output volatility (%)	2.38	**1.99**	**3.18**	2.58

Notes:

1. Financial openness refers to the Chinn & Ito (2002) Index, which codifies the tabulation of restrictions on cross-border financial transactions reported in the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions (AREAER). It is constructed as the first principal component of four binary variables, which indicate i) presence of multiple exchange rates, ii) restrictions on current account transactions, iii) restrictions on capital account transactions; and iv) requirement of the surrender of export proceeds, respectively. Higher value indicates higher degree of openness for the capital account. The measure spans from 0 (least open) to 1 (most open).
2. Political risk, government effectiveness and regulatory quality come from World Bank Development Indicators. It spans from a scale of -2.5 (least favourable) to 2.5 (most favourable) for each measurement. A higher value indicates a more favourable condition.
3. Figures for each indicator refer to the simple average in each economic grouping. Numbers highlighted in green (red) indicate the group with highest (lowest) value for each indicator.

Correlation matrix of the nine selected macro-economic and financial factors — Table 5

	Public debt as % of GDP	Real GDP per capita	Financial Openness	Government effectiveness	Regulatory quality	Political risk	Real money growth	Inflation volatility	Output volatility
Public debt as % of GDP	1.00								
Real GDP per capita	0.23	1.00							
Financial Openness	0.26	0.76	1.00						
Government effectiveness	0.18	0.89	0.71	1.00					
Regulatory quality	0.11	0.88	0.77	0.96	1.00				
Political risk	0.22	0.80	0.68	0.87	0.85	1.00			
Real money growth	-0.49	-0.56	-0.41	-0.55	-0.49	-0.56	1.00		
Inflation volatility	-0.24	-0.39	-0.50	-0.59	-0.58	-0.53	0.27	1.00	
Output volatility	-0.05	-0.05	-0.13	-0.28	-0.21	-0.23	0.08	0.56	1.00

Principal component analysis of the nine selected macro-economic and financial factors — Table 6

	P1	P2	P3	P4	P5	P6	P7	P8	P9
Public debt as % of GDP	0.15	0.21	**-0.80**	0.35	0.06	0.39	-0.04	0.12	-0.06
Real GDP per capita	**0.40**	0.23	0.17	0.00	0.17	0.20	-0.43	-0.70	-0.12
Financial Openness	**0.42**	-0.01	0.15	-0.12	-0.10	0.25	-0.26	0.35	0.72
Government effectiveness	**0.41**	0.01	0.25	0.00	-0.07	0.08	-0.20	0.53	-0.66
Regulatory quality	**0.36**	0.08	0.10	0.53	0.49	-0.51	0.22	0.08	0.14
Political risk	**0.40**	0.03	0.10	-0.17	-0.13	0.32	0.81	-0.17	-0.03
Real money growth	-0.29	-0.29	0.41	**0.61**	0.05	0.54	0.03	-0.01	0.01
Inflation volatility	-0.29	**0.50**	0.16	-0.30	0.64	0.28	0.08	0.22	0.00
Output volatility	-0.14	**0.75**	0.21	0.29	-0.53	-0.10	0.03	0.03	0.06
Proportion of total variation explained endogenously (%)	**57.42**	**14.33**	**12.63**	**5.88**	3.49	3.06	2.01	0.86	0.32

Note: The figures bonded and underlined for the first 4 principal components (P1 to P4) refer to the variables that are most relevant for each of these components.

Returns from technical trading rules — Table 7

Country/ region	Average Sharpe ratio (annualized and scaled[4]) All rules	Positive rules only	Share of unprofitable rules, %	Bootstrapped p-value	Region
China	5.21	5.64	5.80	***0.00***	Emerging Asia
Philippines	4.73	6.21	16.70	***0.00***	Emerging Asia
Peru	4.46	5.65	14.72	***0.04***	Other EMEs
Greece	3.52	3.88	6.66	***0.01***	Other EMEs
Nigeria	3.01	5.05	24.35	**0.08**	Other EMEs
India	2.89	4.04	19.46	***0.01***	Emerging Asia
Indonesia	2.62	3.76	20.17	***0.04***	Emerging Asia
Brazil	2.37	3.11	15.48	***0.02***	Other EMEs
Hong Kong	1.97	2.79	19.72	**0.05**	AEs
Thailand	1.86	3.82	35.48	**0.08**	Emerging Asia
Taiwan	1.76	2.85	26.16	**0.08**	Emerging Asia
Slovenia	1.41	2.45	24.53	0.13	Other EMEs
Portugal	1.37	2.54	25.64	**0.09**	AEs
Malaysia	1.32	3.02	35.58	0.20	Emerging Asia
Czech Republic	1.30	1.96	20.43	0.12	Other EMEs
Canada	1.00	1.92	29.53	0.16	AEs
Russia	0.90	2.38	32.39	0.28	Other EMEs
Iceland	0.89	2.01	30.28	0.20	AEs
Ireland	0.86	1.95	28.66	0.17	AEs
New Zealand	0.81	1.64	28.33	0.21	AEs
Italy	0.80	1.77	32.06	0.20	AEs
Australia	0.72	1.54	29.37	0.23	AEs
Sweden	0.71	1.68	33.08	0.23	AEs
Finland	0.71	1.60	29.84	0.24	AEs
Morocco	0.70	2.55	39.80	0.31	Other EMEs
US	0.66	1.47	29.88	0.25	AEs
Slovakia	0.66	2.71	38.77	0.29	Other EMEs
Poland	0.65	1.58	32.06	0.25	Other EMEs
Norway	0.60	1.48	32.76	0.28	AEs
Singapore	0.60	2.01	38.63	0.26	AEs
Austria	0.57	1.59	33.54	0.28	AEs
Turkey	0.56	2.10	40.72	0.32	Other EMEs
Spain	0.55	1.61	33.71	0.27	AEs
Netherlands	0.51	1.48	33.25	0.29	AEs
Denmark	0.50	1.48	32.17	0.30	AEs
France	0.47	1.52	35.15	0.30	AEs
Germany	0.41	1.42	34.99	0.33	AEs
Korea	0.39	0.46	11.13	0.06	Emerging Asia
Belgium	0.39	1.59	38.04	0.31	AEs
Egypt	0.35	3.10	48.23	0.35	Other EMEs
Japan	0.19	1.30	42.61	0.40	AEs

Returns from technical trading rules

Table 7 (Cont')

Country/ region	Average Sharpe ratio (annualized and scaled[4]) All rules	Positive rules only	Share of unprofitable rules, %	Bootstrapped p-value	Region
Hungary	-0.10	1.18	52.01	0.53	Other EMEs
South Africa	-0.30	1.04	55.89	0.61	Other EMEs
Luxembourg	-0.55	1.56	58.86	0.62	AEs
Mexico	-0.60	1.11	65.63	0.72	Other EMEs
UK	-0.71	0.70	69.09	0.76	AEs
Switzerland	-1.04	1.01	75.69	0.81	AEs
Chile	-2.30	1.69	83.50	0.89	Other EMEs
Average	***1.06***	***2.34***	***34.17***		

Notes:

1. Positive rules refer those trading rules that could generate positive excess return over benchmark measure (buy-and-hold strategy)
2. Share of unprofitable rules refers to the ratio of the number of rules with negative returns to the total number of effective trading rules. Effective rules are those that generate at least 1 buy or sell signal in the sample period.
3. Bootstrapped p-value refers to the p-value from the SPA test applied on average return from all effective rules, as outlined in the methodology section.
4. The Sharpe ratio is scaled up by the average annualized SD of the returns from applying trading rules on the 48 sovereign bond market indices.

Number of sovereign bond markets that have an additional return by incorporating a machine learning algorithm into the trading rule strategy

Table 8

Region	Improved by machine learning? No	Yes (% of improved markets)	All economies
All economies	17	31 (65%)	48
Emerging Asia	2	6 (75%)	8
Other EMEs	6	10 (63%)	16
AEs	9	15 (63%)	24

Note: "Improved by machine learning" refers to higher average excess returns from a machine learning algorithm, compared with the benchmark strategy where all 27,000 trading rules are included and equally weighted.

Regression results of profitability on principal components derived from nine potential macro-economic and financial factors

Table 9

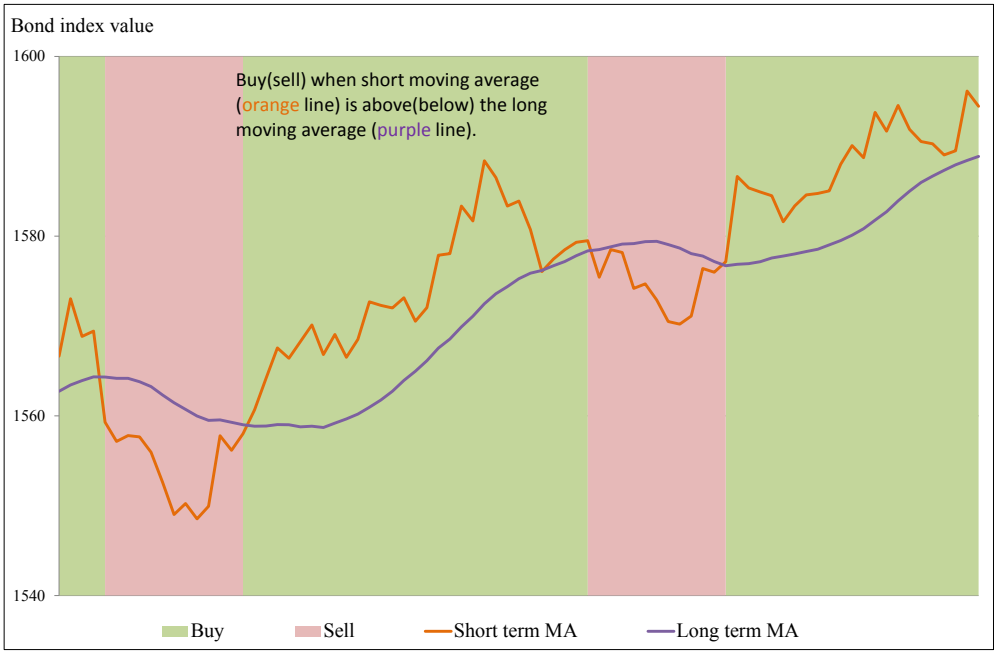
Explanatory Variable	Least square regression		Logistic regression		
	Full model	Selected model	Full model	Selected model	Odds ratio
P1	-0.69*	-0.69*	-1.9*	-1.41*	-75.50%
P2	-0.15		-0.11		
P3	0.05		0.01		
P4	0.46*	0.46*	1.01*	0.82*	127.23%
P5	-0.22		-0.4		
P6	0.21		0.51		
P7	0.17		0.51		
P8	0.01		0.86		
P9	0.24		0.35		
Constant	1.07*	1.07*	-1.92*	-1.61*	
Adjusted R-squared / McFadden R-squared	0.31	0.37	0.45	0.32	
Akaike info criterion	3.38	3.25	1.04	0.88	
Schwarz criterion	3.78	3.41	1.44	1	
Hannan-Quinn criteria	3.53	3.31	1.19	0.93	
F-statistic / LR statistic	3.21	8.24	22.83	15.99	
Prob (F-statistic / LR statistic)	0.01	0	0.01	0	

Notes:

1. Selected model is chosen by using a "stepwise" method based on F-statistic.
2. "*" denotes significant at a 5% level.

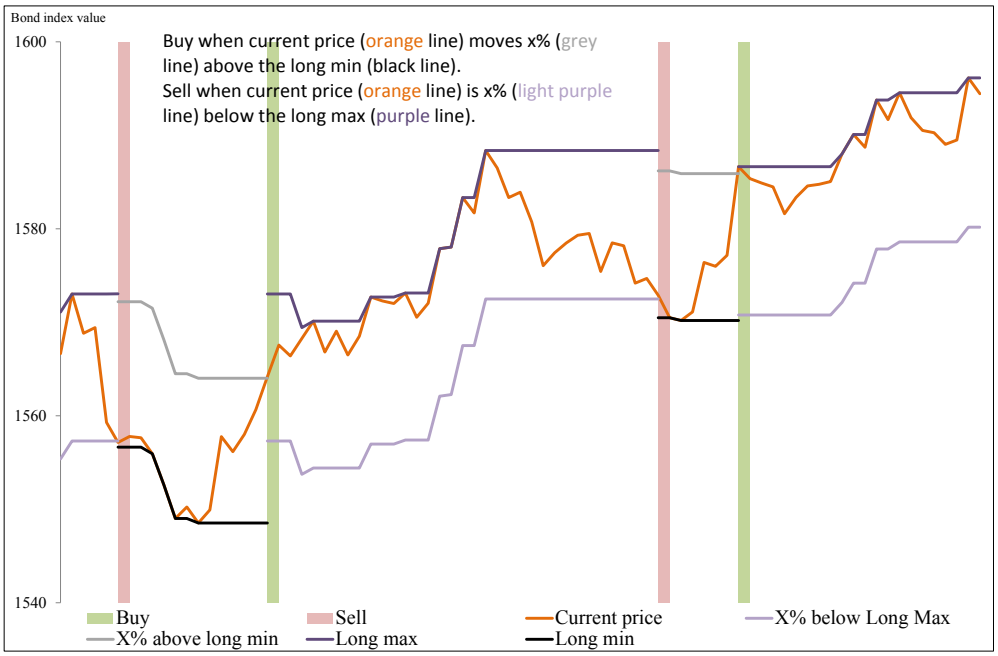
Graphical illustration of MA rule

Figure 1



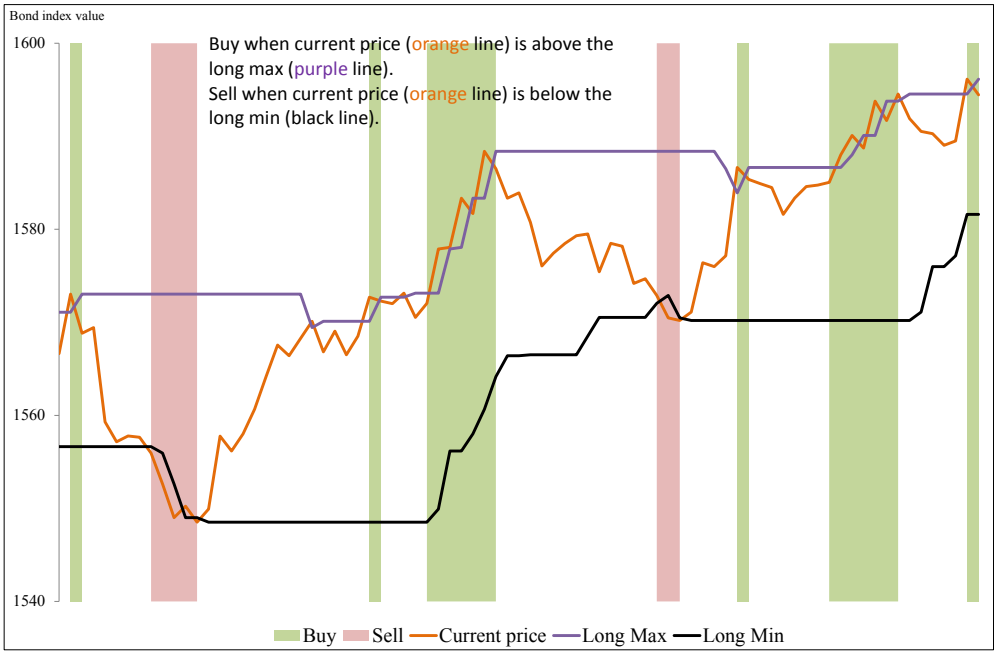
Graphical illustration of FL rule

Figure 2



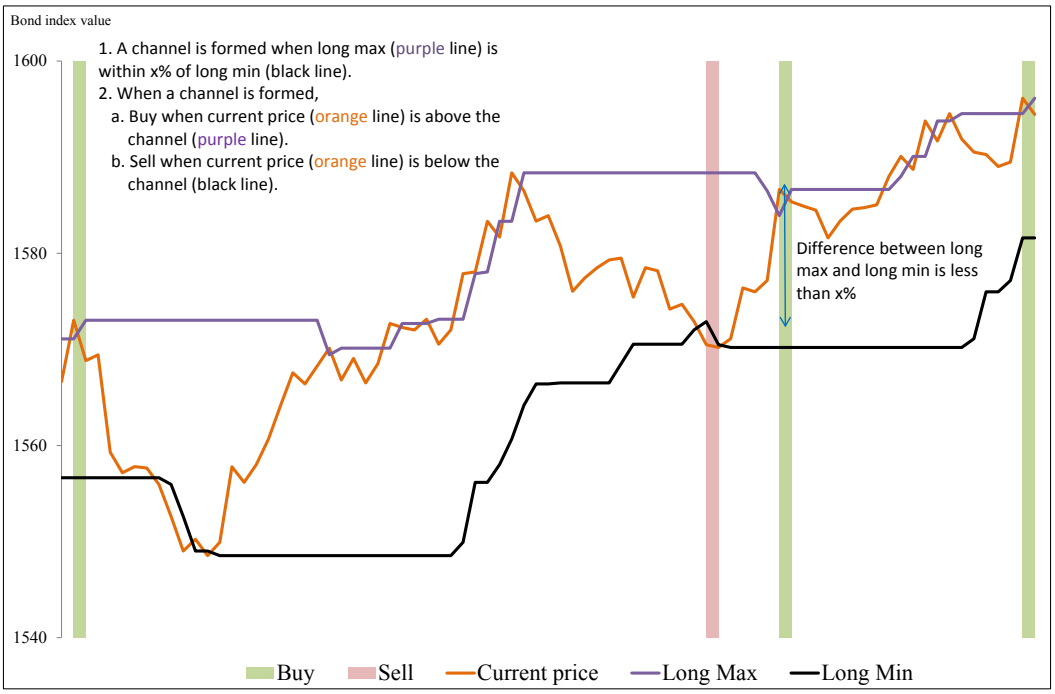
Graphical illustration of SR rule

Figure 3



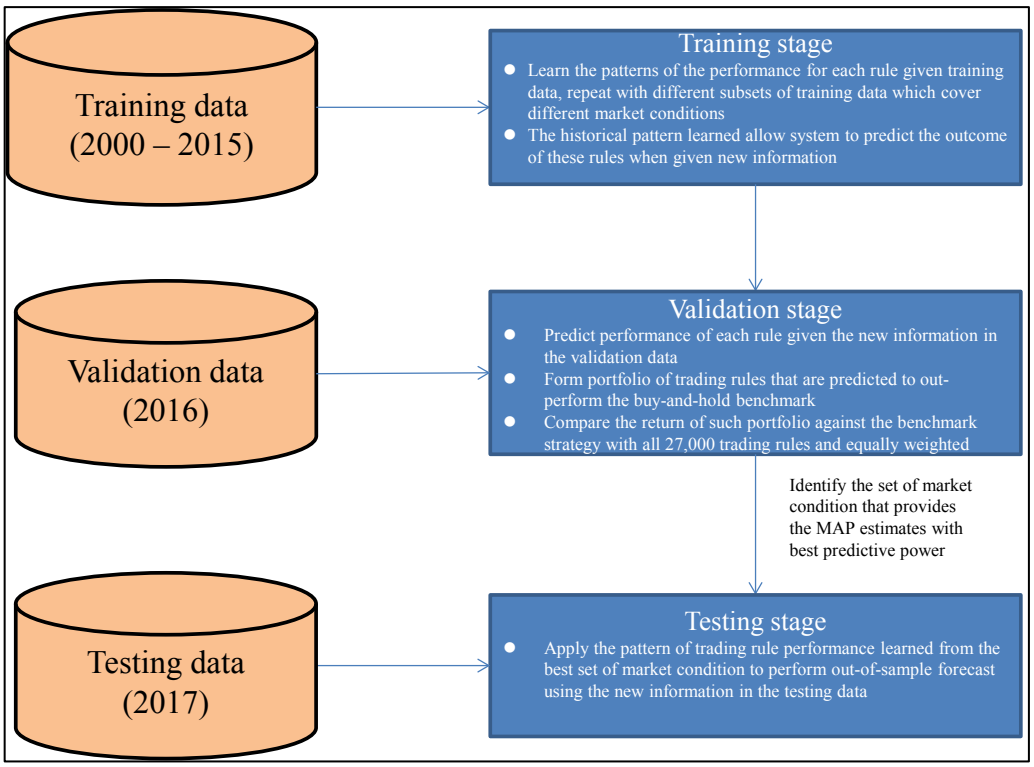
Graphical illustration of CB rule

Figure 4



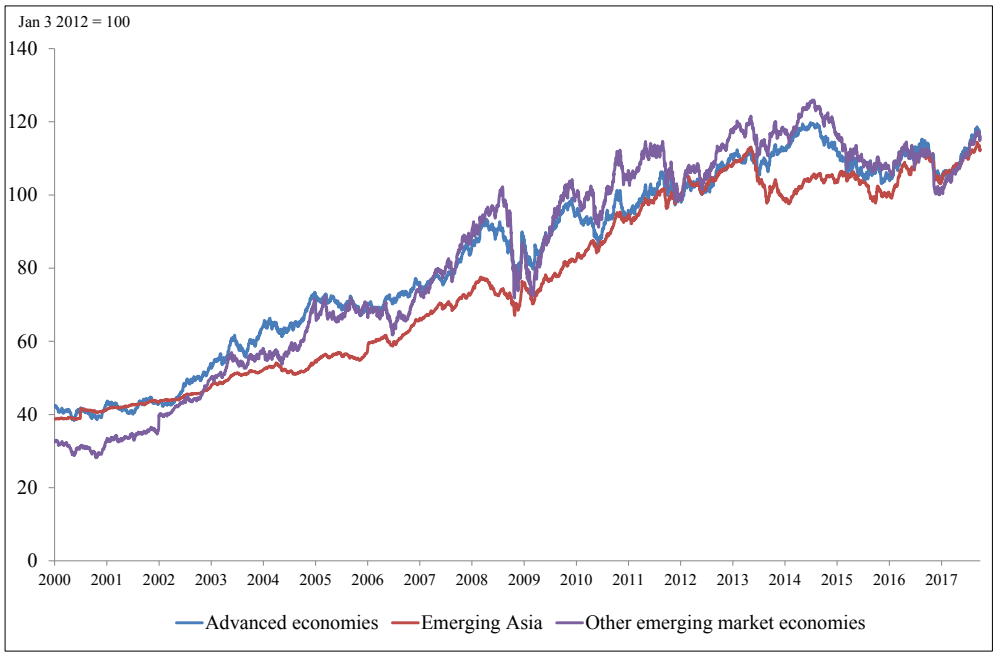
Machine learning system for each sovereign bond market

Figure 5



Time series plot of sovereign bond indices

Figure 6



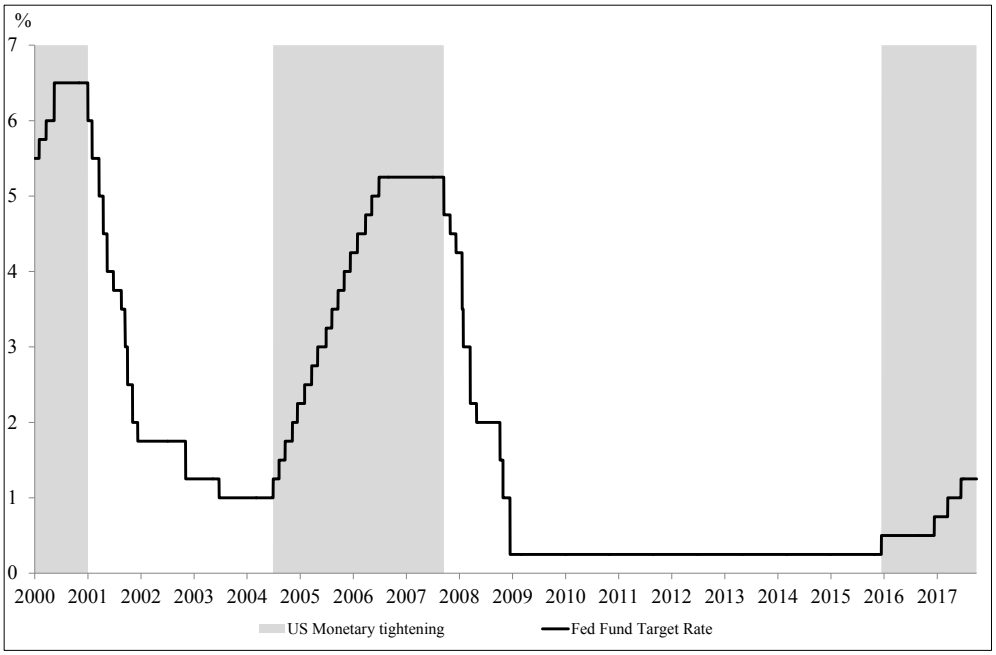
Notes:

1. The time series plots refer to the average bond index values for sovereign bond markets under each economic group.
2. All bond indices are rebased with value at Jan 3 2012 equals 100.
3. Greece is excluded from the calculation for other emerging market economies due to a much more volatile index series when compared to its peers.

Source: Bloomberg

US monetary cycle

Figure 7

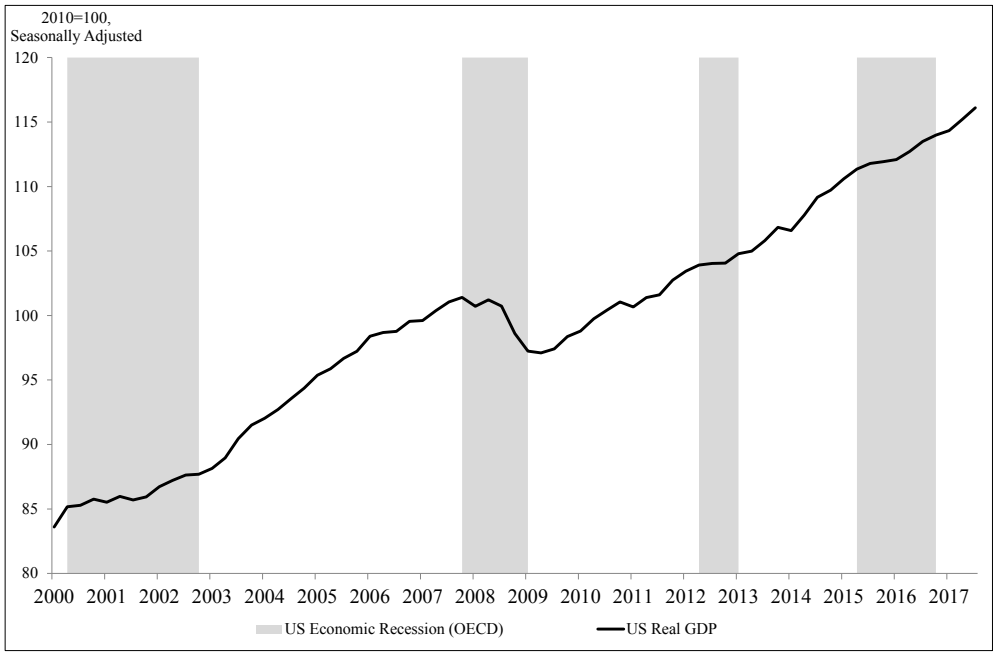


Note: Areas not shaded denote US monetary easing phase.

Sources: Federal Reserve Bank of St. Louis and author estimates.

US business cycle based on OECD definition

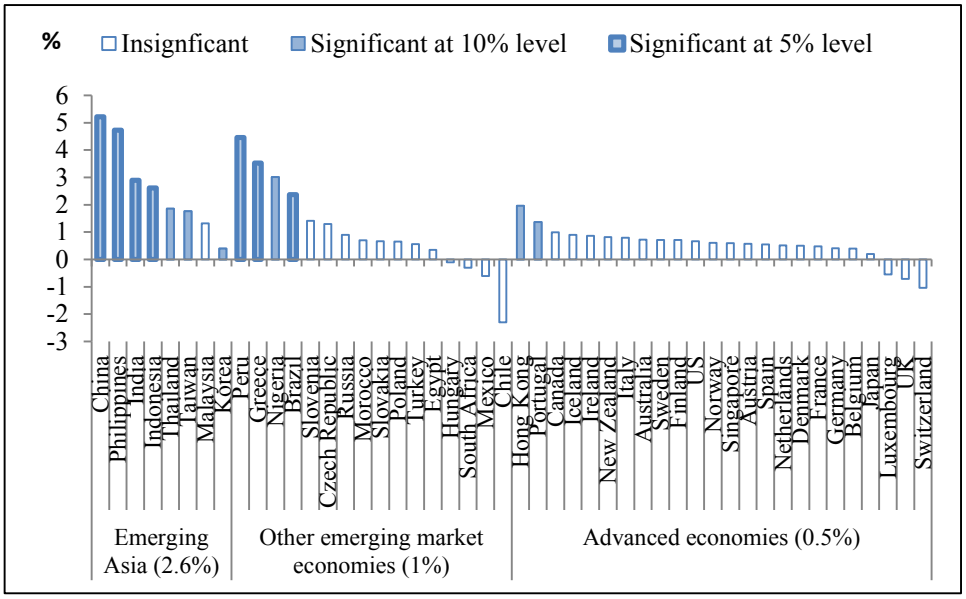
Figure 8



Note: Areas not shaded denote US economic expansion phase.
Sources: Federal Reserve Bank of St. Louis and OECD.

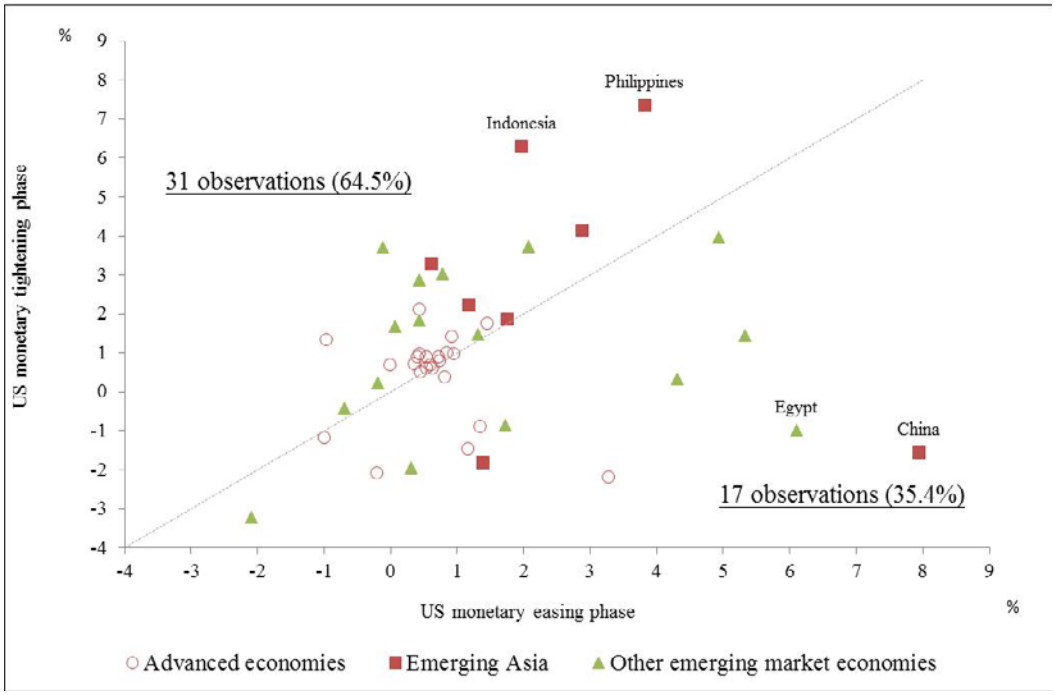
Annualised risk-adjusted average returns from trading-rule-based investment in sovereign bond markets during the sample period from 2000 Q1 to 2017 Q3

Figure 9



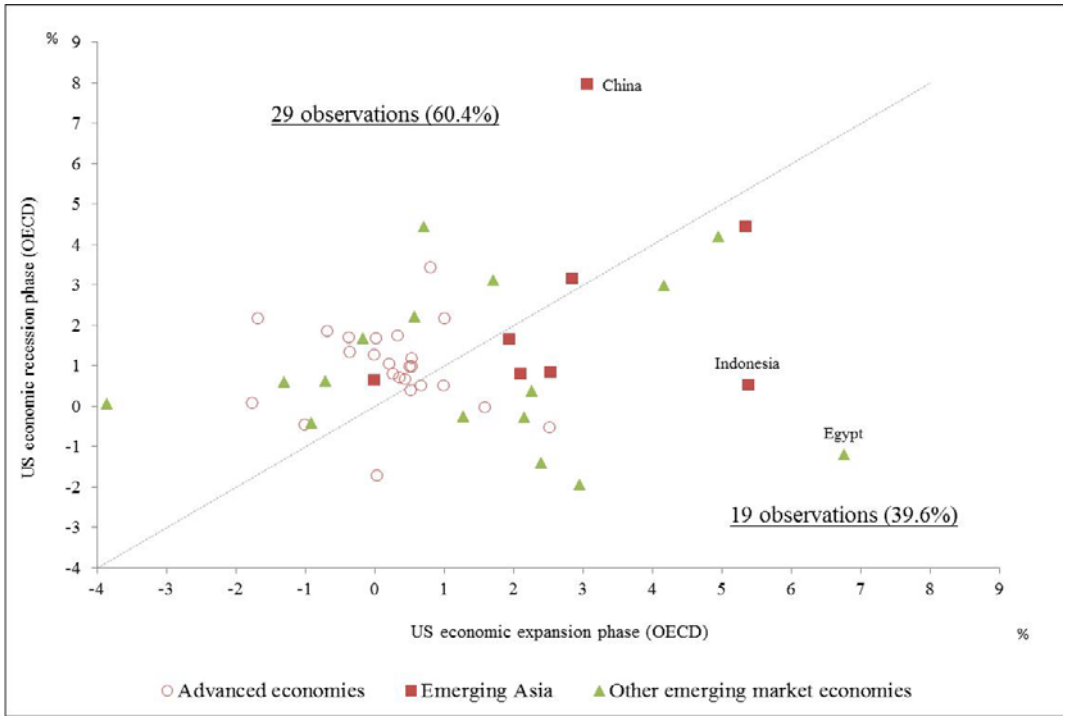
Note: All the returns are risk adjusted, annualized, and scaled up by the average annualized SD of the daily returns from apply trading rules on the 48 sovereign markets.

Scatter plot of conditional returns on the US monetary conditions Figure 10



Note: It refers to the additional excess returns of the trading rule strategy formulated by the machine learning algorithm, over the benchmark strategy where all 27,000 trading rules are included and equally weighted.

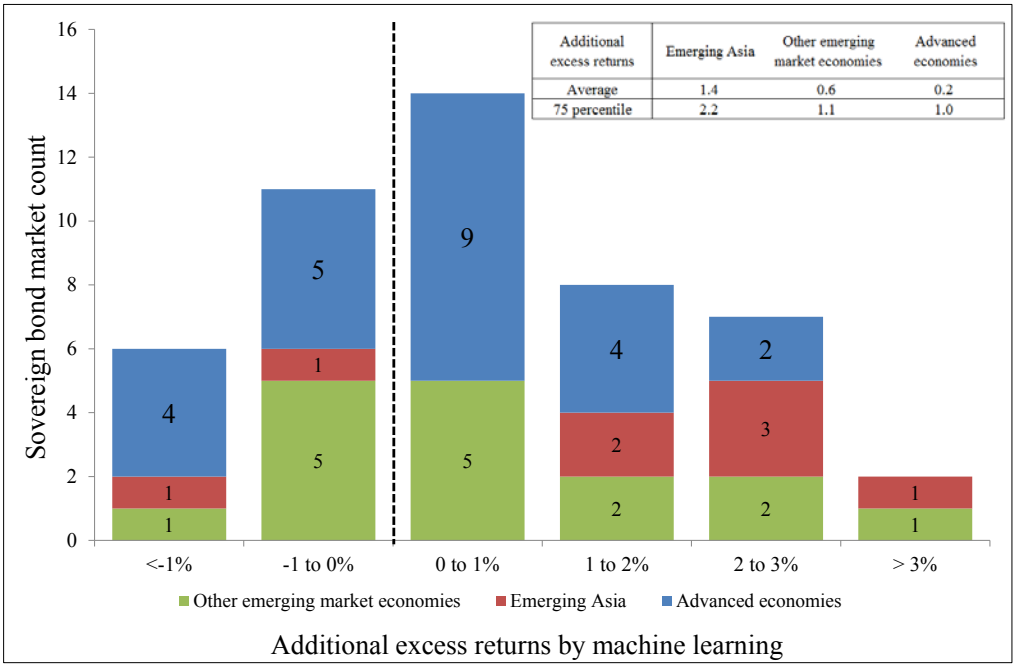
Scatter plot of conditional returns on the US economic conditions Figure 11



Note: All the returns are risk adjusted, annualized, and scaled up by the average annualized SD of the daily returns from apply trading rules on the 48 sovereign markets.

Additional returns earned from trading rule strategy formulated by a machine learning algorithm

Figure 12



Notes: It refers to the additional excess returns of the trading rule strategy formulated by the machine learning algorithm, over the benchmark strategy where all 27,000 trading rules are included and equally weighted.

Appendix: Universe of trading rules

This appendix describes in detail the logic for each class of trading rule, and lists out the parameters and combinations applied, which all follows Shynkevich (2016).

Moving average (MA)

A moving average rule is implemented by first constructing two moving averages: a short term moving average (SMA, average of recent x days' closing prices including current closing price) and long term moving average (LMA, average of recent y days' closing prices including current closing price) where x must be strictly less than y. Buy and sell signals are generated when the SMA crosses the long term LMA. An investor should buy when the SMA is greater than the LMA, and sell when the SMA is less than the LMA.

Three variations on the basic MA rule are also considered:

1. Fixed holding period: all changes in positions are held for a minimum of c days irrespective of the trading signals generated during that time.
2. Fixed percentage band filter: a trading signal is generated only when the difference between the current closing price and the maximum or minimum exceeds a predefined percentage (b),
3. Time delay filter: the trading signal is only generated when it is maintained for d days.

The specifications for different parameters are described as below;

x: number of days in a short moving average
y: number of days in a long moving average
z: number of x–y combinations where y is strictly less than x
b: fixed band multiplicative value
d: number of days for the time delay filter
c: number of days a position is held, ignoring all other signals during that time
x = 1, 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 175 (14 values)
y = 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 175, 200 (14 values)
z = x + x * (y-1)/2 = 14 + 14 * 13 / 2 = 105
b = 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05 (10 values)
d = 2, 3, 4, 5 (4 values)
c = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 (11 values)

Number of rules in MA class = $z \times(1 + b + d + c + b \times c) = 105 \times(1 + 10 + 4 + 11 + 10 \times 11) = 14{,}280$

Filter rules (FL)

A standard filter rules generates a buy (sell) signal when current closing price increases (decreases) by at least x percent above (below) subsequent minimum (maximum). In basic form, subsequent maximum (minimum) is defined as the highest (lowest) price while holding a buy (sell) position (not including current closing price).

Three variations on the standard FL rule are also considered:

1. Allow for neutral positions to be held if the increase or decrease in the price

is more than another predefined threshold (y, where y<x).
2. Redefine high (low) prices to be highest (lowest) closing price for the previous e days (not including current closing price), where e is a predefined number.
3. Fixed holding period: all changes in positions are held for a minimum of c days irrespective of the trading signals generated during that time.

The specifications for different parameters are described as below;

x: percentage change in price to initiate a position
y: percentage change in price to liquidate a position
z: number of x–y combinations where y is strictly less than x
e: number of days to define a local high (low)
c: number of days a position is held, ignoring all other signals during that time
x = 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.22, 0.24, 0.26, 0.28, 0.3 (24 values)
y = the same 24 values as
z = x * (y −1)/2 = 24 * 23/2 = 276
k = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 (11 values)
c = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 (11 values)

Number of rules in FL class = x ×(1 + k + k ×c) + z ×(1 + k) = 24 ×(1 + 11 + 11 ×11) + 253 ×(1 + 11) = 6504

Support and resistance (SR)
Rules based on support and resistance level involve buying the asset when the current closing price exceeds a local maximum (resistance) and selling when the closing price is less than a local minimum (support). The local maximum (minimum) is defined as the highest (lowest) closing price over the previous n days (excluding current closing price).

Three variations on the basic SR rule are also considered:
1. Fixed holding period: all changes in positions are held for a minimum of c days irrespective of the trading signals generated during that time.
2. Fixed percentage band filter: a trading signal is generated only when the difference between the current closing price and the maximum or minimum exceeds a predefined percentage (b),
3. Time delay filter: the trading signal is only generated when it is maintained for d days.

The specifications for different parameters are described as below;

n: number of days in the support and resistance range
b: fixed band multiplicative value
d: number of days for the time delay filter
c: number of days a position is held, ignoring all other signals during that time
n = 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 175, 200 (14 values)
b = 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05 (10 values)
d = 2, 3, 4, 5 (4 values) c = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 (11 values)

Number of rules in SR class = n ×(1 + b + d + c + b ×c + d ×c) = 14 ×(1 + 10 + 4 + 11 + 10 ×11 + 4 ×11) = 2520

Channel breakout (CB)

Channel breakout can be considered as a variation of support and resistance with additional "channel" criteria on local maximum and minimum. A buy (sell) signal is triggered when the current closing price moves above (below) a channel, where a channel is defined as the occasion where the local maximum is within x percent of the local minimum. The local maximum and minimum are defined in the same way as that under the support and resistance rule.

The following variation on the basic CB rule is also considered:

1. Fixed holding period: all changes in positions are held for a minimum of c days irrespective of the trading signals generated during that time.

The specifications for different parameters are described as below;

n: number of days for a channel
x: difference between the high price and the low price as a percentage of the low price required to form a channel
c: number of days a position is held, ignoring all other signals during that time
n = 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 175, 200 (14 values)
x = 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.22, 0.24, 0.26, 0.28, 0.3 (24 values)
c = 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 (11 values)

Number of rules in CB class = n ×x ×c = 14 ×24 ×11 = 3696

Total number of rules = 6,504 (24.1%) + 14,280 (52.9%) + 2,520 (9.3%) + 3,696 (13.7%) = **27,000**

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Predictability in sovereign bond returns using technical trading rule: do developed and emerging markets differ?[1]

Tom Fong and Gabriel Wu,

Hong Kong Monetary Authority

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Predictability in sovereign bond returns using technical trading rules: do developed and emerging markets differ?

Presented by *Tom Fong*
Hong Kong Monetary Authority

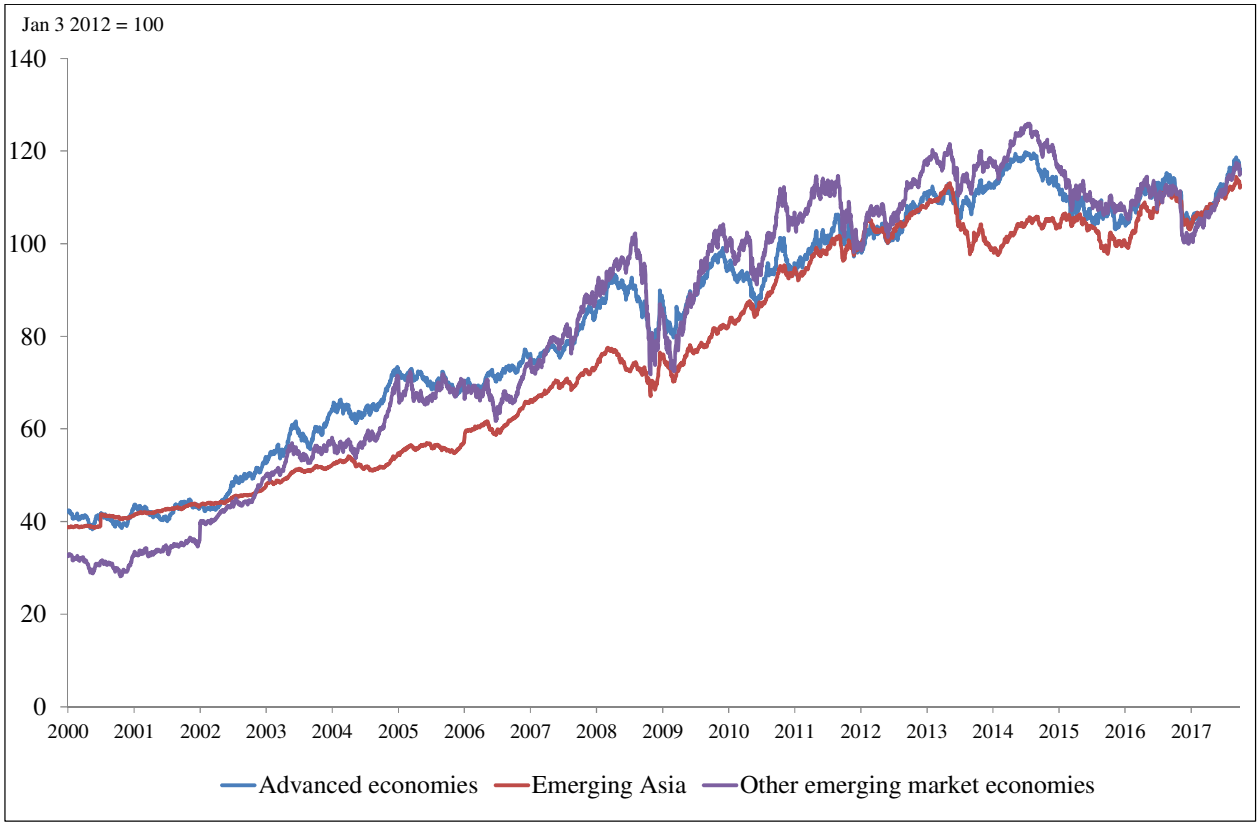
(collaborated with Gabriel Wu)

Bank Indonesia / IFC
"International Workshop on Big Data for Central Bank Policies"
Bali, 25 July 2018

Agenda

- Objective of the study
- Major findings
- Analytical framework
- Empirical results
- Conclusion

Sovereign bonds appear to be more responsive since major AEs begin their monetary policy normalization



Notes:

1. The time series plots refer to the average bond index values for sovereign bond markets under each economic group.
2. All bond indices are rebased with value at Jan 3 2012 equals 100.
3. Greece is excluded from the calculation for other emerging market economies due to a much more volatile index series when compared to its peers.

Source: Bloomberg.

Our objective

- Evaluate the predictability of sovereign bond markets using technical trading rules
- Evaluate the robustness of the predictability using both a statistical test and machine learning technique
- Assesses whether the predictability is affected or not by changes in the monetary policy and economic business cycle in the US.
- Identify potential factors driving the predictability

Major findings

- Investing in sovereign bond markets of EMEs, particularly of emerging Asian economies, are significantly profitable
- Profits from investing in advanced economies' remains very thin even though we use an advanced (machine-learning) technique in profit optimization
- The profit is notably higher when the US tightens its monetary policies or undergoes an economic recession
- Several domestic factors, including government effectiveness, regulatory quality, political risk, financial openness, income level and real money growth, can significantly affect the predictability

Analytical framework

- 4 classes of trading rules considered
  1. Moving average (MA)
  2. Filter (FL)
  3. Support and resistance (SR)
  4. Channel breakout (CB)

  ➢ A total of **27,000** trading rules considered

- Performance measured by excess return over “buy-and-hold” strategy

- Robustness of the return is tested by two advanced methods
  1. Superior Predictive Ability (SPA) test
  2. Naïve Bayes Classifier (NBC) technique

Data

- 48 Bond index data
  - Bank of America (BofA) Merrill Lynch sovereign bond index
  - Local currency, fixed rate nominal sovereign debt with maturity over 1 year
  - Weighted by market capitalization
  - Total return index
  - Covering both AEs and EMEs
  - With sample period from Jan 2000 to Sep 2017

Notable differences in the returns of sovereign bond markets among different economic groups

Economic group	Mean	SD	Sharpe ratio
AE	5.42	10.47	0.56
Emerging Asia	6.22	8.13	0.87
Other EMEs	4.46	14.68	0.34

Notes:

1. AEs and EMEs (including emerging Asia and other EMEs) classified according to the MSCI classification of developed and emerging markets
2. Mean" denotes annualized average daily return on respective sovereign bond index, while "SD" denotes annualized standard deviation of index's daily return.
3. Sharpe ratio is calculated as mean return divided by the standard deviation of returns.

- Emerging Asian markets have the highest return than other markets, after adjusting for risk (i.e. Sharpe ratio)

These economic groups also display vastly different economic, social and financial conditions

Variable	AE	Emerging Asia	Other EMEs
Public debt as % of GDP (%)	**77.74**	**40.51**	58.87
Real GDP per capita (ln US$)	**10.77**	**9.60**	9.80
Government effectiveness	**1.58**	0.39	**0.17**
Regulatory quality	**1.57**	**0.19**	0.21
Financial Openness	**0.95**	**0.36**	0.58
Political risk	**0.86**	**-0.54**	-0.29
Real money growth (%)	**3.33**	**7.45**	6.15
Inflation volatility (%)	**1.42**	1.95	**2.68**
Output volatility (%)	2.38	**1.99**	**3.18**

Note: Figures for each indicator refer to the simple average in the each economic grouping. Numbers highlighted in **green** (**red**) indicate the group with highest (lowest) value for each indicator.

- AEs are characterised with a deeper market for public bonds, a more effective government and a higher degree of financial openness
- Emerging Asia displays stronger growth of money in real terms
- While other EMEs show a larger volatility in both inflation and output

Principal component analysis is applied to group similar factors together

Variable	1st PC	2nd PC	3rd PC	4th PC
Public debt as % of GDP (%)	0.15	0.21	**-0.80**	0.35
Real GDP per capita (ln US$)	**0.40**	0.23	0.17	0.00
Government effectiveness	**0.42**	-0.01	0.15	-0.12
Regulatory quality	**0.41**	0.01	0.25	0.00
Financial Openness	**0.36**	0.08	0.10	0.53
Political risk	**0.40**	0.03	0.10	-0.17
Real money growth (%)	-0.29	-0.29	0.41	**0.61**
Inflation volatility (%)	-0.29	**0.50**	0.16	-0.30
Output volatility (%)	-0.14	**0.75**	0.21	0.29
Proportion of total variation explained endogenously (%)	**57.42**	**14.33**	**12.63**	**5.88**

Note: For brevity, only the first 4 principal components (which in total explain 90% of the variations) are shown.

- Based on principal component analysis, we can classify the first 4 components as below
  - 1st component: Stage of social and economic development
  - 2nd component: Economic uncertainty
  - 3rd component: Market depth of sovereign bond market
  - 4th component: Strength of money

Superior Predictive Ability test

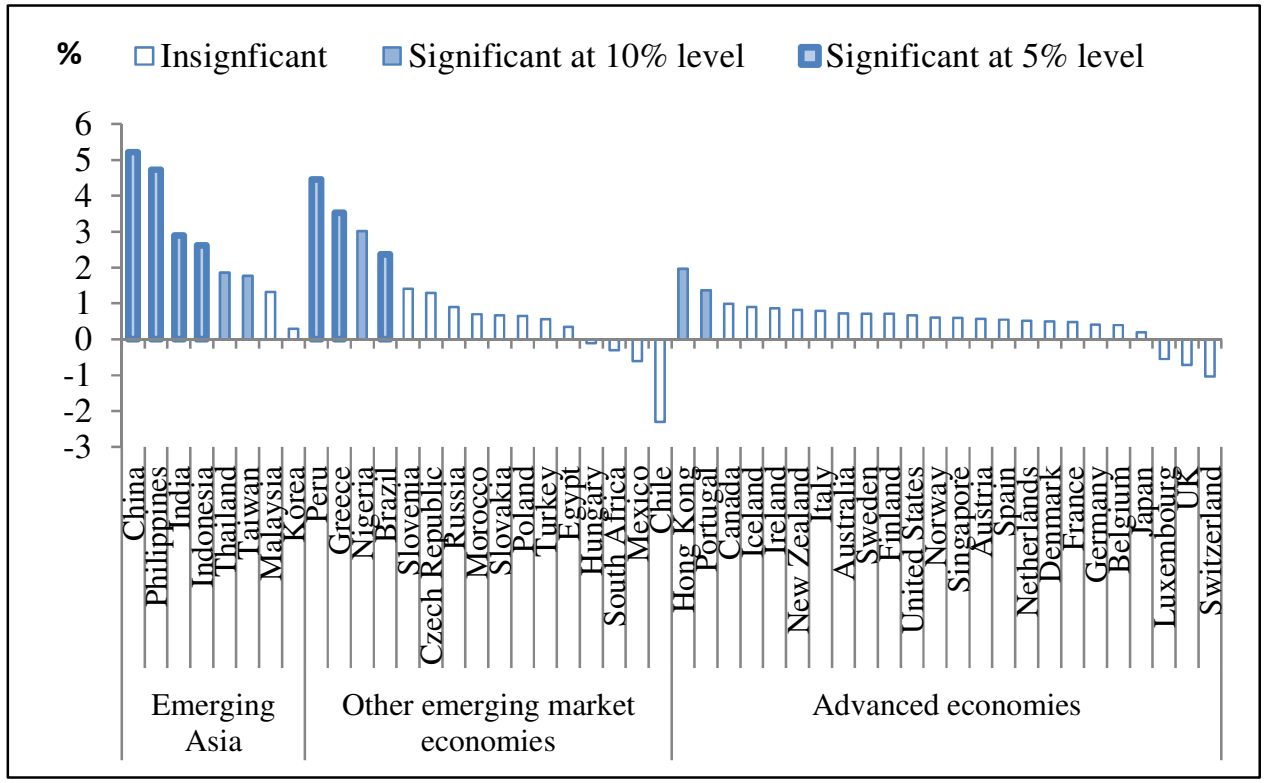
- Applying a large number of trading rules on predicting returns are susceptible to unintentional data mining or data snooping problem

- Hansen (2005)'s Superior Predictive Ability (SPA) test address such problem, as it takes into account the joint distribution of all trading rules.

- To obtain the distribution, the test involves a large scale bootstrapping simulation on the trading rule returns:
  - 27,000 trading rules
  - 1,000 bootstrapping simulations
  - 48 sovereign bond indices in daily frequency
  - Thousands of observations for each index

- The output of the SPA test is the p-value, which indicates the level of significance that trading rule returns are greater than 0

SPA test indicates significant trading rule returns for most of the emerging Asian markets



Notes

1. All the returns are risk adjusted, annualized, and scaled up by the average annualized SD of the daily returns from apply trading rules on the 48 sovereign markets.
2. Statistically significance of average trading rule returns is determined by the p-value of SPA test.

- 12 out of 48 markets record significant trading rule returns at 10% level
- Among them, half of them are from emerging Asia

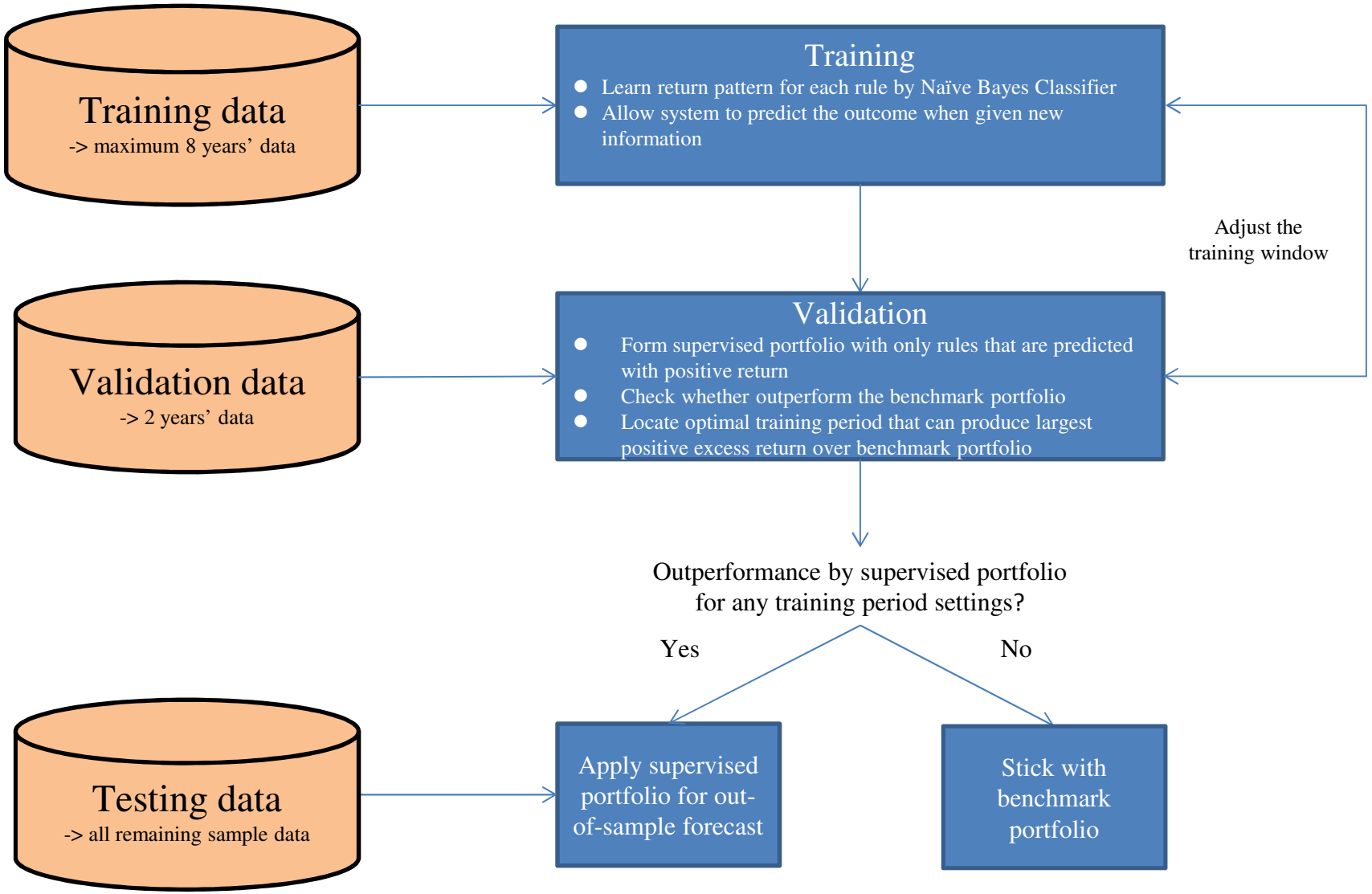
Machine learning – Naïve Bayes Classifier

- Robustness of the profitability of each sovereign bond market is also tested by a simple machine learning system, which determines rooms for increasing the average returns acquired from the trading-rule investment.

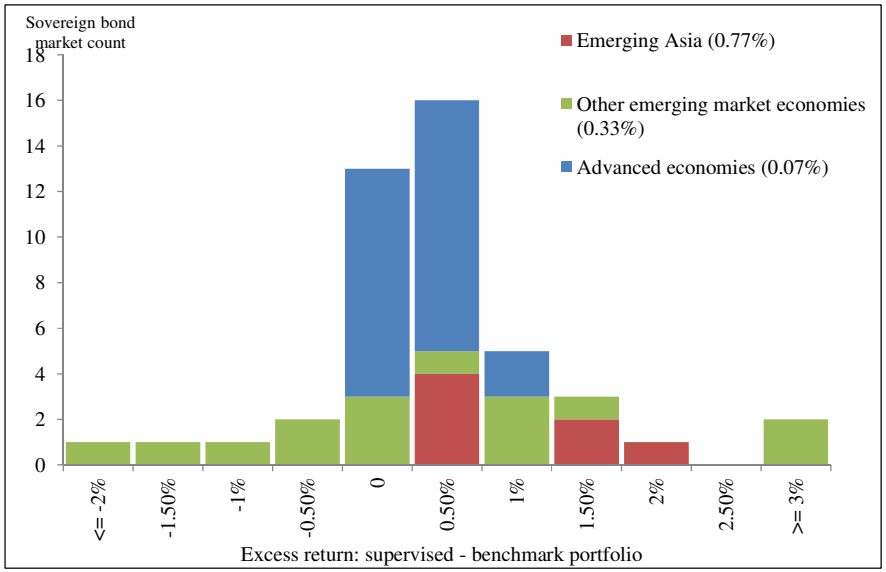
- The training system involves 2 parts; training and validation stage.
  - In the training stage, Naïve Bayes Classifier technique is adopted for system to "learn" the pattern of each trading rule's return given its historical investment performance.
  - The knowledge learned is then applied in the validation stage in forming a "supervised" portfolio, where only trading rules predicted with good performance are included, with higher weights assigned to those that are more likely to be profitable. The supervised portfolio is then evaluated to see whether it outperforms a benchmark portfolio which uses all 27,000 trading rules with equal weighting.

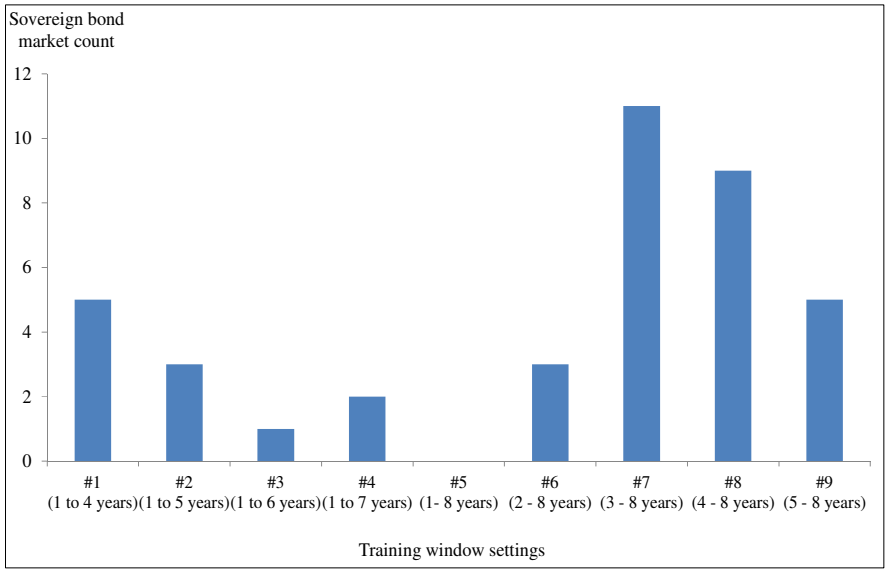
- Once the system is validated to be successful in improving returns for certain sovereign bond market, it will be applied to a set of testing data as out-of-sample forecasting (testing stage).

Machine learning set-up



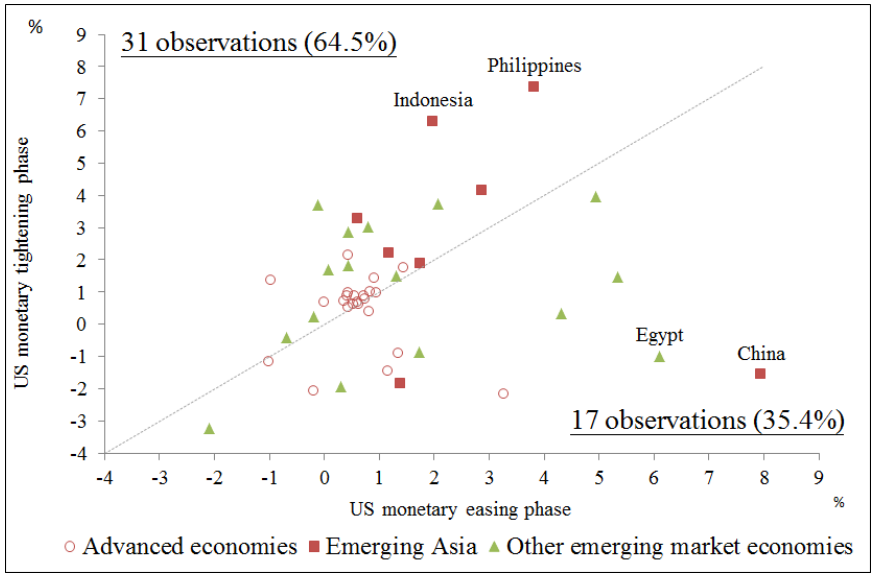
Improvements in returns for EMEs with machine learning technique, but not really for AEs

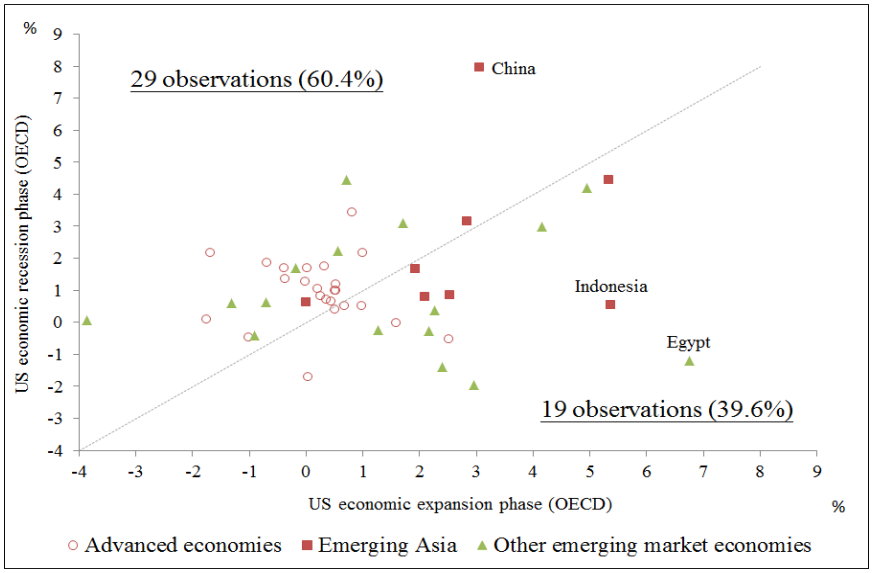




- Returns of emerging markets can substantially increase when applying machine-learning technique in profit optimization.
- However, returns of advanced markets cannot be increased further despite applications of the advanced technique.
- Optimal learning period usually fall at the later part of the training data

Trading rules usually attain higher returns when US tightens its monetary policy or undergoes economic recession





- Returns for AEs do not differ substantially during different US monetary cycles, while there are a larger dispersion among both emerging Asia and other EMEs.
- Chine earns a substantially higher returns during US economies recession than expansion.

Return predictability is associated with level of social and economic development, and rate of real money growth

Explanatory Variable	Least square regression		Logistic regression		
	Full model	Selected model	Full model	Selected model	Odds ratio (selected model)
P1	-0.69*	-0.69*	-1.9*	-1.41*	-75.50%
P2	-0.15		-0.11		
P3	0.05		0.01		
P4	0.46*	0.46*	1.01*	0.82*	127.23%
P5	-0.22		-0.4		
P6	0.21		0.51		
P7	0.17		0.51		
P8	0.01		0.86		
P9	0.24		0.35		
Constant	1.07*	1.07*	-1.92*	-1.61*	
Adjusted R-squared / McFadden R-squared	0.31	0.37	0.45	0.32	
Akaike info criterion	3.38	3.25	1.04	0.88	
Schwarz criterion	3.78	3.41	1.44	1	
Hannan-Quinn criteria	3.53	3.31	1.19	0.93	
F-statistic / LR statistic	3.21	8.24	22.83	15.99	
Prob (F-statistic / LR statistic)	0.01	0	0.01	0	

Notes:

1. Dependent variable used in each model: trading rule returns (least square regression); binary variable which equals 1 when trading rule is statistically significant at 10% level (logistic regression).
2. Selected model is chosen by using a "stepwise" method based on F-statistic. "*" denotes significant at a 5% level.

- Both ordinary least square regression and logistic regression are estimated.
- Principal components extracted from the observable factors as explanatory variables
- The 1st PC (stage of social and economic development) and 4th PC (strength of money) are found to be significant under both specifications

Conclusion

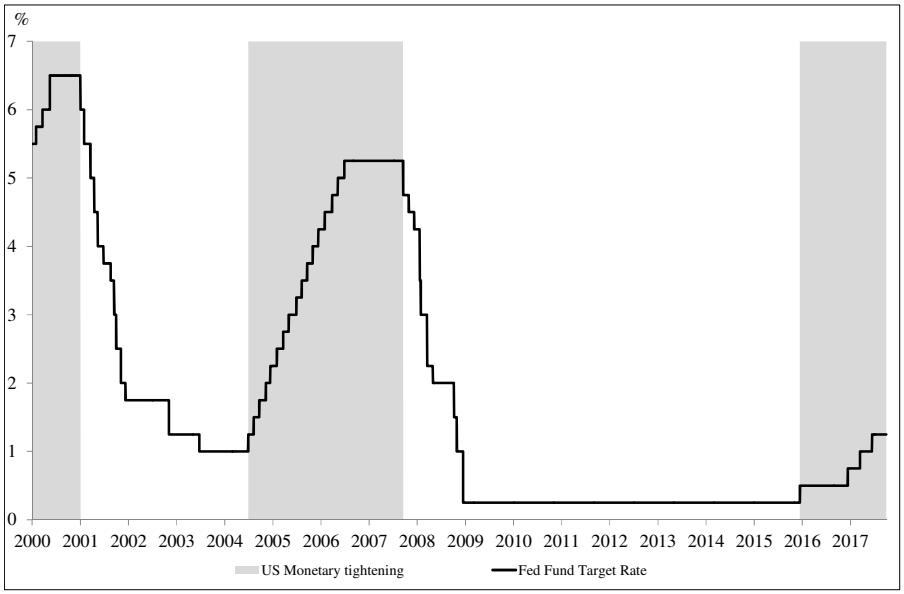
- Sovereign bond markets could be predictable with trading rule strategies
- Robustness of the predictability is tested by advanced data mining and machine learning technique
- Policy implications:
  - Considerable spillover impact from the US to inefficient markets such as emerging Asian ones
  - Promoting not only the financial development but also the social and economic development

Appendix: Classification of countries/regions by economic grouping

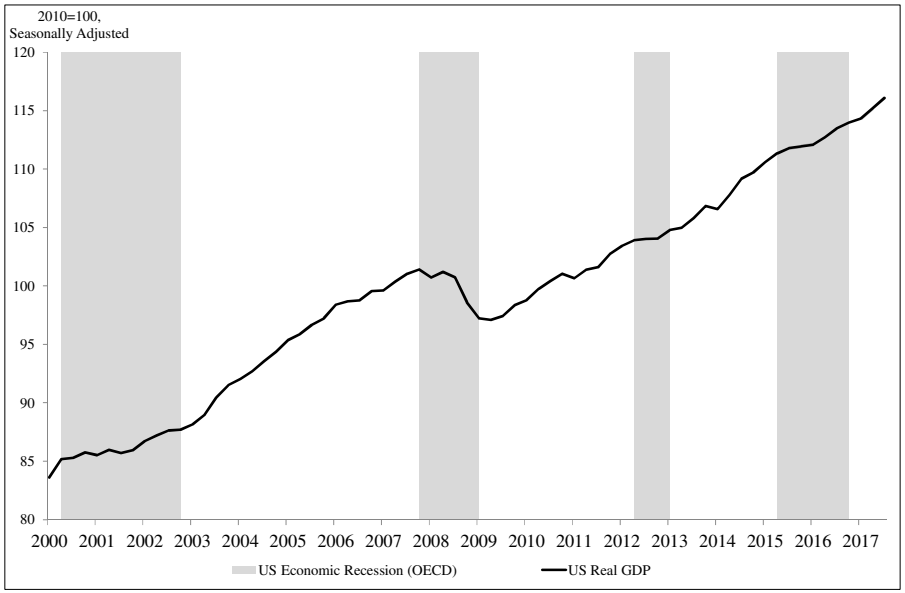
Advanced economies			Emerging Asia	Other emerging market economies	
Australia	Hong Kong	Norway	China	Brazil	Nigeria
Austria	Iceland	Portugal	India	Chile	Peru
Belgium	Ireland	Singapore	Indonesia	Czech Republic	Poland
Canada	Italy	Spain	Korea	Egypt	Russia
Denmark	Japan	Sweden	Malaysia	Greece	Slovakia
Finland	Luxembourg	Switzerland	Philippines	Hungary	Slovenia
France	Netherlands	UK	Taiwan	Mexico	South Africa
Germany	New Zealand	US	Thailand	Morocco	Turkey

Note: Classification according to the MSCI classification of developed and emerging markets, see https://www.msci.com/market-classification for details.

Appendix: US monetary and business cycles



Note: Areas not shaded denote US monetary easing phase.
Sources: Federal Reserve Bank of St. Louis and author estimates.



Note: Areas not shaded denote US economic expansion phase.
Sources: Federal Reserve Bank of St. Louis and OECD.

Appendix: Technical details of the SPA test

- The SPA test in this study is based on the following test statistics

$$\overline{V_l} = \frac{1}{K}\sum_{k=1}^{K}\left(\sqrt{N} * \overline{ER_k}\right) / \sigma_k \qquad (1)$$

where $\overline{ER_k} = \sum_{t=201}^{T} ER_{k,t} / N$ is the average excess return for the *k*-th trading rule out of *K* trading rules and *N*=*T*-200 is the sample size, and σ_k is a consistent estimator for the standard deviation of $\sqrt{N} * \overline{ER_k}$

- The joint distribution of all trading rules is empirically drawn by applying stationary bootstrap method of Politis and Romano (1994) to the observed values of $ER_{k,t}$

- In each bootstrapping simulation, we compute the sample average of the bootstrapped returns denoted by $\overline{ER^*_{k,i}}$ The process is repeated B times and we construct the following bootstrap test statistics to form the distribution for $\overline{V_l}$;

$$\overline{V_{l,i}} = \frac{1}{K}\sum_{k=1}^{K}\left(\sqrt{N} * (\overline{ER^*_{k,i}} - \overline{ER_k} * I_{((\sqrt{N} * \overline{ER_k}) / \sigma_k > -A)})\right) / \sigma_k \qquad (2)$$

where i = 1,2,….B and I is an indicator function which equals one when the condition is satisfied and zero otherwise, and A = $\sqrt{2 \ln \ln N}$

- The test's p-value is subsequently obtained by comparing $\overline{V_l}$ with the quantiles of $\overline{V_{l,i}}$.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Measuring market and consumer sentiment and confidence[1]

Stephen Hansen,

University of Oxford

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Measuring Market and Consumer Sentiment and Confidence

Bank Indonesia—IFC Workshop

Stephen Hansen
University of Oxford

Introduction

The first lecture focused mainly on recovering the underlying structure of text.

The methods we discussed did not seek to directly explain any label associated with text.

In many cases, we are interested in mapping the content of text into some outcome variable of interest.

One prominent example of this is sentiment analysis, in which we wish to associate text with the sentiment it reflects.

We will again see a distinction between word-count exercises and machine learning approaches.

Tetlock (2007)

Tetlock (2007) is a highly cited paper that applies dictionary methods to the Wall Street Journal's "Abreast of the Market" column.

Uses Harvard IV-4 dictionaries
`http://www.wjh.harvard.edu/~inquirer`.

Large number of categories: positive, negative, pain, pleasure, rituals, natural processes, etc. 77 in all.

Count number of words in each dictionary in each column from 1984-1999.

Principal components analysis shows most variation on dimensions that reflect pessimism: negative, weak, fail, fall.

Tetlock (2007)

Tetlock (2007) is a highly cited paper that applies dictionary methods to the Wall Street Journal's "Abreast of the Market" column.

Uses Harvard IV-4 dictionaries
`http://www.wjh.harvard.edu/~inquirer`.

Large number of categories: positive, negative, pain, pleasure, rituals, natural processes, etc. 77 in all.

Count number of words in each dictionary in each column from 1984-1999.

Principal components analysis shows most variation on dimensions that reflect pessimism: negative, weak, fail, fall.

Main result: pessimism predicts low short-term returns (measured with the Dow Jones index) followed by reversion.

Loughran and McDonald (2011)

Following Tetlock (2007), popular to use just negative word dictionary from Harvard IV-4.

This includes words like ‘tax’, ‘cost’, ‘capital’, ‘liability’, and ‘vice’.

Unclear that these are appropriate for describing negative content in financial context.

Loughran and McDonald (2011) use 10-K filings to define their own finance-specific word lists, available from `http://www3.nd.edu/~mcdonald/Word_Lists.html`.

Negative list includes words like ‘restated’, ‘litigation’, ‘termination’, ‘unpaid’, ‘investigation’, etc.

Loughran and McDonald (2011)

Following Tetlock (2007), popular to use just negative word dictionary from Harvard IV-4.

This includes words like 'tax', 'cost', 'capital', 'liability', and 'vice'.

Unclear that these are appropriate for describing negative content in financial context.

Loughran and McDonald (2011) use 10-K filings to define their own finance-specific word lists, available from `http://www3.nd.edu/~mcdonald/Word_Lists.html`.

Negative list includes words like 'restated', 'litigation', 'termination', 'unpaid', 'investigation', etc.

Main result: the context-specific list has greater predictive power for return regressions than the generic one.

Social Media Data

Social media data is another data source for measuring sentiment.

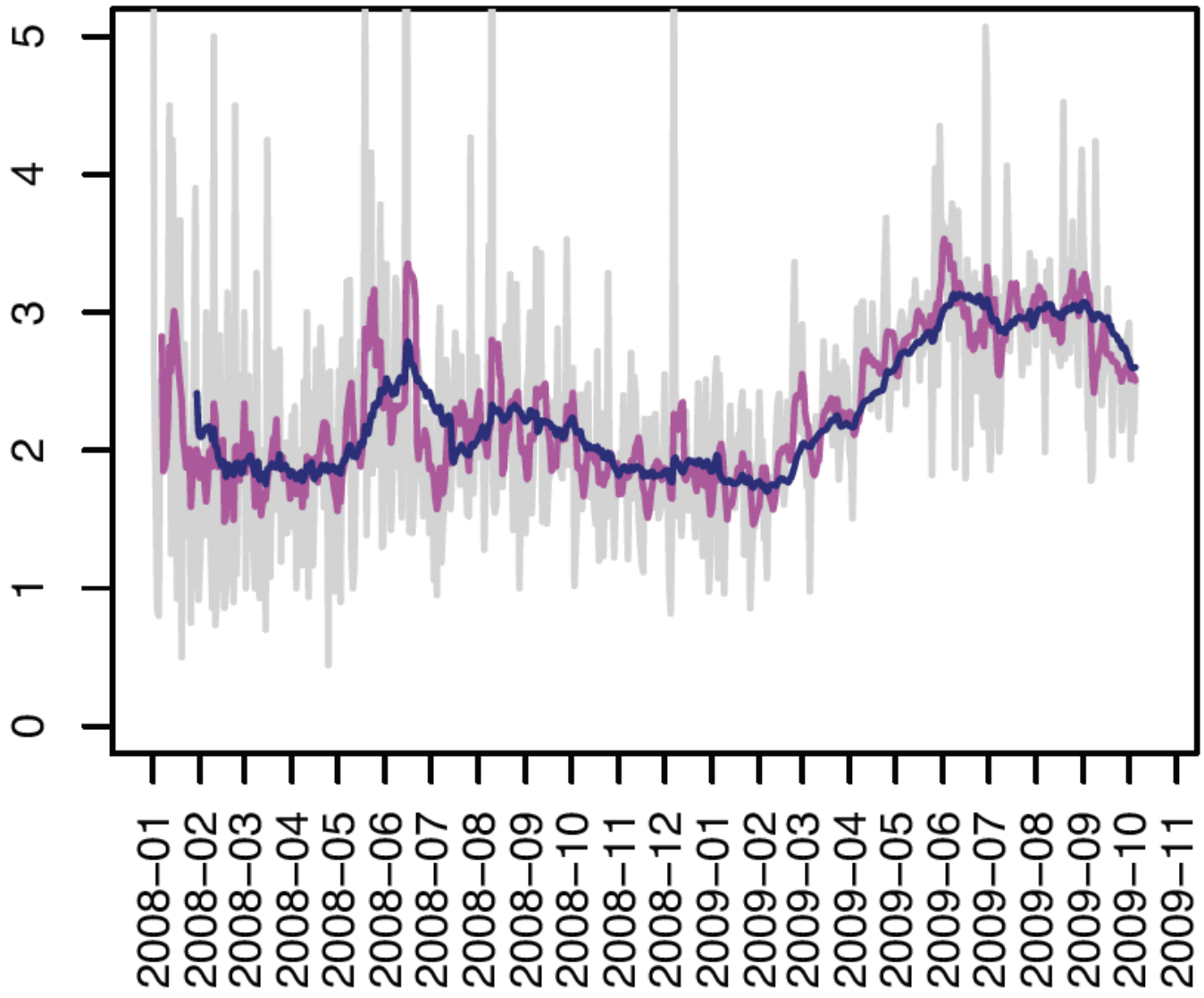
O'Connor et. al. (2010) use Twitter data to track consumer confidence, as measured by the US Index of Consumer Sentiment.

Two challenges: (1) identify relevant tweets; (2) measure sentiment within relevant tweets.

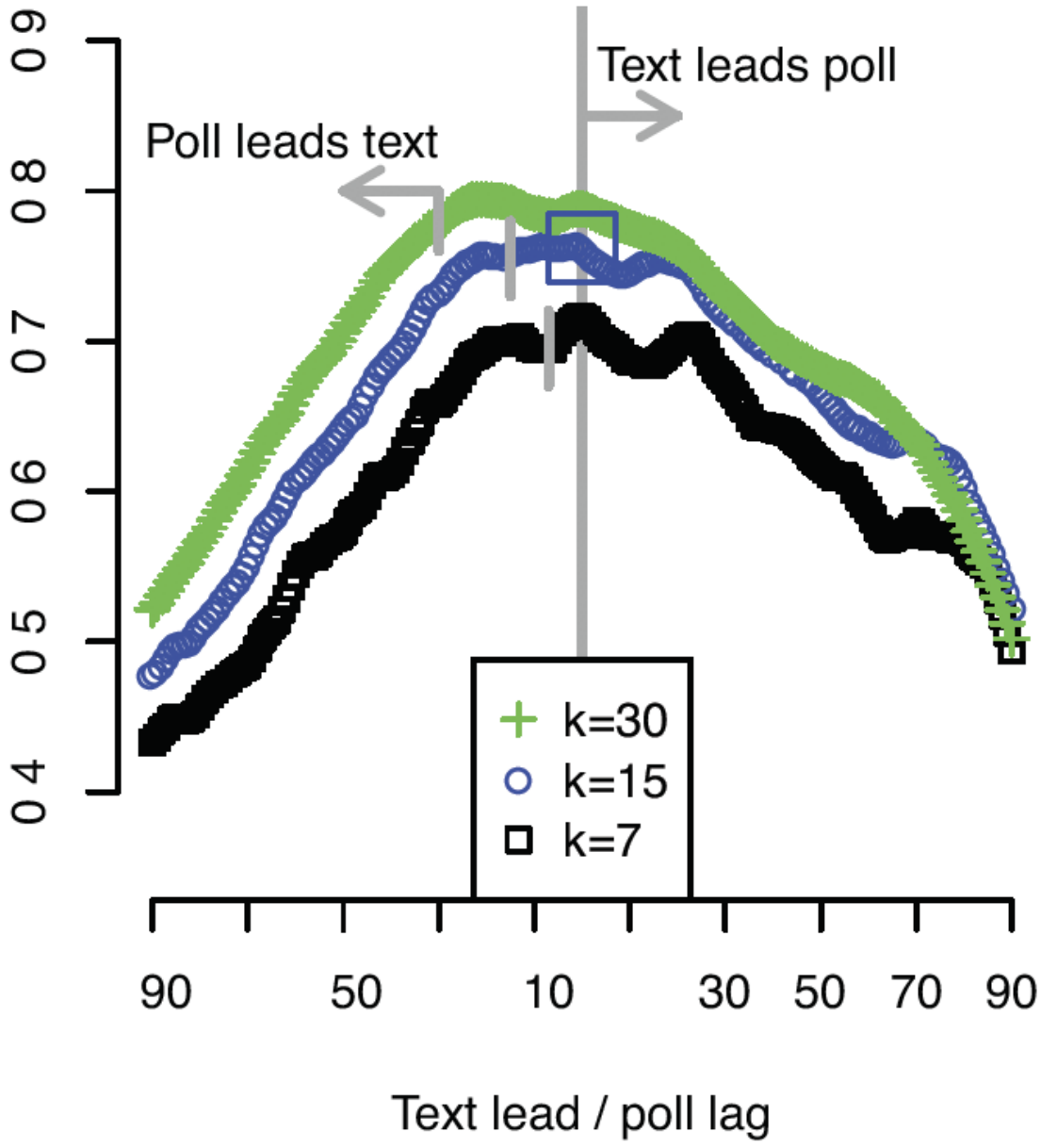
For (1), use all tweets that contain word 'economy', 'job', and 'jobs'.

For (2), use positive and negative words from OpinionFinder. Tweet is positive (negative) if it contains any positive (negative) word; day t sentiment score is ratio of positive to negative messages.

Sentiment Index (Daily, Weekly and Monthly Smoothing)



Correlation with ICS



Theoretically Grounded Dictionaries

Nyman et. al. (2018) use dictionaries grounded in psychological theory to characterize emotional states that ground people's actions.

The index is applied to three different sources of text:

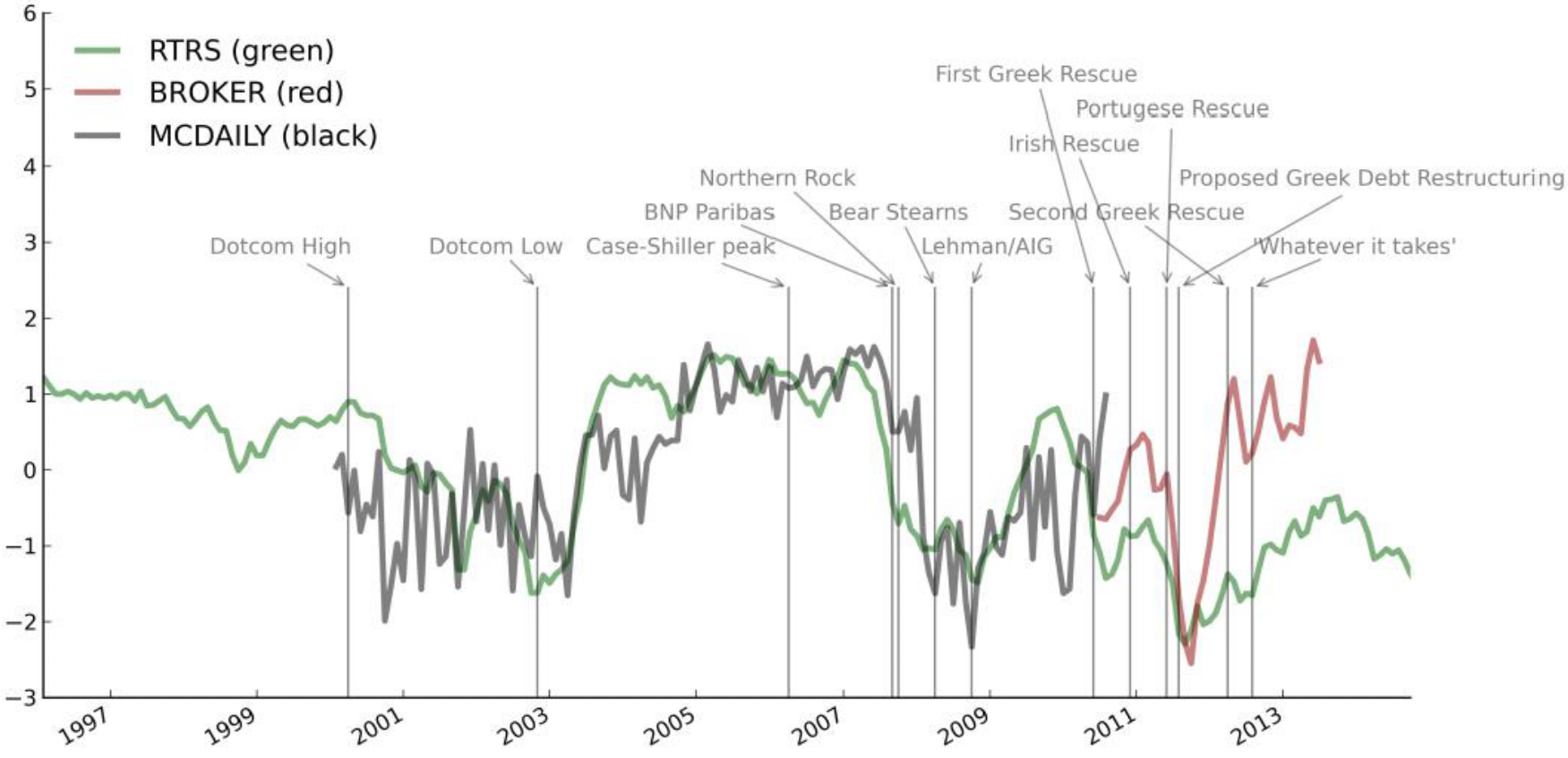
1. Bank of England market commentary.
2. Broker reports.
3. Reuters news archive.

Dictionary

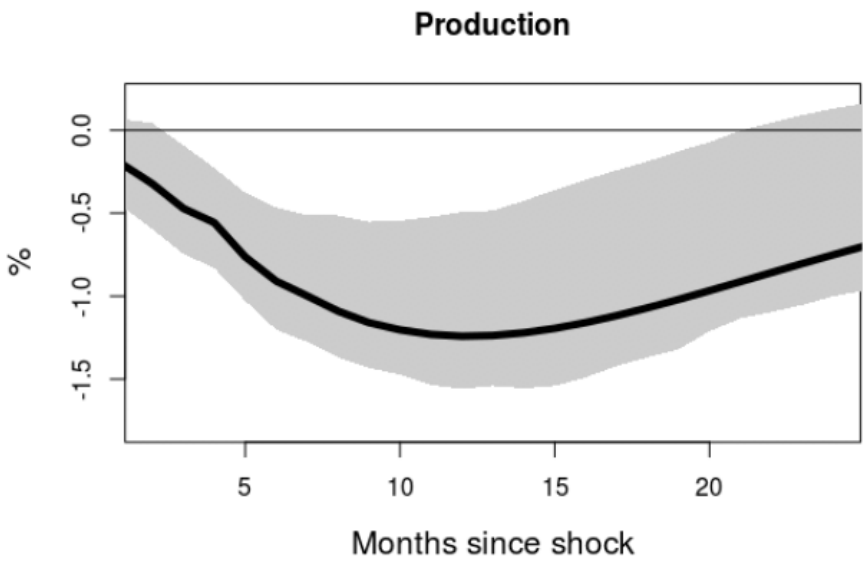
Table 1: Emotion dictionary samples

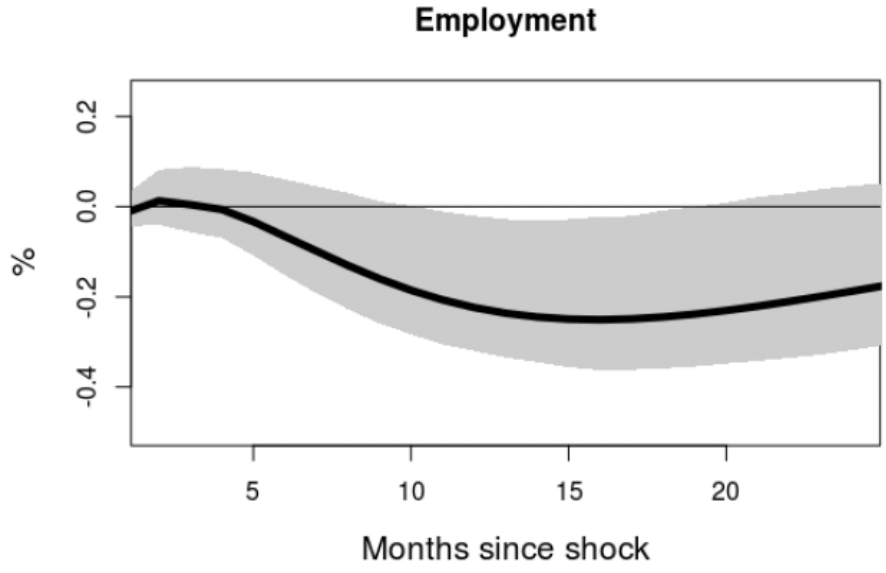
Anxiety		Excitement	
Jitter	Terrors	Excited	Excels
Threatening	Worries	Incredible	Impressively
Distrusted	Panics	Ideal	Encouraging
Jeopardized	Eroding	Attract	Impress

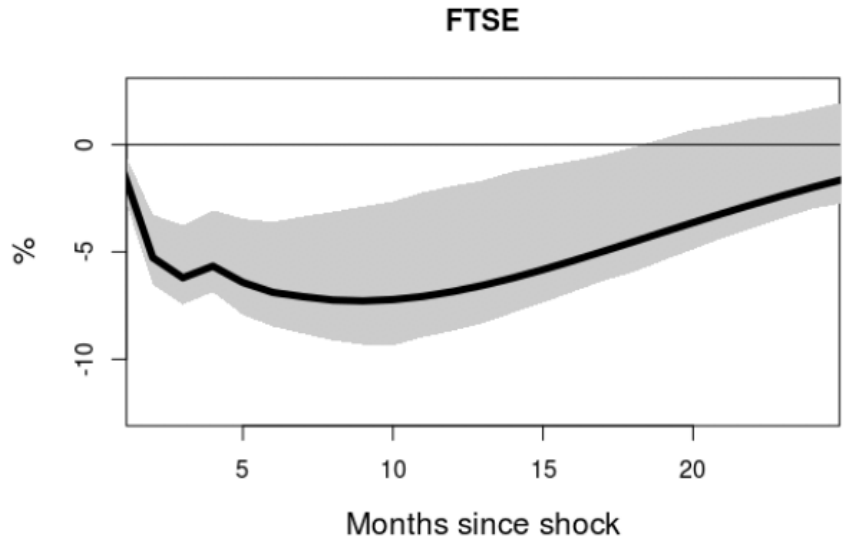
Index


RTRS (green)
BROKER (red)
MCDAILY (black)
Dotcom High
Dotcom Low
Case-Shiller peak
BNP Paribas
Northern Rock
Bear Stearns
Lehman/AIG
First Greek Rescue
Irish Rescue
Portugese Rescue
Second Greek Rescue
Proposed Greek Debt Restructuring
'Whatever it takes'
6
5
4
3
2
1
0
−1
−2
−3
1997
1999
2001
2003
2005
2007
2009
2011
2013

VAR Results


Production
%
0.0
-0.5
-1.0
-1.5
5
10
15
20
Months since shock


Employment
%
0.2
0.0
-0.2
-0.4
5
10
15
20
Months since shock


FTSE
%
0
-5
-10
5
10
15
20
Months since shock

Dictionary Methods + LDA

Sometimes the meaning of a dictionary can vary depending on the topic it discusses.

One can combine dictionary methods with the output of LDA to weight words counts by topic.

Recent application to minutes of the Federal Reserve to extract index of economic situation and forward guidance.

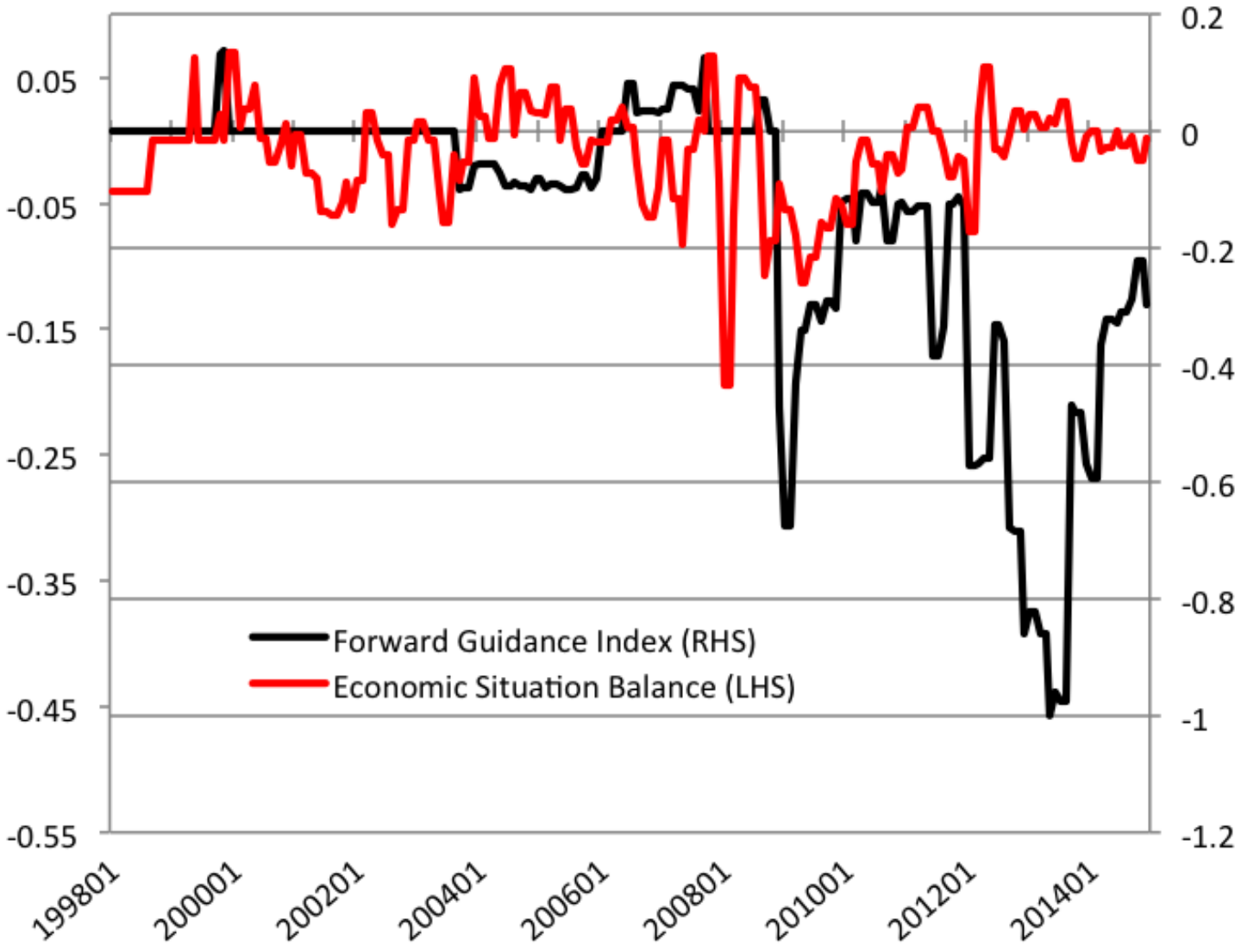
We run 15-topic model and identify two separate kinds of topic.

Monetary Measures of Tone from Apel/Blix-Grimaldi

Contraction	**Expansion**
decreas*	increas*
decelerat*	accelerat*
slow*	fast*
weak*	strong*
low*	high*
loss*	gain*
contract*	expand*

Example Topic

Indices


0.05
-0.05
-0.15
-0.25
-0.35
-0.45
-0.55
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
199801
200001
200201
200401
200601
200801
201001
201201
201401
Forward Guidance Index (RHS)
Economic Situation Balance (LHS)

Supervised Learning

One advantage of supervised learning over dictionary methods is that they are targeted directly at maximizing predictive accuracy, which is often what we care most about.

Suppose we have text features for document d represented as $\mathbf{x}_d$ along with an associated sentiment variable y_d.

Two options for supervised machine learning:

1. Discriminative classifier that models $p(\, y_d \mid \mathbf{x}_d \,)$: e.g. LASSO, ridge regression.
2. Generative classifier that models the full joint distribution $p(\, y_d, \mathbf{x}_d \,)$: e.g. Naive Bayes, supervised LDA.

Generative classifiers have a higher asymptotic error than discriminative (Efron 1975) but can achieve their error faster (Ng and Jordan 2001).

Feature Selection for Discriminative Classifier

One question is how to represent text: can use unigram, bigram, trigrams counts (and even more complex structures as in Shapiro et. al. 2018).

One can also apply a dimensionality-reduction algorithm to map $\mathbf{x}_d$ into a K-dimensional latent space, and use these as the features.

This technique is related to principal components regression, and is particularly appropriate when terms are highly correlated.

Can also use non-labeled texts along with labeled texts in topic modeling, since LDA uses no information from labels in estimation of topic shares.

Blei et. al. (2003) show that topic share representation is competitive with raw counts in classification.

Example

In recent work with Michael McMahon and Matthew Tong, we study the impact of the release of the Bank of England's Inflation Report on bond price changes at different maturities.

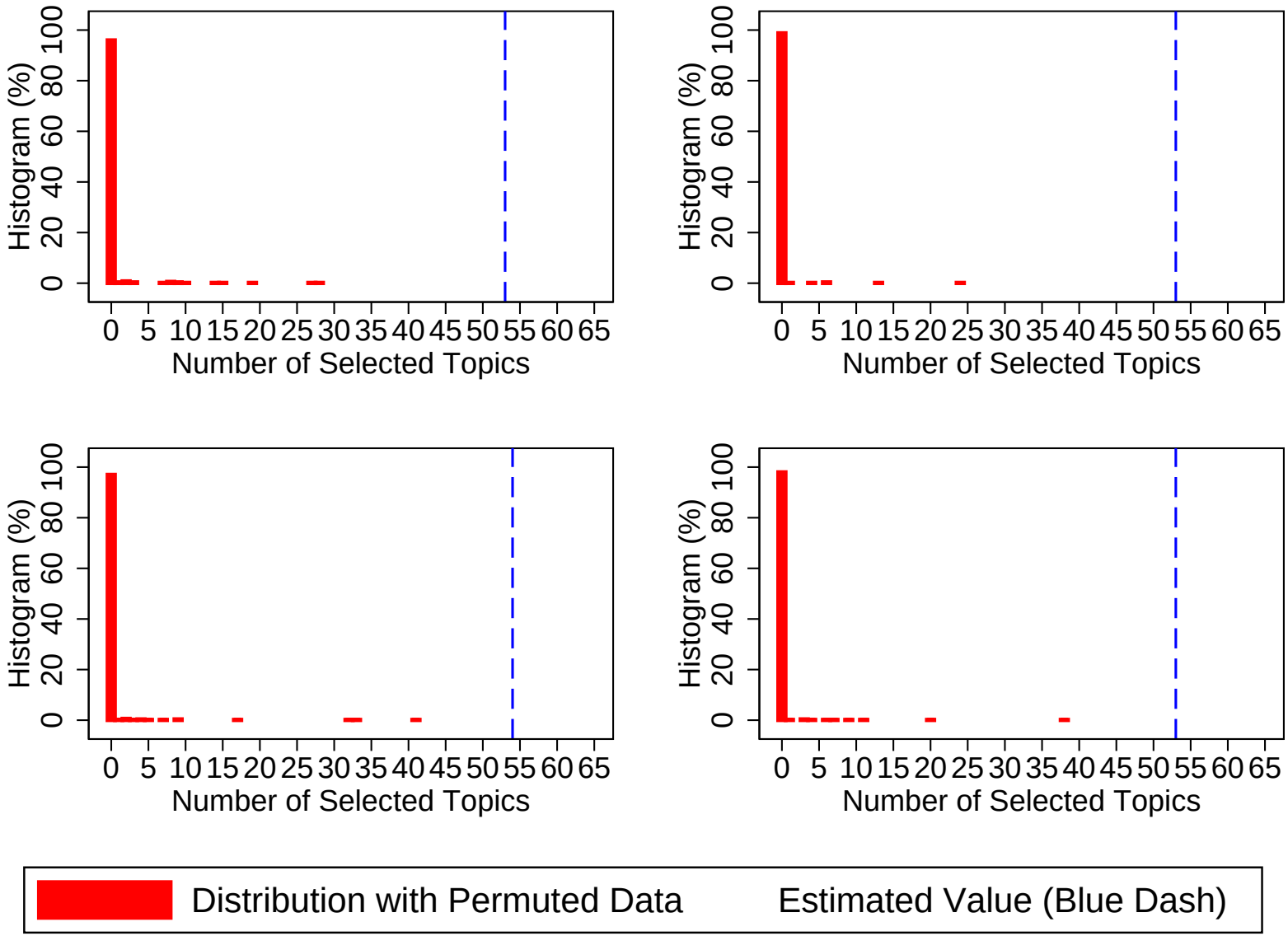
IR contains forecast variables we use as controls: (i) mode, variance, and skewness of inflation and GDP forecasts; (ii) their difference from the previous forecast.

To represent text, we estimate a 30-topic model and represent each IR in terms of (i) topic shares and (ii) evolution of topic shares from previous IR.

First step in the analysis is to partial out the forecast variables from bond price moves and topic shares by constructing residuals.

We are then left with 69 bond price moves (number of IRs in the data) and 60 text features.

LASSO-based Test of Information Content of Narrative



Which Information Matters?

LASSO selects dozens of features at all maturities: standard over-selection problem (Meinshausen and Bühlmann 2006, Annals).

How to identify key topics?

We apply a non-parametric bootstrap to simulate the "inclusion probabilities" of topic features at different maturities.

Draw with replacement from our 69 observations to obtain new sample, perform LASSO, and record whether each feature is included.

Repeat 500 times, and rank topics according to the fraction of bootstrap draws in which they appear.

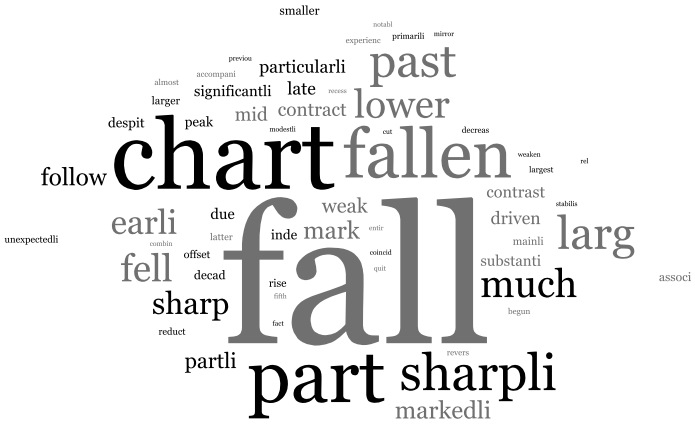
Results: Top Topic

Top Topics for Different Yields (L=Level; D=Change)

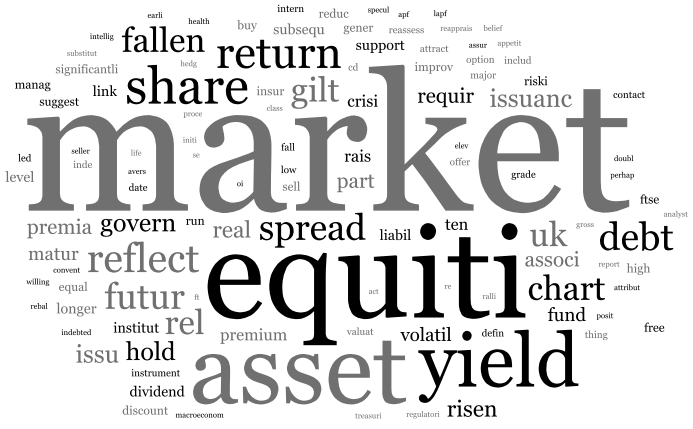
$\lvert\Delta i_{0:12,t}\rvert$		$\lvert\Delta f_{36,t}\rvert$		$\lvert\Delta f_{60,t}\rvert$		$\lvert\Delta f_{60:120,t}\rvert$	
Var	Selection %	Var	Selection %	Var	Selection %	Var	Selection %
L25	0.958	D24	0.858	L28	0.876	D17	0.91
D24	0.954	D25	0.844	D17	0.784	D18	0.896
L5	0.932	L28	0.826	D18	0.772	L20	0.836
L26	0.91	D14	0.76	L20	0.722	D13	0.808

Results: Top Topics

1-Year Spot Rate



(a) L25



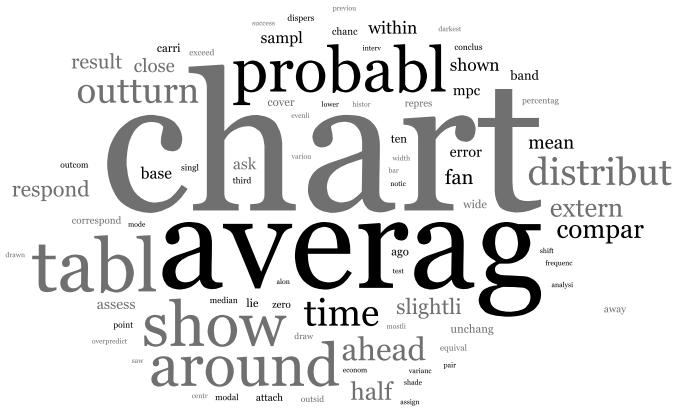
(b) D24

(c) L5

(d) L26

Results: Top Topics

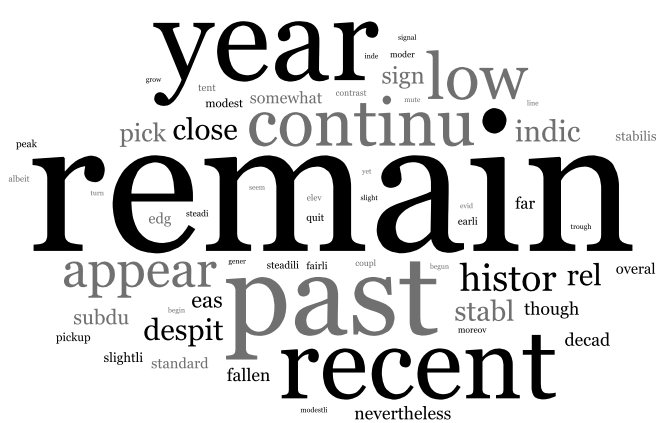
5-Year, 5-Year Forward Rate



(a) D17

(b) D18

(c) L20



(d) D13

Monetary Shocks & Fed Statements

We examine the relationship between Fed statements and the direction of the Romer and Romer (2004) shocks.

To do so, we compute all unique two- and three-word phrases in Fed statements (bigrams/trigrams), and count their frequency in each documents.

Let x_v^- (x_v^+) be the count of term v among statements associate with negative (positive) shocks.

We rank terms according to their informativeness by $\log(x_v^-) - \log(x_v^+)$, and select the top 1,000.

Most Informative Terms—Statements

negative shock	**positive shock**
lower.target.feder	rais.target.feder
lower.target	rais.target
committe.continu	increas.discount.rate
continu.believ	increas.discount
committe.continu.believ	rise.energi
basi.point.reduct	point.increas.discount
point.reduct	point.increas
committe.decid	basi.point.increas
today.lower.target	action.stanc.monetari
today.lower	growth.price

Naive Bayes for Text

We can represent each statement as a length-1000 vector $\mathbf{x}_t$, where $x_{t,v}$ is the count of term v in the time t statement.

Let $RR_t \in \{-, +\}$ represent the direction of the shock in period t.

Suppose that term v appears with probability β_v^y when the shock is y, and that shock y occurs with probability ρ_y. The log-likelihood of observing the data is then

$$\sum_t \mathbb{1}(RR_t = y) \log(\rho_y) + \sum_t \sum_v \mathbb{1}(RR_t = y) x_{t,v} \log\left(\beta_v^y\right).$$

Maximum likelihood estimation gives

$$\widehat{\rho}_y = \frac{N_y}{N} \text{ and } \widehat{\beta}_v^y = \frac{x_v^y}{\sum_v x_v^y}.$$

Classification

One can use the MLE estimates to associate out-of-sample document $\mathbf{x}_d$ with label y_d. By Bayes' Rule we have

$$\Pr[\, y_d = y \mid \mathbf{x}_d \,] \propto \Pr[\, \mathbf{x}_d \mid y_d = y \,] \Pr[\, y_d = y \,]$$

and we can select

$$y_d = \arg\max_y \; \log(\widehat{\rho}_y) + \sum_v x_{d,v} \log\left(\widehat{\beta}_v^y\right).$$

Classification

One can use the MLE estimates to associate out-of-sample document $\mathbf{x}_d$ with label y_d. By Bayes' Rule we have

$$\Pr[\, y_d = y \mid \mathbf{x}_d \,] \propto \Pr[\, \mathbf{x}_d \mid y_d = y \,] \Pr[\, y_d = y \,]$$

and we can select

$$y_d = \arg\max_y \; \log(\widehat{\rho}_y) + \sum_v x_{d,v} \log\left(\widehat{\beta}_v^y\right).$$

To evaluate the quality of the classification, a standard exercise is to:

1. Draw some fraction (one half in our case) of the data, and estimate parameters on it.
2. Use the estimates to classify the held-out documents.
3. Compare the predicted and actual labels.

We perform this exercise using 1,000 random draws for the training set.

Classification Results—Statements

	predicted	
actual	0	1
0	17.706	2.658
1	6.911	13.725

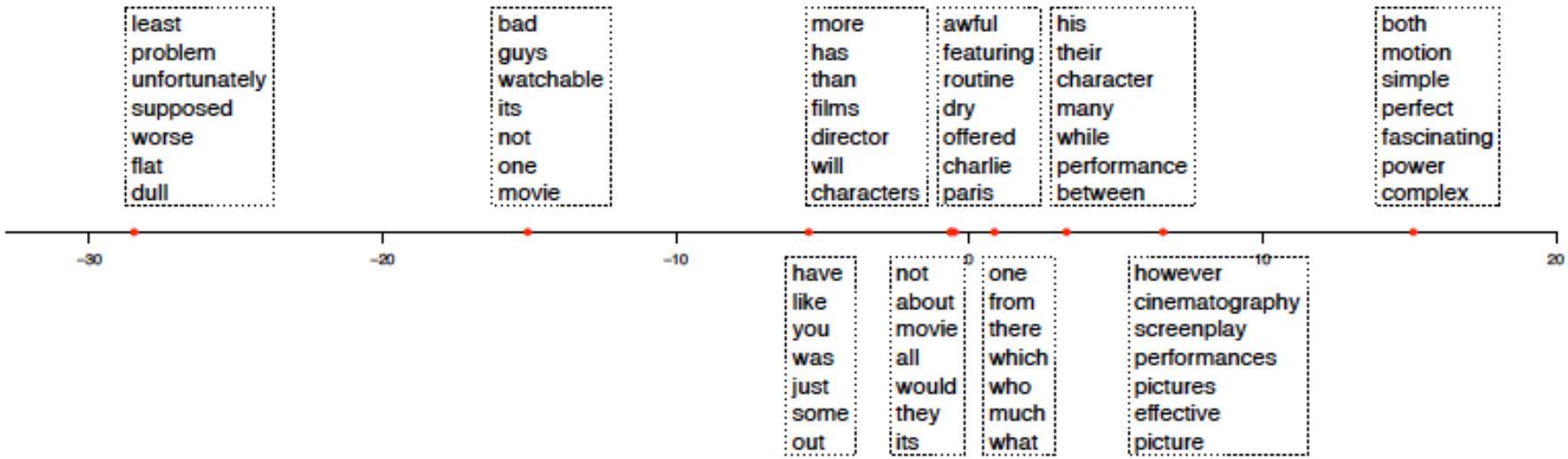
76% average classification accuracy, but asymmetry across shock values.

Supervised LDA (Blei and McAuliffe)

1. Draw $\boldsymbol{\theta}_d$ independently for $d = 1, \ldots, D$ from Dirichlet($\boldsymbol{\alpha}$).
2. Each word $w_{d,n}$ in document d is generated from a two-step process:
   - 2.1 Draw topic assignment $z_{d,n}$ from $\boldsymbol{\theta}_d$.
   - 2.2 Draw $w_{d,n}$ from $\boldsymbol{\beta}_{z_{d,n}}$.
3. Draw y_d from $\mathcal{N}(\boldsymbol{\phi}^T \bar{\mathbf{z}}_d, \sigma^2)$ where $\bar{\mathbf{z}}_d = (n_{d,1}/N_d, \ldots, n_{d,K}/N_d)$ and $z_{d,k}$ is the number of allocations to topic k in document d.

Essentially plain LDA with a linear regression linking topic allocations with observed variables.

Example of Supervised LDA with Movie Review Data



Conclusion

Sentiment analysis is an example of a broader problem with big data: associate a label to a document based on its content.

Dictionaries allow the researcher to control the content that guides the classification, but supervised learning should generally perform better for classification accuracy.

When there are relatively few documents, generative models can perform very well even if they are more complex to estimate.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

A robust machine learning approach for credit risk analysis of large loan-level datasets using deep learning and extreme gradient boosting[1]

Anastasios Petropoulos, Vasilis Siakoulis,
Evaggelos Stavroulakis and Aristotelis Klamargias,

Bank of Greece

[1] This paper was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting

Anastasios Petropoulos

Vasilis Siakoulis

Evaggelos Stavroulakis

Aristotelis Klamargias

Abstract

In the aftermath of global financial crisis of 2007–2008, central banks have put forward data statistics initiatives in order to boost their supervisory and monetary policy functions which will lead to central banks possessing big databases increasing the need for robust data mining processes and financial statistical modelling to support more informed decision making. Conventional econometric methods fail to capture efficiently the information contained in the full spectrum of the datasets. To address these challenges, in this work we investigate the analysis of a corporate credit loans big dataset using cutting edge machine learning techniques and deep learning neural networks.

The novelty of our approach lies in the combination of a data mining algorithms that aim to reduce dimensionality in the data and increase accuracy in predicting the future behaviour of corporate loans, to facilitate a more effective micro and macro supervision of credit risk in the Greek banking system. Our analysis is based on a large dataset of loan level data, spanning a 10 year period of the Greek economy with the purpose of performing obligor credit quality classification and quantification of Probability of Default under a through the cycle setup.

We perform extensive comparisons of the classification and forecasting accuracy of the proposed methods, using a 3-years' period out-of-time sample. Our experimental results are benchmarked against other traditional methods, like logistic regression and discriminant analysis methods, yielding significantly superior performance. In the final stage of our analysis, a robust through the cycle financial credit rating is developed which can offer a proactive monitoring mechanism of the credit risk dynamics in a financial system. Finally the methodological framework introduced can support a more in depth analysis of database initiatives like ECB AnaCredit.

Keywords: Credit Risk, Neural Networks, Deep Learning, Extreme Gradient Boosting.

JEL classification: G24, C38, C45, C55

Contents

1. Introduction

In the aftermath of global financial crisis of 2007–2008, central banks have put forward data statistics initiatives in order to boost their supervisory and monetary policy functions. In the coming years central banks will possess big databases increasing the need for robust data mining processes and financial statistical modelling to support more informed decision making. Under this era, central banks should simultaneously enrich their statistical techniques in order to accommodate the increase availability of data, and to exploit all possible dimensions of information collected. Big financial datasets usually pose significant statistical challenges because they are characterized by increased noise, heavy-tailed distributions, nonlinear patterns and temporal dependencies. Conventional econometric methods fail to capture efficiently the information contained in the full spectrum of the datasets. To address these challenges, in this work we focus on the analysis of a corporate credit loans big dataset using cutting edge machine learning techniques, like Extreme Gradient Boosting (XGBoost) and deep learning neural networks (MXNET).

The novelty of our approach lies in the combination of a data mining algorithms that aim to reduce dimensionality in the data and increase accuracy in predicting the future behaviour of corporate loans, to facilitate a more effective micro and macro supervision of credit risk profile in the Greek banking system. Our analysis is based on a large dataset of loan level data, spanning in a 12 year period of the Greek economy. Data are collected by Bank of Greece for statistical and banking supervision activities. The dataset is comprised of more than 200k records of corporate and SME loans of the Greek banking system, with information related to the one-year-ahead delinquency behaviour. Features collected for analysis include companies' historical data of properly selected set of financial ratios, along with historical data of macro variables relevant to the Greek economy. To ameliorate the issue of high dimensionality in the data we used an advanced machine learning algorithm, called Boruta, to perform the variable importance selection in a multivariate holistic approach. Extreme gradient Boosting and Deep neural networks are used for performing obligor credit quality classification and quantification of Probability of Default under a through the cycle setup.

We perform extensive comparisons of the classification and forecasting accuracy of the proposed methods, using a 3-years' period out-of-time sample. Our experimental results are benchmarked against other traditional methods, like logistic regression, and discriminant analysis methods, yielding significantly superior performance. Furthermore, it is also found that the performance of deep neural-network models depend on the choice of activation function, the number and structure of the hidden layers, and the inclusion of dropout and batch normalization layers signalling increase flexibility in addressing complex datasets and potential increased classification capabilities. In the final stage of our analysis, a robust through the cycle financial credit rating scale is developed which can accommodate the efficient benchmarking of A-IRB models and offer a proactive mechanism of the credit risk dynamics in a financial system. In addition, it can support top down stress testing exercises offering a more risk sensitive and accurate forecasting framework.

In all the methodological framework introduced can support a more in depth analysis of database initiatives like ECB AnaCredit[1].

2. Literature Review

In the domain of credit risk modelling, more accurate and robust systems to drive expert decisions have been employed in recent years, exploring new statistical techniques especially from the field of machine and deep learning. In the last decades, a plethora of approaches has been developed to address the problem of modelling the credit quality of a company, using both quantitative and qualitative information.

Several studies have explored the utility of probit models (Mizen and Tsoukas, 2012) and linear regression models (Avery, et al., 2004). These models however, suffer from their clear inability to capture non-linear dynamics, which are prevalent in financial ratio data (Petr and Gurný, 2013). Another class of statistical models used for credit rating is hazard rate models. These models extend the time horizon of a rating system, by looking at the probability of default during the life cycle of the examined loan or portfolio (Chava and Jarrow, 2004 & Shumway, 2001).

A Bayesian inference-based analogous to support vector machines (SVMs) (Vapnik, 1998), namely Gaussian processes, has been considered by Huang (2011). A drawback of this approach is its high computational complexity, which is cubic to the number of available data points, combined with the assumption of normally distributed data. Yeh et al. (2012) applied Random Forests (Breiman, 2001) in credit corporate rating determination, Zhao et al, (2015) employed feed forward neural networks in the same domain whereas Petropoulos et al (2016) made use of Student's-t hidden Markov models

Addo et al. (2018) focus on credit risk scoring where they examine the impact of the choice of different machine learning and deep learning models in the identification of defaults of enterprises. They also study the stability of these models relative to a choice of subset of variables selected by the models. More specifically, they build binary classifiers based on machine and deep learning models on real data in predicting loan default probability. The top features from these models are selected and then used for testing the stability of binary classifiers by comparing their performance on separate data. They observe that the tree-based models are more stable than the models based on multilayer artificial neural networks.

Khandani et al. (2010) apply machine learning techniques (generalized classification and regression trees (CART)-like algorithm (Breiman et al., 1984)) to construct nonlinear nonparametric forecasting models of consumer credit risk. They combine customer transactions and credit bureau data from January 2005 to April 2009 for a sample of a major commercial bank's customers; thus, they are able to

[1] https://www.ecb.europa.eu/stats/money/aggregates/anacredit/shared/pdf/explanatorynoteanacreditregulation.en.pdf

construct out-of-sample forecasts that significantly improve the classification rates of credit-card-holder delinquencies and defaults.

Butaru et al. (2016) use account-level credit card data from six major commercial banks from January 2009 to December 2013; they combine consumer tradeline, credit bureau, and macroeconomic variables to predict delinquency, employing C4.5 decision trees, logistic regression and random forests. They find substantial heterogeneity in risk factors, sensitivities, and predictability of delinquency across banks, implying that no single model applies to all six institutions. The results suggest the need for a more customized approached to the supervision and regulation of financial institutions, in which capital ratios, loss reserves, and other parameters are specified individually for each institution according to its credit risk model exposures and forecasts.

Galindo and Tamayo (2000) test CART decision-tree models on mortgage-loan data to detect defaults. They also compare their results to the Neural Networks (ANN), the k-nearest neighbor (KNN) and probit models, showing that CART decision-tree models provide the best estimation. Huang et al. (2004) provides a survey of corporate credit rating models showing that Artificial Intelligence (AI) methods achieve better performance than traditional statistical methods. The article introduces a relatively new machine learning technique, support vector machines (SVM), to the problem in attempt to provide a model with better explanatory power. They used backpropagation neural network (BNN) as a benchmark and obtained prediction accuracy around 80% for both BNN and SVM methods for the United States and Taiwan markets.

Motivated from all the aforementioned research endeavours we revisit the issue of credit risk modelling following a different venue. We explore two state of the art techniques namely Extreme Gradient Boosting (XGBoost) and deep learning neural networks in order to obtain at first maximum information gain from a loan level large size data source and secondly to create a useful, from a regulatory scope, credit rating grade system measuring credit risk in supervised banks portfolios.

3. Data collection processing and variable selection

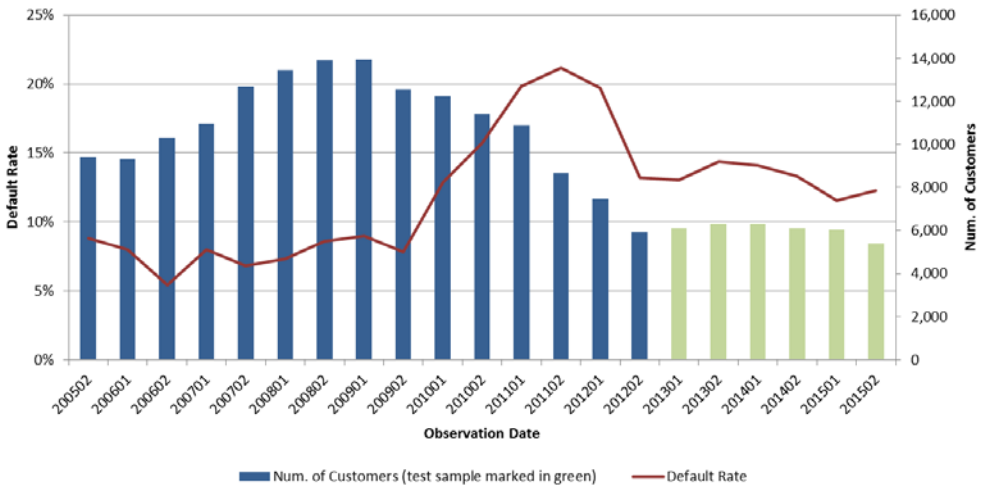
We have collected loan level information on Corporate and SME loans of the Greek banking system, from the supervisory database of the Central Bank of Greece. The data collection procedure excludes special cases of obligors from the financial sector, including banks, insurance, leasing, and factoring companies, due to the very unique nature of their business models, which deviate quite a lot from the business models of commercial companies.

The adopted definition of a default event in this dataset is in line with the rules of the Credit Risk Regulation (CRR). Specifically, a loan is flagged as delinquent if it is either 90 days past due or it gets rated as delinquent based on each bank's internal rating rules. At each observation snapshot, all performing loans are considered. At the end of the 12-month observation period, each obligor is categorized as either good (i.e., performing) or bad (i.e., non-performing). At the end the dependent variable in our dataset is a binary indicator, with the value of

one flagging a default event (i.e., the obligor is categorized as bad at the end of the 12-month observation period).

The dataset covers the 2005-2015 period; a 10 years' period with semi-annual information (i.e. semi-annual snapshots). The selected time period, seems to approximate a full economic cycle, in terms of the default rate evolution. Figure 1, shows the number of customers included in each snapshot and the corresponding default rate. The overall dataset includes approximately 200.000 unique customers, resulting in even more records on a facility level, as one customer may have more than one facility in one or more than one banks with different risk characteristics (for example the average facility number in the credit risk supervisory database reaches 120.000 records per quarter). It is clear that the default rates have elevated in the most recent period, i.e. from the second half of 2010 and onwards, compared to the older observations, i.e. up to 2010. Specifically, the default rates follow an increasing trend in the 2010-2011 periods, where they peak at 21.2% in the second half of third quarter of 2011. Thereafter, they follow a decreasing trend. The default rates seem to have flattened out since 2013, remaining stable at around 12%-13%.

Figure 1: Greek banking system business portfolio metric evolution



In order to perform the modelling and prediction methodology, our approach incorporates the companies' 5 year lagged historical data of properly selected set of financial ratios along with 10 quarters lagged historical macro variables relevant to the Greek economy (both shown analytically in the Appendix). This is based on the assumption that financial ratios carry all the information necessary to describe and predict the internal state of a company, providing adequate insights on how profitable an examined company is, what the trends are.

The combined dataset of lagged financial ratios and macro variables along with some data transformations, led to a set of 354 predictor variables (distinct time-series) as potential candidates for our modelling procedures. Fitting a machine learning model to such a huge number of independent variables (relative to the size of the dataset) is doomed to suffer from the so-called curse of dimensionality problem, whereby the fitted classifier may seem to yield very good performance in the training dataset, but it turns out to generalize very poorly, yielding a

catastrophically low performance outcome in the test data. Thus, to ensure a good performance outcome for our model, we need to implement a robust independent variable (feature) selection stage, so as to limit the number of used features to the absolutely necessary. Besides, apart from increasing the generalization capabilities of the fitted models, such a reduction is also important for increasing the computational efficiency of the explored machine learning algorithms.

We employ the Boruta algorithm to independently assign importance to the available features. The Boruta algorithm is based on a postulated Random Forest model. Based on the inferences of this Random Forest, features are removed from the training set, and model training is performed anew. Boruta infers the importance of each independent variable (feature) in the obtained predictive outcomes by creating shadow features. Specifically, the algorithm performs the following steps: First, it adds randomness to the given dataset by creating shuffled copies of all features (shadow features). Then, it fits a Random Forest on the extended dataset and evaluates the importance of each feature. In every iteration, it checks whether a real feature has a higher importance than the best of its shadow features, and constantly removes features which are deemed unimportant. The comparison is done based on Z score. The algorithm stops when all features are classified as important or are rejected as noise. In our study, we employ the Boruta Package, provided by the R programming language, to implement variable selection. In this way, all features relevant to both dependent variables are selected based on error minimization for the fitted Random Forest models, in each iterative step of the algorithm. From the Boruta variable selection process 65 variables out of 354 candidates were selected for the moment development alleviating dimensionality issues. The so-obtained dataset was split into three parts:

- An in-sample train dataset, comprising data pertaining to the 70% of the examined companies, obtained over the observation period 2005-2012, which was used for model development.
- An in-sample test dataset, comprising the data pertaining to the rest 30% of the companies for the period 2005-2012 which was employed for assessing the parameter calibration.
- An out-of-time dataset that comprises all the data pertaining to the observation period of year 2013-2015 (marked in green in Figure 1) which was employed for validation purpose and testing the generalization capacity of all candidate models.

4. Model Development

Given the extended number of employed predictors and the large scale dataset employed we resort to a methodology from the general domain of Machine Learning techniques called Extreme Gradient Boosting (henceforth XGBoost) and a Deep Learning Technique used to train, and deploy deep neural networks (MXNET). The supervisory motivation for employing such types of methodologies rests on the availability of large scale supervisory data, which are expected to further augment in the near future (e.g. ECB's AnaCredit project), upon which the capability of pattern

detection by traditional statistical methodologies is limited due to multicollinearity, dimensionality and convergence issues.

The XGBoost is a boosting tree algorithm that is an enhancement over tree bagging methodologies, such as Random Forests (Breiman 2000), which have gained significant ground and are frequently used in many machine learning applications across various fields of the academic community. The basic philosophy of bagging is based on combining three concepts: i) Creation of multiple datasets; ii) building of multiple trees and iii) bootstrap aggregation or bagging. It adopts a divide-and-conquer approach to capture non-linearities in the data and perform pattern recognition. Its core principle is that a group of "weak learners" combined, can form a "strong predictor" model.

For example in the case of Random Forests the algorithm is based on the random generation of a number of classification trees which is the so called Forest. Tree generation is randomly performed in an iterative mode so in each iteration, a random subsample of the included features is selected from the dataset by means of bootstrap. Then, a tree is generated from using the CART algorithm which contains a relatively limited number of features. After constructing the random trees, prediction is performed using Bagging. Each input is entered through each decision tree in the forest and produces a forecast. Then, all predictions of each tree are aggregated either as a (weighted) average or majority vote, depending whether the underlying problem is a regression or a classification, respectively.

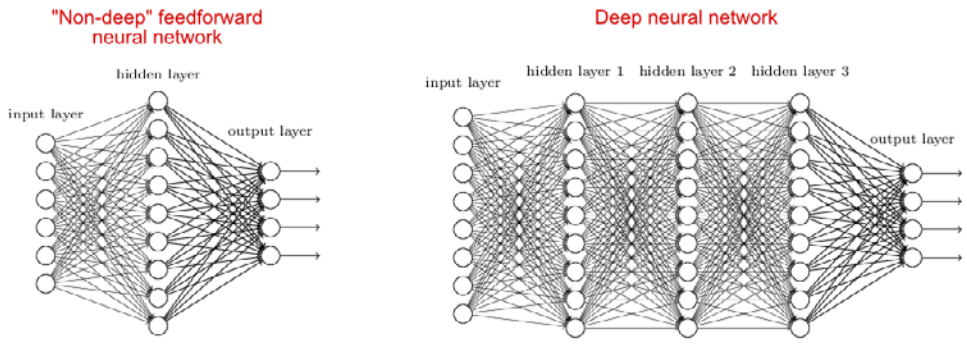
Gradient Boosting trees model is proposed by Friedman (1999) and has the advantage of reducing both variance and bias. It reduces variance because multiple models are used (bagging), whereas it additionally reduces bias in training the subsequent model by telling him what errors the previous models made (boosting). In gradient boosting each subsequent model is trained using the residuals (the difference between the predicted and true values) of previous models. XGBoost (Extreme Gradient Boosting) is an advanced implementation of gradient boosting algorithm, offering increased efficiency, accuracy and scalability over simple bagging algorithms. It supports fitting various kinds of objective functions, including regression, classification and ranking. XGBoost offers increased flexibility, since optimization is performed on an extended set of hyperparameters, while it fully supports online training.

We developed XGBoost in the context of our study by utilizing the XGBoost R package. We performed an extensive cross-validation procedure to select a series of entailed hyper parameters, including the maximum depth of trees generated, the minimum leaf nodes size to perform a split, and the size of sub-sampling for building the classification trees and the variables considered in each split. The objective function used for the current problem was logistic due to the binary nature of the dependent variable while the area under the curve (AUROC) metric was used for model selection in the context of cross-validation. The AUROC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. In practice, the value of AUROC varies between 0.5 and 1, with a value above 0.8 denoting a very good performance of the algorithm. To reduce overfitting tendencies, we tuned the γ hyper parameter, which controls model complexity by imposing the requirements that node splits should yield a minimum reduction in the loss function, as well as the α and λ hyper parameters, which perform regularization of model weights similar to shrinkage techniques such as LASSO.

Besides Extreme Gradient Boosting we implement also a Deep Neural Network (henceforth DNN) to address the issue of corporate default forecast. Deep learning has been an active field of research in the recent years, as it has achieved significant breakthroughs in the fields of computer vision and language understanding. In particular they have been extremely successful in as diverse time-series modelling tasks as machine translation (Cho et al., 2014, Tu et al., 2016.), machine summarization (See et al., 2017) and recommendation engines (Quadrana et al., 2017). However, their application in the field of finance is rather limited. Specifically, our paper constitutes one of the first works presented in the literature that considers application of deep learning to address this challenging financial modelling task.

Deep Neural Networks differ from Shallow Neural Networks (one layer) on the multiple internal layers employed between the input values and the predicted result (Figure 2). Constructing a DNN without nonlinear activation functions is impossible, as without these the deep architecture collapses to an equivalent shallow one. Typical choices are logistic sigmoid, hyperbolic tangent and rectified linear unit (ReLU). The logistic sigmoid and hyperbolic tangent activation functions are closely related; both belong to the sigmoid family. A disadvantage of the sigmoid activation function is that it must be kept small due to their tendency to saturate with large positive or negative values. To alleviate this problem, practitioners have derived piecewise linear units like the popular ReLU, which are now the standard choice in deep learning research ReLU, (Vinod & Hinton, 2010).

Figure 2: Shallow and Deep Neural Networks



On a different perspective, since DNNs comprise a huge number of trainable parameters, it is key that appropriate techniques be employed to prevent them from overfitting. Indeed, it is now widely understood that one of the main reasons behind the explosive success and popularity of DNNs consists in the availability of simple, effective, and efficient regularization techniques, developed in the last few years. Dropout has been the first, and, expectably enough, the most popular regularization technique for DNNs (Srivastava et al., 2014). In essence, it consists in randomly dropping different units of the network on each iteration of the training algorithm. This way, only the parameters related to a subset of the network units are trained on each iteration. This ameliorates the associated network overfitting tendency, and it does so in a way that ensures that all network parameters are effectively trained.

Inspired from these merits, we employ Dropout DNNs with ReLU activations to train and deploy feed forward deep neural networks. More precisely we employ the Apache MXNET toolbox of R[2]. We postulated deep networks that are up to five hidden layers deep and comprise various numbers of neurons. Model selection using cross-validation was performed by maximizing the AUROC metric.

We benchmark the abovementioned techniques versus traditional statistical techniques employed in Probability of Default modelling, such as Logistic regression (Logit) and Linear Discriminant Analysis (LDA). Logistic regression is an approach broadly employed for building corporate rating systems and retail scorecards, due to its parsimonious structure. It was first used by Ohlson (1980) to predict corporate bankruptcy based on publicly available financial data. Logistic regression models determine the relative importance of each independent variable in the classification outcome using the fitting dataset. In order to account for non-linearities, and to relax the normality assumption, a sigmoid likelihood function is typically used (Kamstra et al. 2001).

Linear discriminant analysis (LDA) is a method to find a linear combination of features that characterizes or separates two or more classes of objects or events. The main assumptions are that the modelled independent variables are normally distributed and that the groups of modelled objects (e.g. good and bad obligors) exhibit homoscedasticity. LDA is broadly used for credit scoring. For instance, the popular Z-Score algorithm proposed by Altman (1968) is based on LDA to build a rating system for predicting corporate bankruptcies. The normality and homoscedasticity assumptions are hardly ever the case in real-world scenarios, thus, being the main drawbacks of this approach. As such, this method cannot effectively capture nonlinear relationships among the modelled variables, which is crucial for the performance of a credit rating system. We implemented this approach in R using the MASS R package. Before estimating both the logit and the LDA model we dropped collinear variables based on correlation cut-off threshold of 50%.

5. Model Evaluation

Classification accuracy, as measured by the discriminatory power of a rating system, is the main criterion to assess the efficacy of each method and to select the most robust one, in terms of discriminatory power and performance misinterpretation. We tested a series of metrics that are broadly used for quantitatively estimating the discriminatory power of each scoring model, such as the Area Under the ROC curve metric, as well as the Kolmogorov Smirnov (KS) statistic as performance measures.

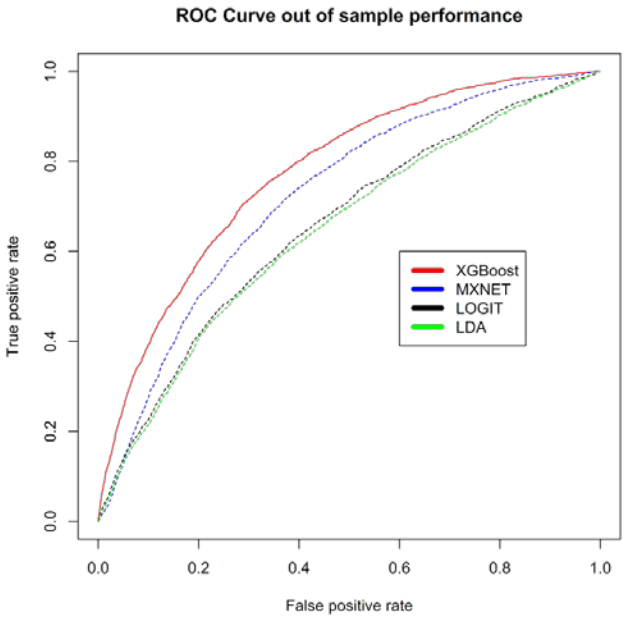
[2] https://mxnet.incubator.apache.org/api/r/index.htm

Classification Accuracy Model Comparison	KS	AUROC	Table 1
Logit	24%	66%	
LDA	23%	65%	
XGBoost	42%	78%	
MXNET	35%	72%	

Classification Accuracy Metrics: Kolmogorov - Smirnov (KS), Area Under ROC curve (AUROC).

Further, we present in Figure 3 the ROC curveS corresponding to the methodologies analysed. This curve is created by plotting the true positive rate against the false positive rate at various threshold settings. As such, they illustrate the obtained trade-offs between sensitivity and specificity, as any increase in sensitivity will be accompanied by a decrease in specificity. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the modelling approach. The corresponding ROC curve of extreme gradient boosting (XGBoost) is higher over all the considered competitors supporting the high degree of efficacy and generalization capacity of the proposed employed machine learning system.

Figure 3: ROC curve for forecasting a default event on 1 year horizon

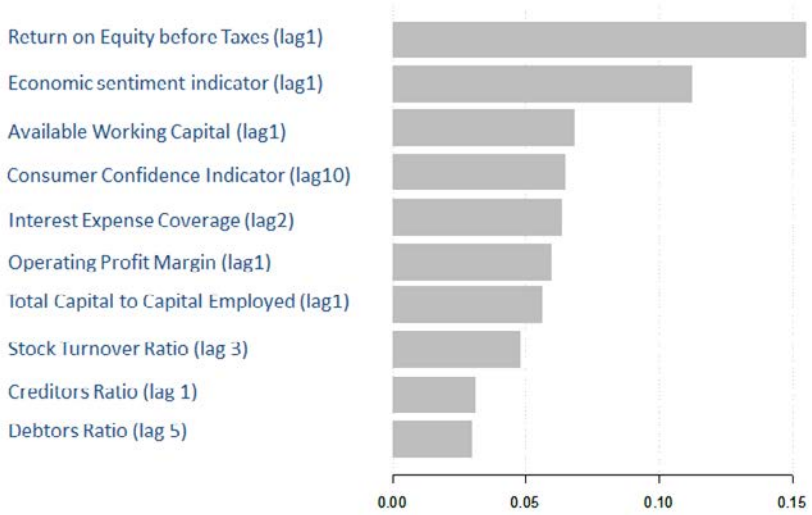


From Table 1 and Figure 3 we deduce that the XGBoost and MXNET algorithms provide better classification accuracy compared to traditional classification methods such as Logistic Regression and Linear Discriminant analysis. As a robustness check other widely employed classification techniques were employed namely the CART algorithm, Random Forests and One Layer Neural Networks (Shallow) but their performance did not surpass the XGBoost and MXNET which is logical if you

consider that the first two are subcases of XGBoost whereas Shallow Neural Network are subcases of MXNET algorithm.

Boosting and Bagging algorithms, even though they are computation intensive, have the relative advantage that they are not "black boxes" regarding the factors affecting the final result, since they provide a module for calculating variable importance measures through reshuffling. In other words after predicting with the benchmark model the reshuffling technique predicts hundreds of times for each variable in the model while randomizing that variable. If the variable being randomized hurts the model's benchmark score, then it is an important variable. If, on the other hand, nothing changes, then it is a useless variable. We run the variable importance algorithm and we show in Figure 4 the ranked list of first ten more important variables

Figure 4: XGBoost variable importance plot. The x-axis describes the percentage contribution of the predictor in the "real" model.



It appears that the most important financial ratio predictor for the default probability of a company is Return on Equity followed by the availability of working capital and Interest Expense Coverage. In essence the company return, the availability of financial resources and the prudent leverage policy may assure the viability of a business. In addition the economic climate, seem to play an additional important role in business viability since the Economic sentiment indicator and the Consumer confidence indicator are rendered important in the model whereas other widely employed factors such as GDP growth seem not to be predominant. What is important is that XGBoost includes both macro variables and financial ratios capturing both the systemic and idiosyncratic behaviour in obligor's credit quality, thus both discriminatory and calibration test exhibit stability and steady performance.

6. Rating System Calibration

An essential aspect of each classification system lies in the creation of a way to represent the classification results to a rating system which can be employed for supervisory purposes in the course central banking operations. For this purpose, we apply a credit rating system calibration process. Calibration of a credit rating system is a mapping process under which each score value is matched to rating grade, which is then associated with a probability of default. To perform the calibration of our systems, the development sample population of each scoring model was split into groups. Specifically, 50 groups (i.e. ranges of scores) were created of equal size, each one including 2% of the total population. Each group is associated with the default rate observed in the development sample.

When necessary, ranges of scores were grouped together, in order to ensure monotonicity of the obtained default rates, maximum intra-rate homogeneity of the observed default rates, and maximum inter-range heterogeneity. In order to overcome overfitting issues and create a reasonable system, each grade included at least 4% of the development population. Grouping optimization was performed based on the Information Value metric.

The following graphs visualize the calibration performed for each rating system. In specific, the first graph present the default rate associated to each of the 50 groups initially created, while the second graph show the default associated to the final selected grades.

Figure 5: Estimated Default Rate of the Initial Grouping (50 Groups)

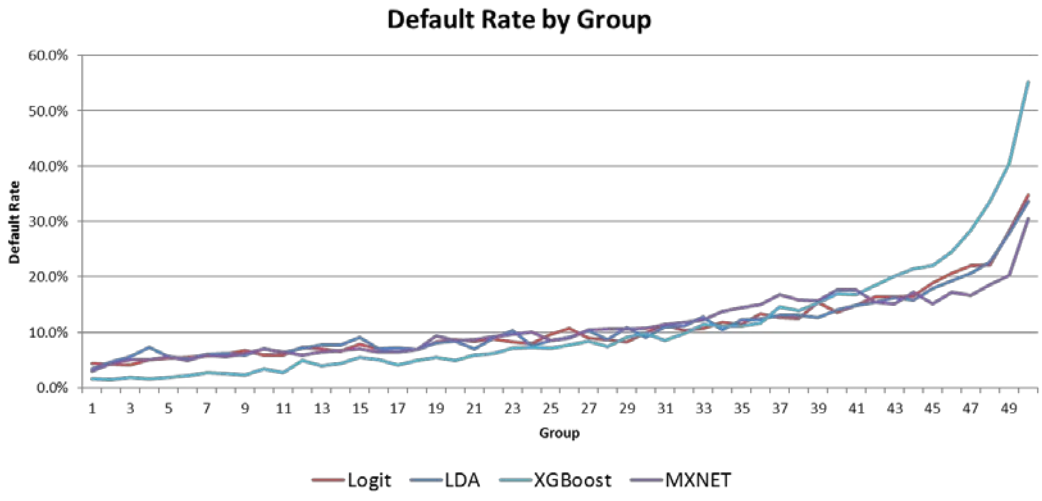
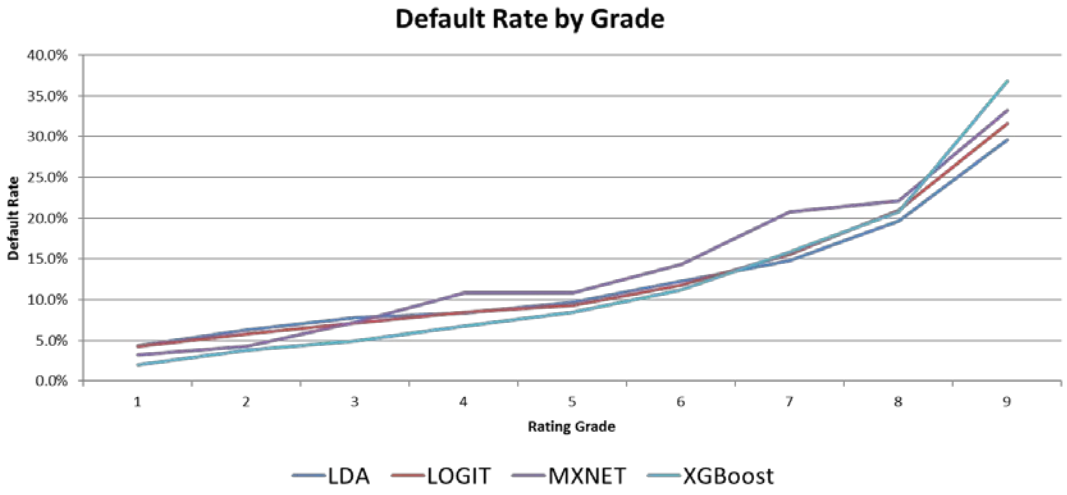


Figure 6: Estimated Default Rate of the Final Selected Grades (9 Grades)



Based on the Figures 5 and 6 it is clear that the XGBoost model is able to produce a more granular calibration. This means that the XGBoost calibration can assign lower default rates to the low grades, and vice versa higher default rates to higher grades, compared to the other models. In order to assess the calibration of the rating systems developed, the binomial test, the normal, the sum of square error, and the brier score validation metrics are utilized. The estimated probability of default was also compared with the out-of-sample observed default rate. The validation methodologies that are typically being applied in the industry include the most, if not all, of the metrics included in our analysis. Additional tests, such as the Bayesian error rate, Chi-square (or Hosmer-Lemeshow) test, that could also have been used, were omitted mainly due to the fact that they produce similar results and conclusions.

Performance Metrics — Table 2

Credit Rating System

	SSE	BRIER
Logit	4.3%	11.3%
LDA	4.8%	11.4%
XGBoost	0.2%	10.1%
MXNET	0.6%	11.0%

Rating System Calibration Metrics: Sum of Square Error (SSE), Brier's score (BRIER).

In the Appendix (Tables 4-7) are shown the calibration results for each evaluated model, i.e. the estimated probability of default per rating grade (based on the default rate on the development sample) and the observed default rate in the out-of-time period. We deduce that Binomial and Normal tests fail for the rating systems developed based on the LDA and Logit models. The estimated PDs are lower than the observed default rates for almost all grades. On the other hand, the MXNET and XGBoost rating systems perform better as the estimated PDs are not statistically different to the observed default rates.

Estimated and Actual default frequency metrics Table 3

	Estimated Probability of Default	Observed Default Rate (Out of sample)	Observed Default Rate (In sample)
Logit	8.20%	13.10%	11.00%
LDA	7.80%		
XGBoost	13.50%		
MXNET	15.00%		

Estimated Probability of Default vs observed Default Rate in out-of-sample and in-sample population

Based on Table 3 it is clear that the rating system developed based on the XGBoost model, is more accurate in terms of PD quantification, compared to the other candidate models due to the more granular calibration achieved by XGBoost. Analytically, XGBoost has marginally overestimated the observed default rate in the validation sample whereas MXNET overestimated the default rate which is good from regulatory perspective. On the other hand, LDA and Logit based systems significantly underestimated the observed default rate.

Deep neural networks provide promising results even though they do not outperform the XGBoost algorithm. The fact is that this methodology provides the opportunity of creating a large combination of different structures based on the number of layers, the selection of activation functions, the number of perceptrons and normalization layers which can be inserted in the optimization process. In the appendix (Figure 7) some illustrative alternative employed structures is shown. Therefore the potentials for Deep Neural Networks algorithms (such as MXNET) in pattern detection in the era of "big data" in which the central banking system is entering are enormous, given that the flexibility of structures is much greater than Boosting and Bagging mechanistic algorithms.

7. Conclusion

In order to tackle the issue of pattern detection in large loan level datasets for extracting information regarding credit risk and exposure credit quality, we employ a combination of data mining algorithms that aim to reduce dimensionality in the data and increase accuracy in predicting the future behaviour of corporate loans. Our analysis is based on a large dataset of loan level data, spanning in a 10 year period of the Greek economy with the purpose of performing obligor credit quality classification and quantification of Probability of Default under a through the cycle setup.

We perform extensive comparisons of the classification and forecasting accuracy of the proposed methods, using a 3-years' period out-of-time sample and we deduce that the Extreme Gradient Boosting technique along with Deep Neural Networks provide better performance in terms of classification accuracy and credit rating system calibration compared to widely employed techniques in credit risk

modelling such as Logistic Regression and Linear Discriminant Analysis. In addition the inclusion of both macro variables and financial ratios captures both the systemic and idiosyncratic behaviour in obligor's credit quality, thus both discriminatory and calibration test exhibit stability and steady performance.

Our findings provide significant oversight for regulatory purposes given that in the coming years, central banks will possess big databases increasing the need for robust data mining processes and financial statistical modelling to support more informed decision making. For example the proposed approaches could find fruitful ground on the European Central Bank's AnaCredit initiative for the collection of loan level data. "Big Data" as referred often entail dimensionality issues, increased noise and other significant statistical challenges which cannot be addressed from traditional statistical techniques.

Regarding the final model selection XGBoost seems to be the methodology marginally outperforming Deep Neural Networks (MXNET) but the latter methodology provides the opportunity of increased flexibility over boosting techniques through a large combination of different structures which may optimize the bias variance trade-off. As a prospect of future research it may be explored whether alternative Deep Neural Network structures, such as recurrent DNN or convolutional networks, may increase the classification accuracy or whether potential forecast combinations among machine and deep learning techniques may further allow boosting of the results.

Appendix

Financial Ratios Employed

- Working Capital
- Employed Capital (Assets minus Current Liabilities)
- Return on Equity before Taxes
- Return on Equity before Interest and Taxes
- Profit before taxes to Employed capital
- Gross Margin to Sales
- Operating Margin to Sales and other income
- Earnings before Interest and Taxes to Sales and other income
- Sales and other income to Employed Capital
- Sales and other income to Equity
- Equity and Long Term Loans to Net Fixed Assets
- Debt to Equity
- Interest Expense Coverage
- Equity to Employed Capital
- Working Capital to Short Term obligations
- Immediate Cash Ratio
- Debtors Ratio
- Creditors Ratio
- Stock turnover Ratio

Macro Variables Employed

- Gross Domestic Product yearly growth
- Investment yearly growth
- Export yearly growth
- Consumption yearly growth
- Economic sentiment indicator
- Consumer Confidence Indicator
- Unemployment Rate
- Inflation
- Stock Market Returns
- Stock Market Volatility
- Deposit Rates
- Loan Rates
- 10 year Government bond spread
- 5 year Government bond spread
- 1 year Government bond spread

Binomial and Normal Validation Tests

Validation Testing - Logistic Regression model — Table 4

Rating Grade	Estimated Probability of Default	Observed Default Rate (Out of sample)	Binomial Test	Normal Test
1	4.21%	6.46%	0.00%	0.00%
2	5.77%	8.83%	0.00%	0.00%
3	7.10%	10.84%	0.00%	0.00%
4	8.38%	13.97%	0.00%	0.00%
5	9.32%	17.46%	0.00%	0.00%
6	11.77%	21.46%	0.00%	0.00%
7	15.53%	23.45%	0.00%	0.00%
8	20.95%	29.64%	0.00%	0.00%
9	31.57%	40.00%	5.07%	4.86%

Binomial and Normal tests examine the null hypothesis that the actual default rate of a credit rating grade is not greater than the forecasted probability of default

Validation Testing - Linear Discriminant Analysis — Table 5

Rating Grade	Estimated Probability of Default	Observed Default Rate (Out of sample)	Binomial Test	Normal Test
1	4.36%	7.37%	0.00%	0.00%
2	6.25%	9.90%	0.00%	0.00%
3	7.76%	12.28%	0.00%	0.00%
4	8.38%	13.80%	0.00%	0.00%
5	9.63%	21.68%	0.00%	0.00%
6	12.19%	20.89%	0.00%	0.00%
7	14.75%	26.01%	0.00%	0.00%
8	19.64%	27.58%	0.00%	0.00%
9	29.65%	30.19%	51.73%	52.55%

Binomial and Normal tests examine the null hypothesis that the actual default rate of a credit rating grade is not greater than the forecasted probability of default

Validation Testing - Extreme Gradient Boosting (XGBoost) Table 6

Rating Grade	Estimated Probability of Default	Observed Default Rate (Out of sample)	Binomial Test	Normal Test
1	1.99%	1.52%	95.15%	94.66%
2	3.78%	2.34%	99.96%	99.92%
3	4.89%	3.88%	96.81%	96.45%
4	6.72%	5.39%	98.28%	98.03%
5	8.41%	6.96%	98.75%	98.58%
6	11.18%	9.75%	99.07%	98.97%
7	15.86%	15.09%	86.36%	86.33%
8	20.82%	22.57%	1.64%	1.57%
9	36.81%	38.43%	5.95%	5.92%

Binomial and Normal tests examine the null hypothesis that the actual default rate of a credit rating grade is not greater than the forecasted probability of default

Validation Testing - MXNET Table 7

Rating Grade	Estimated Probability of Default	Observed Default Rate (Out of sample)	Binomial Test	Normal Test
1	3.2%	0.9%	99.4%	98.5%
2	4.2%	3.5%	80.9%	81.1%
3	7.2%	2.6%	100.0%	100.0%
4	10.8%	6.4%	100.0%	100.0%
5	10.8%	15.6%	32.0%	32.1%
6	14.3%	23.1%	8.2%	8.1%
7	20.8%	28.9%	61.5%	61.9%
8	22.1%	30.2%	90.5%	90.5%
9	33.2%	32.9%	56.6%	56.8%

Binomial and Normal tests examine the null hypothesis that the actual default rate of a credit rating grade is not greater than the forecasted probability of default

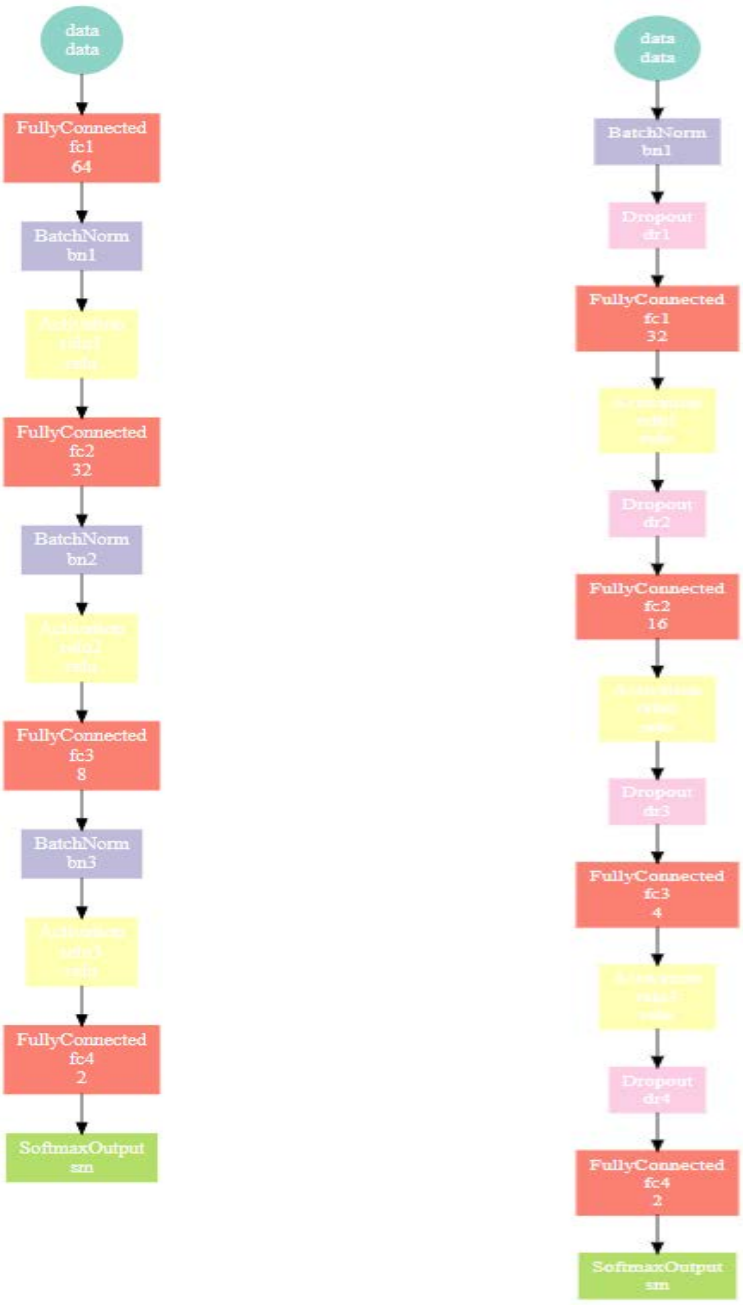


Figure 7: Illustrative structure of some Deep Neural Network structures employed in the optimization process

References

Addo P. M., Guegan D., and Hassani B. (2018). Credit Risk Analysis Using Machine and Deep Learning Models. Risks, 6, 2 (38): 2227-9091.

Altman, E.: "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy." The journal of finance 23.4 (1968): 589-609.

Avery, R. B., Calem, P. S., & Canner, G. B. (2004). Consumer credit scoring: Do situational circumstances matter? Journal of Banking & Finance, 28 (4), 835–856.

Breiman, L. (2001). Random forest. Machine Learning, 45, 5–32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. (1984). Classification and regression trees. CRC press.

Butaru, Florentin, Qingqing Chen, Brian Clark, Sanmay Das, Andrew W. Lo, and Akhtar Siddique. 2016. Risk and risk management in the credit card industry. Journal of Banking and Finance 72: 218–39.

Chava, S., & Jarrow,R.A. (2004). Bankruptcy prediction with industry effects. Re-view of Finance, 8 (4), 537–569.

Cho, K., Van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio. Y. (2014), "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," Proc. EMNLP.

Galindo, Jorge, and Pablo Tamayo. (2000). Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. Computational Economics 15: 107–43.

Huang, S. C. (2011). Using Gaussian process based kernel classifiers for credit rating forecasting. Expert Systems with Applications, 38 (7), 8607–8611.

Huang, Zan, Hsinchun Chen, Chia-Jung Hsu,Wun-Hwa Chen, and SoushanWu. 2004. Credit rating analysis with support vector machines and neural networks: A market comparative study. Decision Support Systems 37: 543–58.

Kamstra, M., Kennedy, p. and Suan, TK. (2001): Combining bond rating forecasts using logit. Financial Review 36.2: 75-96.

Khandani, Amir E., Adlar J. Kim, and Andrew W. Lo. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking and Finance 34: 2767–87.

Mizen, P., & Tsoukas, S. (2012). Forecasting US bond default ratings allowing for previous and initial state dependence in an ordered probit model. International Journal of Forecasting, 28 (1), 273–287.

Ohlson, J.(1980): Financial ratios and the probabilistic prediction of bankruptcy." Journal of accounting research: 109-131.

Petr, G., & Gurný, M. (2013). Comparison of credit scoring models on probability of default estimation for US banks. Prague Economic Papers, 2, 163–181.

Petropoulos A., Chatzis S.P., Xanthopoulos S (2016). A novel corporate credit rating system based on Student's-t hidden Markov models. Expert Systems with Applications, 53, 87-105.

Quadrana, M., Hidasi, B., Karatzoglou, A. and Cremonesi, P. (2017), Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks, Proc. ASM RecSys.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. The Journal of Business, 74 (1), 101–124.

Srivastava, N, Hinton, J., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15 (2014) 1929-1958.

Tu, Z., Lu, Z., Liu. Y., Liu, X., and Li, H. Modeling coverage for neural machine translation. Proc. ACL (2016).

Vapnik, V. N. (1998). Statistical learning theory. New York: Wiley.

Vinod, Nair & Hinton, Geoffrey (2010), Rectified Linear Units Improve Restricted Boltzmann Machines. Proc. ICML.

Yeh, C.-C., Lin, F., & Hsu, C.-Y. (2012). A hybrid KMV model, random forests and rough set theory approach for credit rating. Knowledge-Based Systems, 22, 166–172.

Zhao, Z, Xu, S, Kang, B. H, Kabir, M. M. J, Liu, Y, & Wasinger, R. (2015). Investigation and improvement of multi-layer perception neural networks for credit scoring. Expert Systems with Applications, 42 (7), 3508–3516.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

A robust machine learning approach for credit risk analysis of large loan-level datasets using deep learning and extreme gradient boosting[1]

Anastasios Petropoulos, Vasilis Siakoulis,
Evaggelos Stavroulakis and Aristotelis Klamargias,

Bank of Greece

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting

BIS International Workshop on Big Data for Central Bank Policies

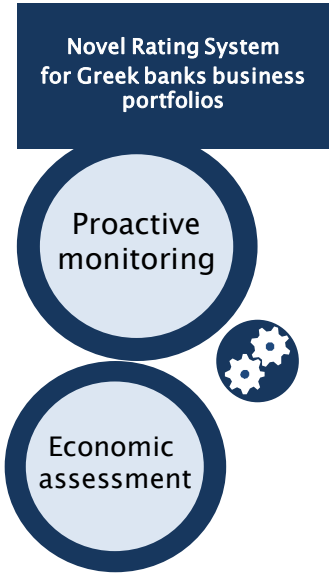
Indonesia, 25 July 2018

BANK OF GREECE
Anastasios Petropoulos
Vasilis Siakoulis
Evangelos Stavroulakis
Aristotelis Klamargias

The views expressed in this paper are those of the authors and not necessarily those of Bank of Greece

Credit Risk Analysis Tool

In a nutshell



Modeling technique

- Extreme Gradient Boosting
- Deep Neural Networks

Main Drivers

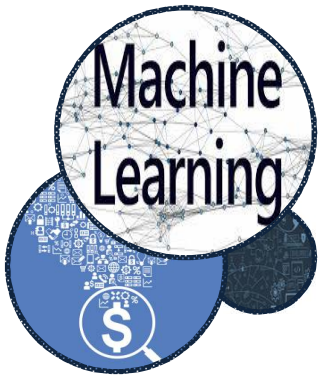
- Company Financial Ratios
- Macroeconomic factors

Implementation

- Corporate and SME loans of the Greek banking system, from the supervisory database of the Central Bank of Greece

Credit Risk Analysis

Machine and Deep learning techniques

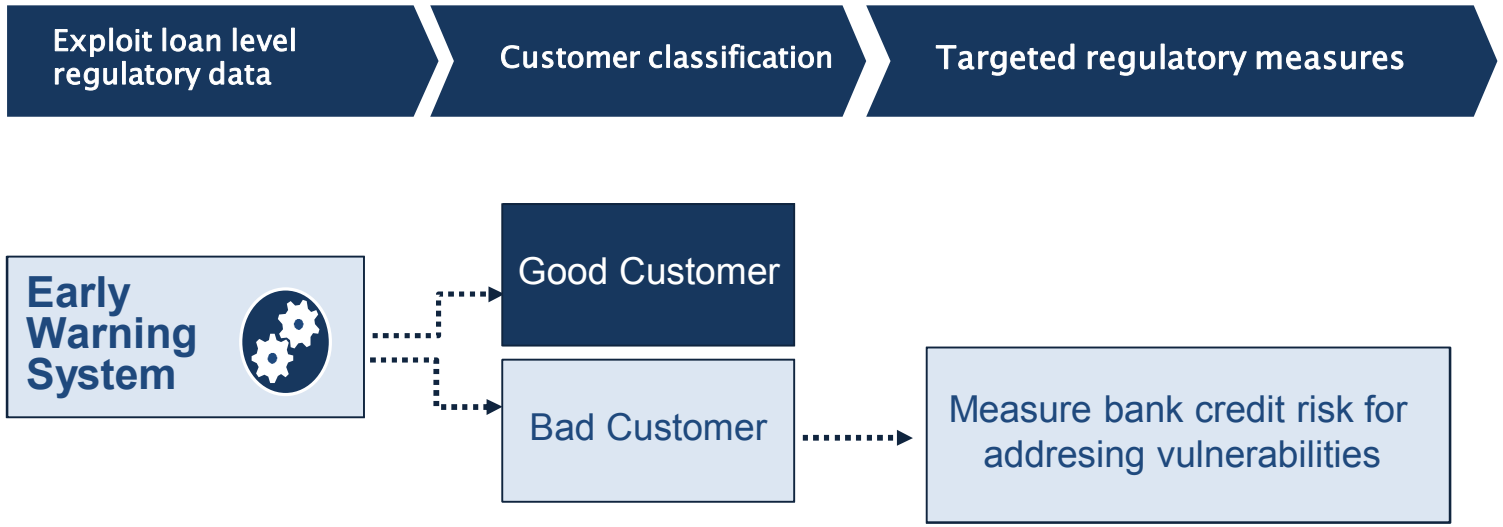


- "Learn" without being explicitly programmed
- Unveiling new determinants and unexpected forms of dependencies among variables.
- Tackling non linear relationships.

- Use of ML and Deep Learning are favored by the technological advances and the availability of financial sector data.
- Supervisory authorities should keep up with the current developments.

Credit Risk Analysis

Bank of Greece – Regulatory Purpose

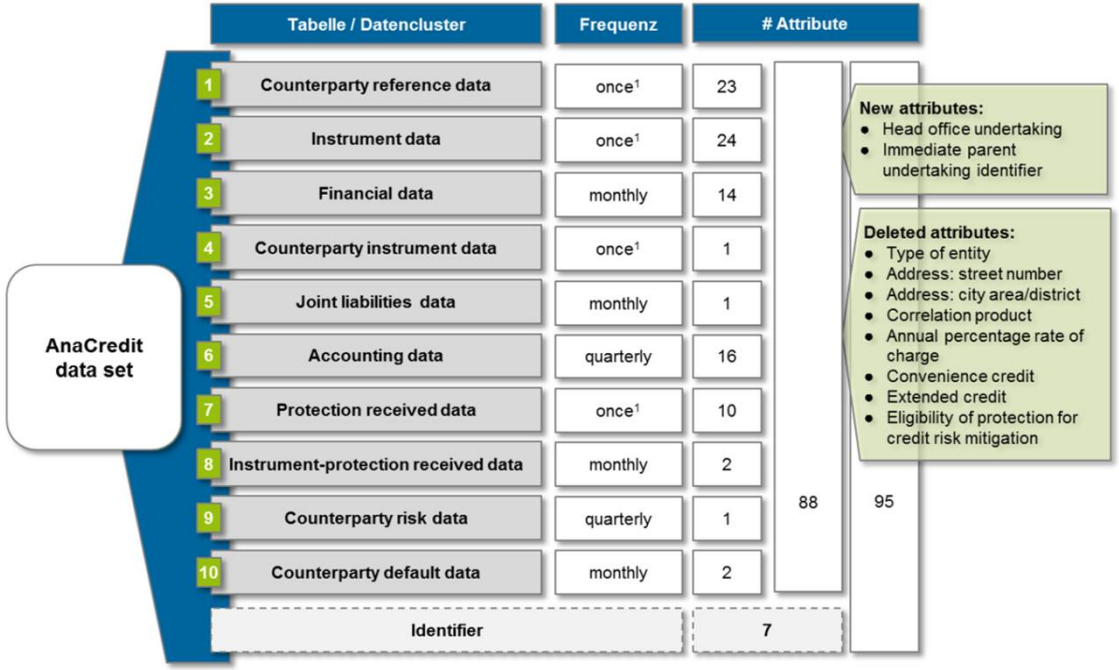


BANK OF GREECE
EUROSYSTEM

Credit Risk Analysis – Big Data

Anacredit project European Central bank

Reporting threshold 25.000 euro



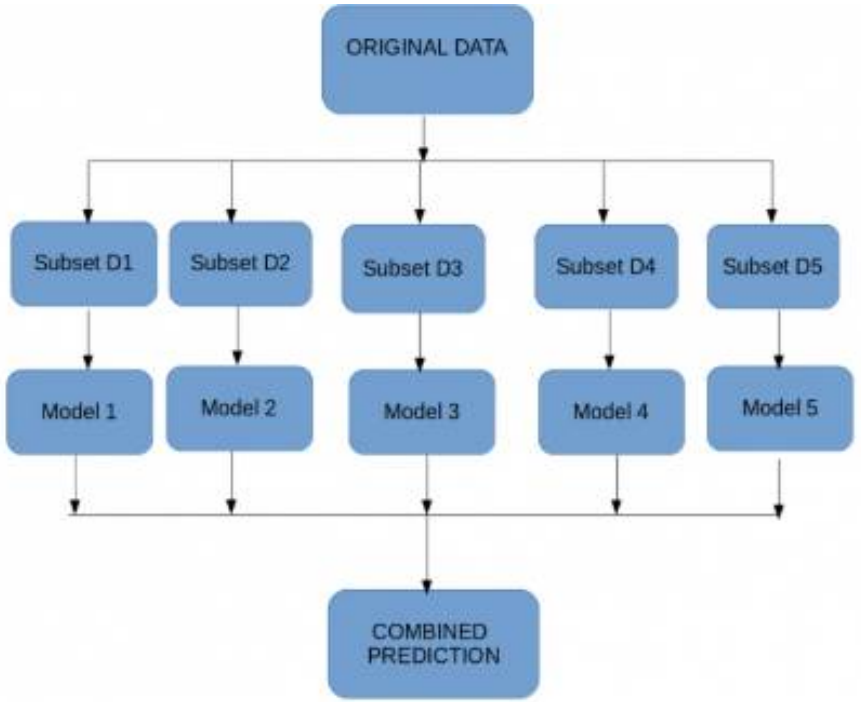
Source: ECB regulation on the collection of granular credit and credit risk data as of May 18th, 2016

- AnaCredit will be a new dataset with detailed information on individual bank loans in the euro area.
- The project was initiated in 2011 and data collection is scheduled to start in September 2018.

BANK OF GREECE
EUROSYSTEM

Credit Risk Analysis

Bagging – Different models vote for the result

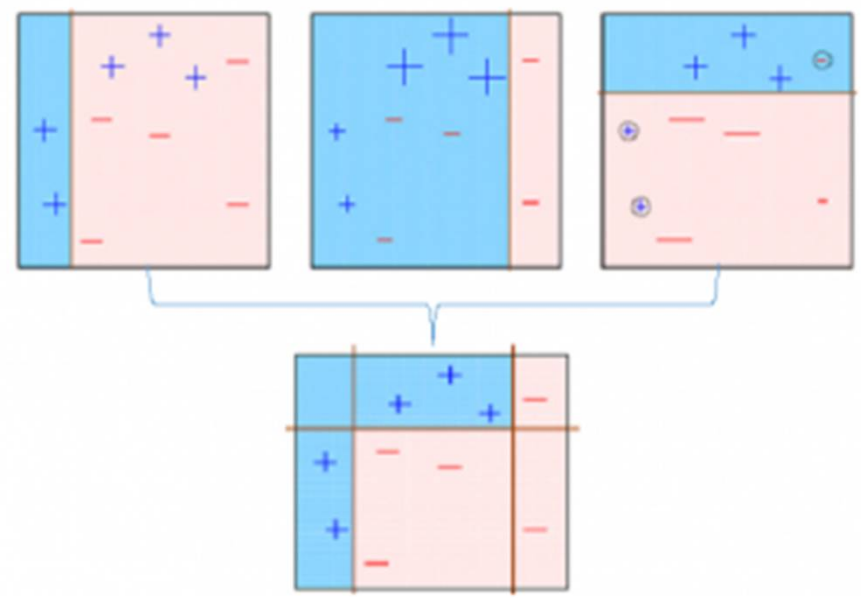


source: Analytics Vidhya

- Multiple subsets are created from the original dataset, selecting observations with replacement and a base model (weak model) is created on each of these subsets.
- The models run in parallel and are independent of each other.
- The final predictions are determined by combining the predictions from all the models.
- Random Forests are common employed bagging techniques.

Credit Risk Analysis

Boosting – Each model learns from the errors of the previous



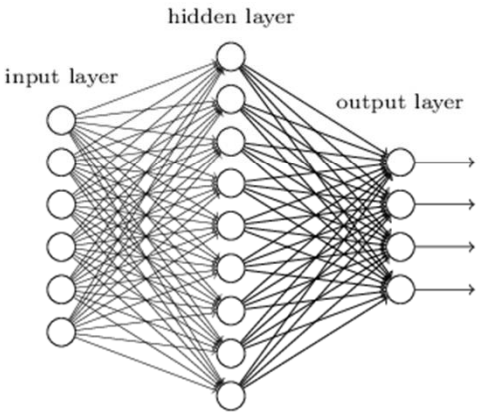
source:
Analytics
Vidhya

- A base model is created based on a subset of the original dataset which is used to make predictions on the whole dataset.
- Errors are calculated and observations which are incorrectly predicted, are given higher weights (large plus signs).
- Another model is created which tries to correct the errors from the previous model.
- Similarly, multiple models are created, each correcting the errors of the previous model.
- The final model (strong learner) is the weighted mean of all the models (weak learners).

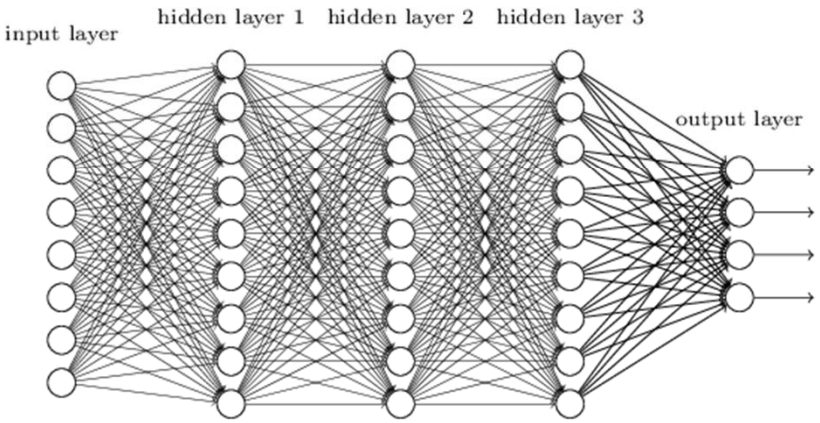
Credit Risk Analysis

Deep Neural Networks-Unlimited potential for Architectures

"Non-deep" feedforward neural network



Deep neural network

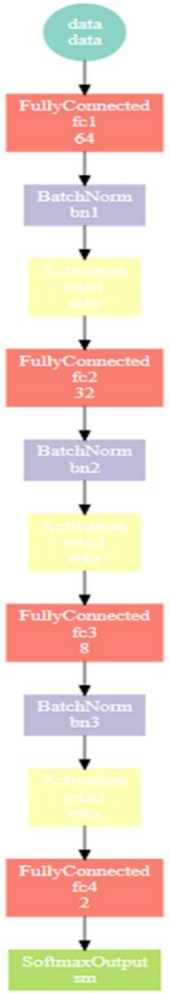


Deep neural network is simply a feedforward network with many hidden layers. It has the following advantages compared to one layer networks ("shallow")

- A deep network needs less neurons than a shallow one
- A shallow network is more difficult to train with our current algorithms (e.g. it has more nasty local minima, or the convergence rate is slower)

Credit Risk Analysis

Deep Neural Networks-Unlimited potential for Architectures



This methodology provides the opportunity of creating a large combination of different structures based on

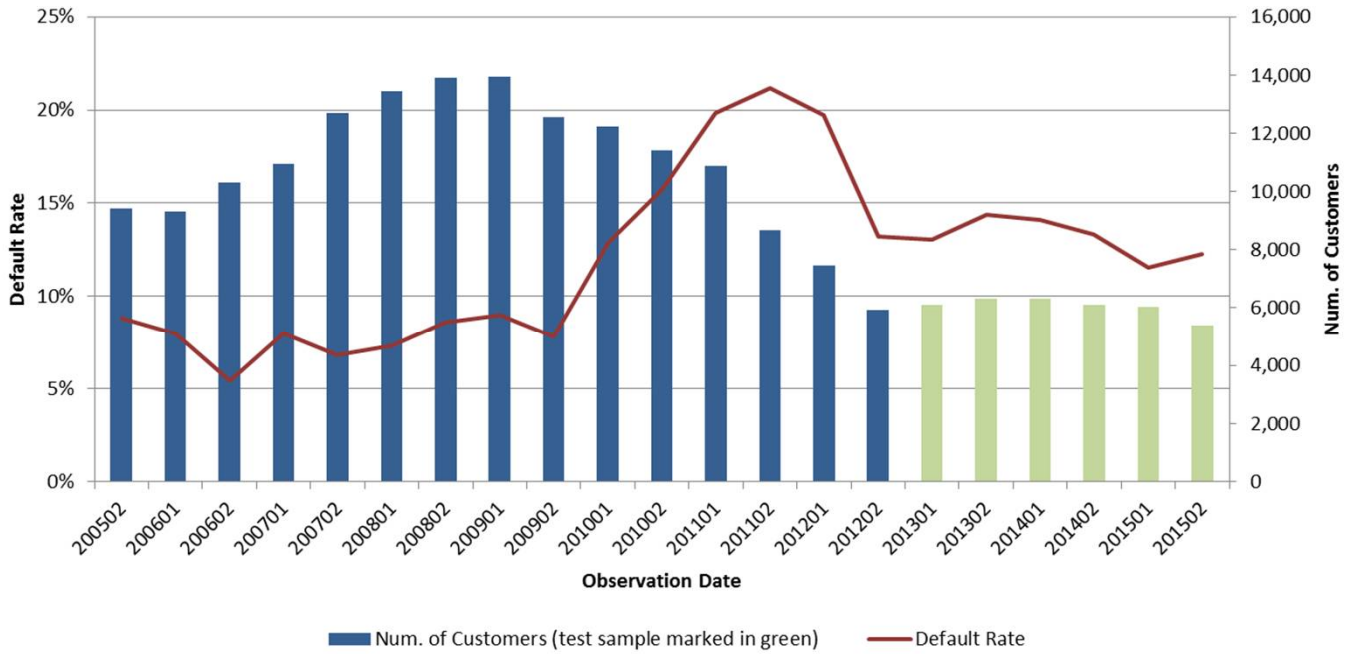
- Number of layers,
- Selection of activation function
- Number of perceptrons
- Normalization layers
- Dropout adjustments

Which can be employed in the optimization process

BANK OF GREECE
EUROSYSTEM

Credit Risk Analysis

Problem at hand



- We have collected loan level information on Corporate and SME loans of the Greek banking system, from the supervisory database of the Central Bank of Greece.
- A loan is flagged as delinquent if it is either 90 days past due or it gets rated as delinquent based on each bank's internal rating rules.
- The forecast horizon for a default event is 1 year whereas the variables employed include macro data and company specific financial ratios.

Credit Risk Analysis

Many Predictor Candidates - Curse of dimensionality

Boruta (aka Leshy): Slavik deity dueling in forests. 1906 illustration

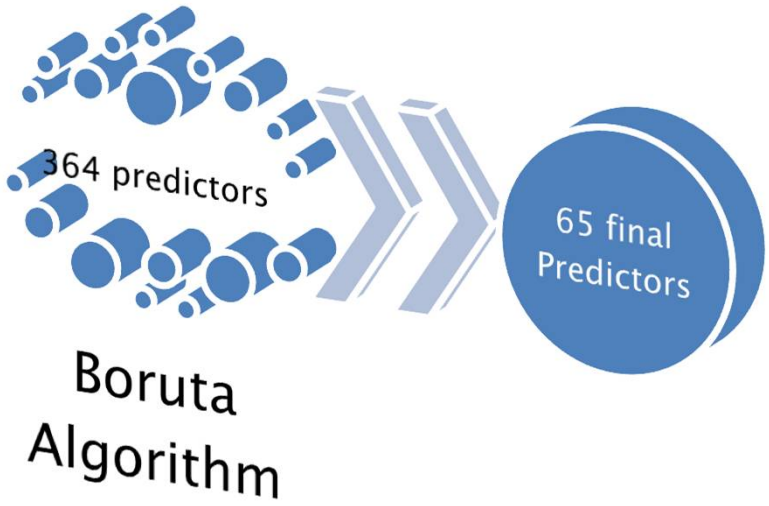
- We employ **Boruta algorithm** for tackling the dimensionality issue. This is sequential Random Forest based algorithm which removes non relevant variables decreasing the dimensionality space.

Credit Risk Analysis

Many Predictor Candidates - Curse of dimensionality

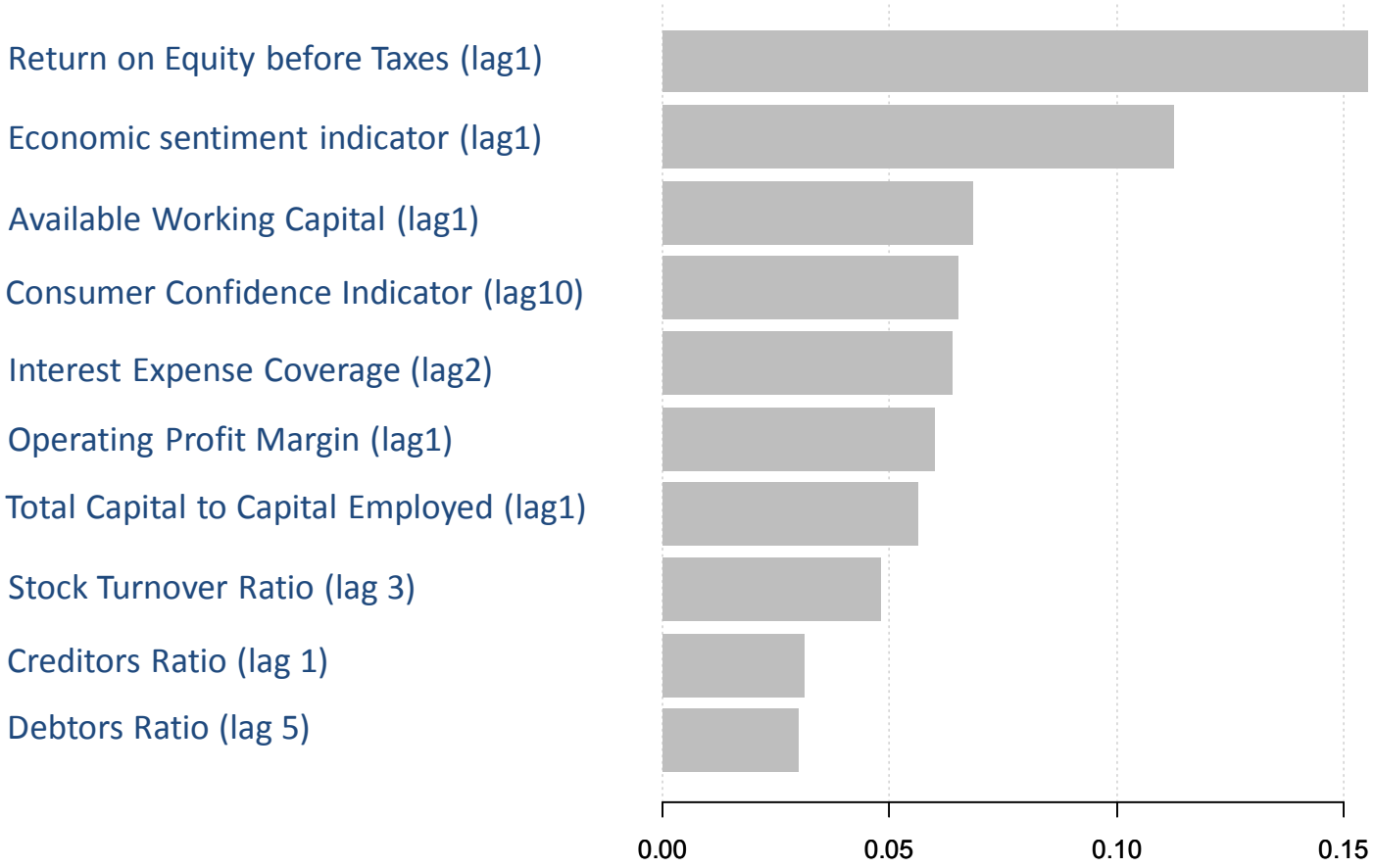
Boruta Algorithm – steps:

- First, it adds randomness to the given dataset by creating shuffled copies of all features (shadow features).
- Then, it fits a Random Forest model (bagging model) on the extended dataset and evaluates the importance of each feature based on Z score.
- In every iteration, it checks whether a real feature has a higher importance than the best of its shadow features, and constantly removes features which are deemed unimportant



Extreme Gradient Boosting

Variable Importance



Extreme Gradient Boosting

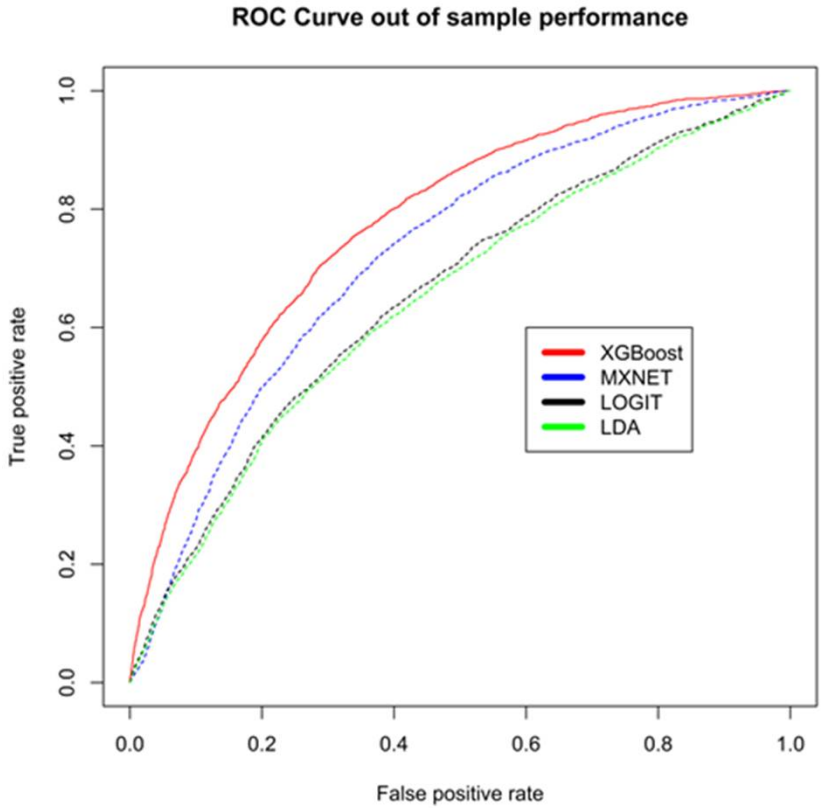
Classification Accuracy

Classification Accuracy Model Comparison	KS	AUROC
Logit	24%	66%
LDA	23%	65%
XGBoost	42%	78%
MXNET	35%	72%

Table 1

Classification Accuracy Metrics: Kolmogorov - Smirnov (KS), Area Under ROC curve (AUROC).

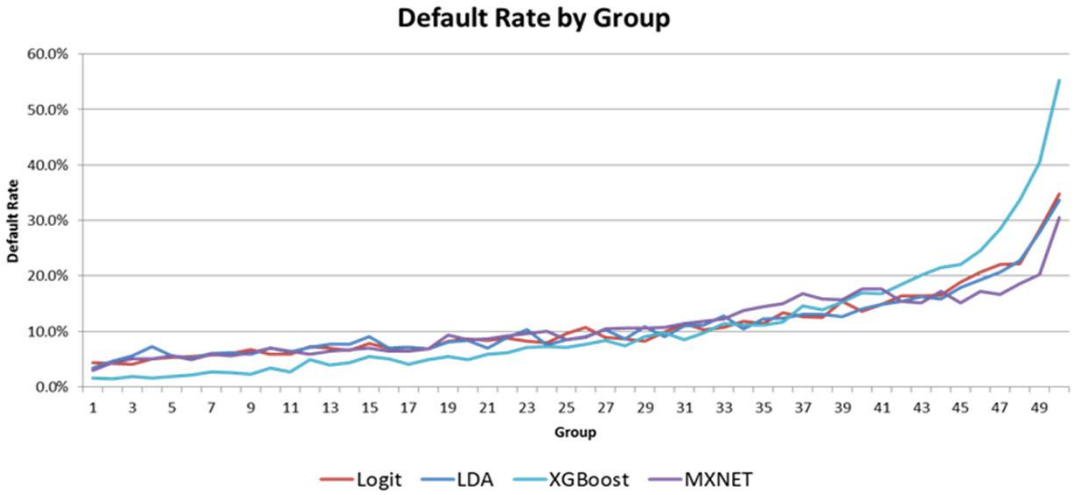
XGBoost and **MXNET** algorithms provide better classification accuracy compared to traditional classification methods such as Logistic Regression and Linear Discriminant analysis.



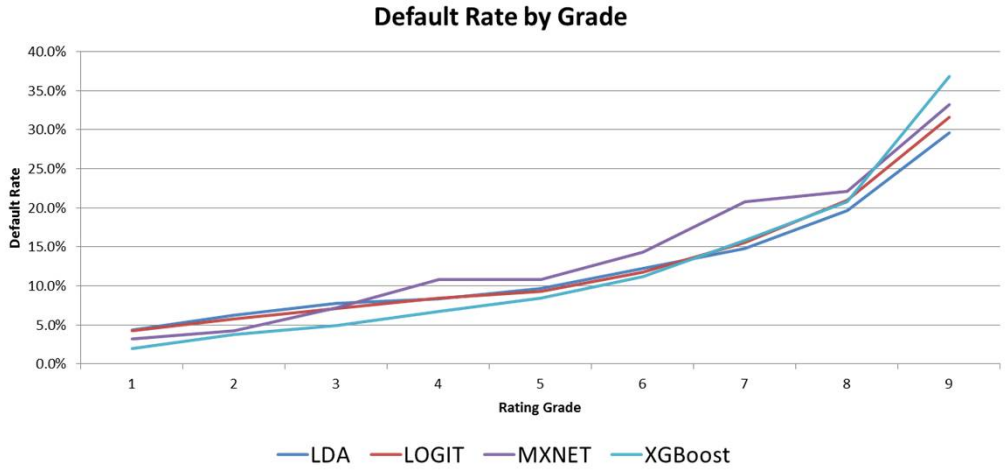
BANK OF GREECE
EUROSYSTEM

Credit Risk Analysis

Calibrating a Rating system



Initial credit rating segmentation in 50 grades



Final credit rating segmentation in 9 grades

BANK OF GREECE
EUROSYSTEM

Deep Neural Networks

Rating System Performance

Performance Metrics — Table 2

Credit Rating System

	SSE	BRIER
Logit	4.3%	11.3%
LDA	4.8%	11.4%
XGBoost	0.2%	10.1%
MXNET	0.6%	11.0%

Rating System Calibration Metrics: Sum of Square Error (SSE), Brier's score (BRIER).

Estimated and Actual default frequency metrics — Table 3

	Estimated Probability of Default	Observed Default Rate (Out of sample)	Observed Default Rate (In sample)
Logit	8.20%	13.10%	11.00%
LDA	7.80%		
XGBoost	13.50%		
MXNET	15.00%		

Estimated Probability of Default vs observed Default Rate in out-of-sample and in-sample population

- Based on SSE and Brier score the MXNET and XGBOOST rating systems perform better than Logistic Regression and Linear Discriminant analysis.
- The estimated PDs for MXNET and XGBOOST are closer to the observed default rates.

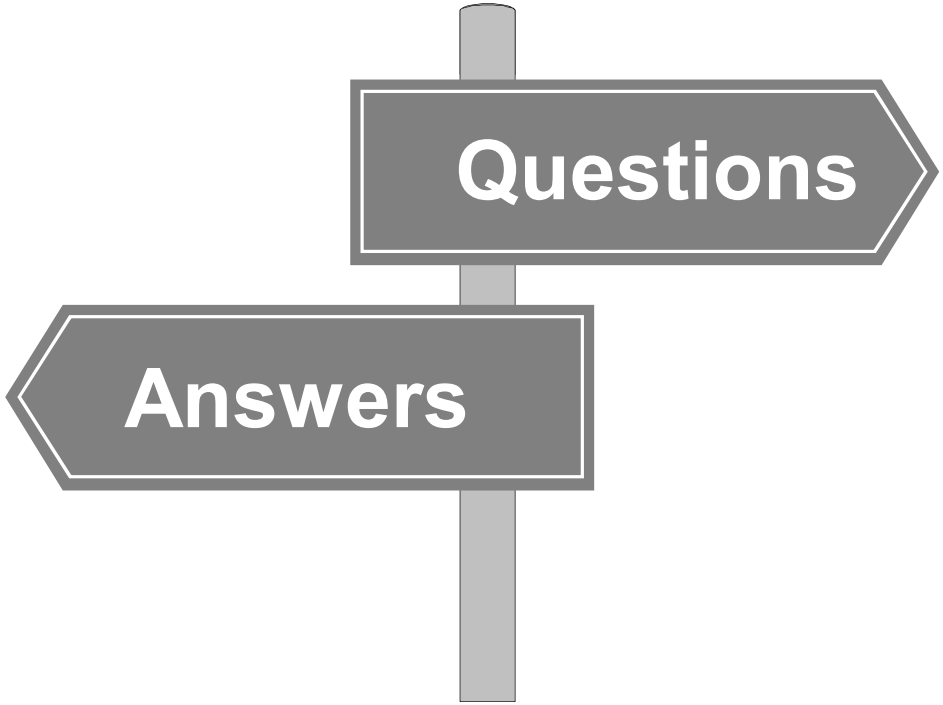
Credit Risk Analysis

Our Contribution

- Extensive exploration of advanced statistical techniques
- An automated algorithm for tackling dimensionality issues
- Application to a regulatory large size dataset
- Robust validation and Performance Measures
- Large potential for application in large datasets (Anacredit)

Credit Risk Analysis

Q&A



BANK OF GREECE
EUROSYSTEM

Thank you!

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Big data and FinRisk[1]

Sanjiv R. Das,
Santa Clara University

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Big Data and FinRisk

Sanjiv R. Das
Santa Clara University

Bank of Indonesia and BIS/IFC
International Conference on Big Data
Bali, Indonesia, July 26, 2018

Outline

1. Overview of challenges in Big Data for Finance
2. Financial networks for Systemic Risk Measurement.
    a. USA
    b. India
3. Zero-revelation prediction of bank malaise.

Research Challenges in Financial Data Modeling and Analysis

Lewis Alexander[1], Sanjiv R. Das[2,*], Zachary Ives[3], H.V. Jagadish[4], and Claire Monteleoni[5]

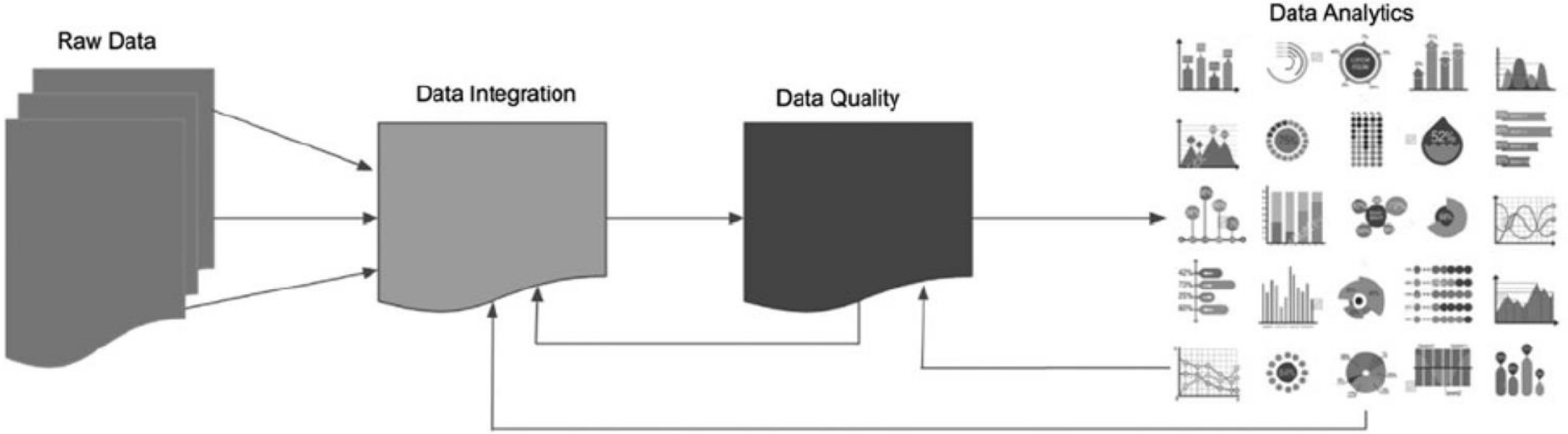
Abstract

Significant research challenges must be addressed in the cleaning, transformation, integration, modeling, and analytics of Big Data sources for finance. This article surveys the progress made so far in this direction and obstacles yet to be overcome. These are issues that are of interest to data-driven financial institutions in both corporate finance and consumer finance. These challenges are also of interest to the legal profession as well as to regulators. The discussion is relevant to technology firms that support the growing field of FinTech.

http://srdas.github.io/Papers/big.2016.0074_FINAL.pdf

Big Data
Volume 5, Number 3, 2017

DOI: 10.1089/big.2016.0074



Staging Template

Areas	Type of issues/problems: Level 1: Curation at the unit level within a firm	Level 2: Curation and aggregation at the firm level	Level 3: Curation and aggregation at the system level	Across levels: Quality issues (privacy, veracity, etc.)
Data integration				
Standards	☐	○	○	☐
Application-specific tools	☐	○	○	○
Text mining tools	☐	☐	○	○
Data quality management				
BSBS239 (14 principles, 4 areas)	☐	○	○	☐
Errors in recording, extraction, entity-matching, interpretation	☐	☐	○	○
Data timeliness: Nowcasting	☐	○	○	☐
Data manipulation	☐	☐	○	○
Data analytics				
Feature selection	☐	☐	☐	○
Model selection	☐	☐	☐	○
Online learning	☐	○	○	○
AI and deep learning	☐	○	○	○
Systemic risk	☐	☐	☐	☐
Consumer finance	☐	☐	☐	☐
Text analytics	☐	☐	☐	☐
High frequency trading	☐	—	○	○
Blockchains	☐	—	○	○
Cybersecurity	☐	☐	☐	○

Codes: ○ represents nascent solutions; ☐ represents work underway, not fully developed; and empty cells represent that decent progress has been made.

Systemic Analysis

The Dodd-Frank Act (2010) and Basel III regulations characterize a systemically risky FI as one that is

1. Large;
2. Complex;
3. Interconnected;
4. Critical, i.e., provides hard to substitute services to the economy.

The DFA does not provide quantification guidance.

Systemic Analysis

Definition: the measurement and analysis of relationships across entities with a view to understanding the impact of these relationships on the system as a whole.

Challenge: requires most or all of the data in the system; therefore, high-quality information extraction and integration is critical.

Attributes of Systemic Risk Measures

Systemic risk is an attribute of the economic system and not that of a single entity. Its measurement should have two important features:

1. Quantifiability (Aggregation): must be measurable on an ongoing basis.

1. Decomposability (Attribution): Aggregate system-wide risk must be broken down into additive risk contributions from all entities in the system.

Financial institutions that make large risk contributions to system-wide risk are deemed “systemically important.”

An Extensive Literature

References

Abbass, P., C. Brownlees, C. Hans, and N. Podlich (2016, April nectedness: What does the market really know? *Journal of* 1–12.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Syst in financial networks. *American Economic Review 105*, 564

Acharya, V., R. Engle, and M. Richardson (2012). Capital sho to ranking and regulating systemic risks. *American Econom*

Acharya, V., L. Pedersen, T. Philippon, and M. Richardson (suring systemic risk. *Review of Financial Studies 30*(1), New York University.

Acharya, V., P. Schnabl, and G. Suarez (2013). Securitization *Journal of Financial Economics 109*, 515–536.

Acharya, V. V., J. A. C. Santos, and T. Yorulmazer (2010, New Bank of New York, August)). Systemic risk and deposit *Economic Policy Review*.

Adrian, T. and M. K. Brunnermeier (2016). Covar. *An view 106*(7), 1705–1741.

Adrian, T. and H. Shin (2010). Liquidity and leverage. *Jour mediation 19*, 418–437.

Ahern, K. R. (2013). Network centrality and the cross-section ing Paper, USC-Marshall School of Business.

Allen, F. and E. Carletti (2013). What is systemic risk? *Jou and Banking 45*, 121–127.

Allen, L., T. Bali, and Y. Tang (2012). Does systemic risk predict future economic downturns? *Review of Financial St*

Anand, K., P. Gai, S. Kapadia, and S. Brennan (2013). A netv system resilience. *Journal of Economic Behavior and Organ*

Avdjiev, S., M. Chui, and H. Shin (2014). Non-financial corpc market economies and capital flows. *BIS Quarterly Review*.

Avramidis, P. and F. Pasiouras (2015). Calculating systemic model approach. *Journal of Financial Stability 16*, 138–150

Benoit, S., J. Colliard, C. Hurlin, and C. Perignon (2017). survey on systemic risk. *Review of Finance 21*(1), 109–152.

Betz, F., N. Hautsch, T. A. Peltonend, and M. Schienle (2016). Syster in the european banking and sovereign network. *Journal of Finan* 206–224.

Bianchiy, D., M. Billio, R. Casarinz, and G. Massimo (2015). Mo and systemic risk. Working Paper, University of Warwick.

Billio, M., M. Getmansky, D. Gray, A. Lo, R. Merton, and L. F Sovereign, bank and insurance credit spreads: Connectedness and Working Paper, International Monetary Fund.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012a). Econc of connectedness and systemic risk in the finance and insurance se *Financial Economics 104*(3), 535–559.

Bisias, D., M. Flood, A. Lo, and S. Valavanis (2012). A survey analytics. *Annual Review of Financial Economics 4*, 255–296.

Black, L., R. Correa, X. Huang, and H. Zhou (2016). The syster pean banks during the financial and sovereign debt crises. *Journa Finance 63*, 107–125.

Bluhm, M. and J. P. Krahnen (2014). Systemic risk in an interco system with endogenous asset markets. *Journal of Financial Stab*

Bonacich, P. (1987). Power and centrality: A family of measures. *A of Sociology 92*(5), 1170–1182.

Bonacich, P. and P. Lloyd (2001, July). Eigenvector-like measures asymmetric relations. *Social Networks 23*(3), 191–201.

Borri, N. (2017). Local currency systemic risk. Working Paper, ssrn

Brownlees, T. and R. Engle (2015). Srisk: A conditional capital sl systemic risk measurement. Working Paper, New York University

Brunetti, C., J. H. Harris, S. Mankad, and G. Michailidis (2015). ness in the interbank market. Finance and Economics Discussion Governors of the Federal Reserve System.

Brunnermeier, M. (2009). Deciphering the liquidity and credit crun *Journal of Economic Perspectives 23*, 77–100.

Brunnermeier, M. and L. Pedersen (2009). Market liquidity and *Review of Financial Studies 22*(6), 2201–2238.

Chan-Lau, J., M. A. Espinosa-Vega, K. Giesecke, and J. Solé (2009 systemic implications of financial linkages. IMF Global Financial Vol. 2.

ChanLau, J. A., C. Chuang, J. Duan, and W. Sun (2016, May). and systemic risk via forwardlooking partial default correlation IMF.

Colliard, J.-E., T. Foucault, and P. Hoffmann (2017). Interbank mented otc market. Working Paper, European Central Bank.

Covitz, D., N. Liang, and G. Suarez (2013). The evolution of Collapse of the asset-backed commercial paper market. *Jour* 815–848.

Das, S. R. (2016). Matrix metrics: Network-based systemic risk *Alternative Investments 18*(4), 33–51.

Das, S. R., S. R. Kim, and D. N. Ostrov (2017). Dynamic syste Working Paper, Santa Clara University.

Das, S. R. and J. Sisk (2005). Financial communities. *Journal of ment 31*(4), 112–133.

De Bandt, O. and P. Hartmann (2000). Systemic risk: A survey European Central Bank.

Demirer, M., F. X. Diebold, L. Liu, and K. Yilmaz (2017). Estin network connectedness. *Journal of Applied Econometrics*.

Diebold, F. and K. Ylmaz (2014). On the network topology of va tions: Measuring the connectedness of financial firms. *Journal of* 119–134.

Duan, J.-C. and W. Miao (2016). Default correlations and lat analysis. *Journal of Business & Economic Statistics 34*(4), 536

Duarte, F. and T. M. Eisenbach (2015). Fire-sale spillovers and syst Paper, Federal Reserve Bank.

Elliott, M., B. Golub, and M. Jackson (2014). Financial netwo *American Economic Review 104*, 3115–3153.

Freeman, L. (1977). A set of measures of centrality based on betw *try 40*, 35–41.

Gabrieli, S. and C.-P. Georg (2014). A network view on interba Working Paper, Banque de France.

Gale, D. M. and S. Kariv (2007). Financial networks. *American Papers and Proceedings*.

Giglio, S., B. Kelly, and S. Pruitt (2016). Systemic risk and the m empirical evaluation. *Journal of Financial Economics 119*(3),

Gobat, J., T. Barnhill, A. Jobst, T. Kisinbay, H. Oura, T. (2011). How to address the systemic part of liquidity ri

Goodhart, C. (2009, August). Liquidity management. Jacl ity and Macroeconomic Policy Symposium, Federal Res

Gorton, G. and A. Metrick (2012). Securitized banking an *of Financial Economics 104*, 425–451.

Hanson, S., A. Kashyap, and J. Stein (2011). A macroprud regulation. *Journal of Economic Perspectives 25*, 3–28.

Härdle, W. K., W. Wang, and L. Yuc (2016). Tenet: Tail- *Journal of Econometrics 192*(2), 499–513.

Hautsch, N., J. Schaumburg, and M. Schienle (2015). Finai contributions. *Review of Finance 19*, 685–738.

Huang, X., H. Zhou, and H. Zhu (2012). Systemic risk *Financial Services Research 42*, 55–83.

Karolyi, A., J. Sedunov, and A. Taboada (2016). Cross-b temic risk. Working Paper, Cornell University.

Kitwiwattanachai, C. (2015). Learning network structu from cds data. Working Paper, University of Connecti

Laeven, L., L. Ratnovski, and H. Tong (2016). Bank size, Some international evidence. *Journal of Banking and F*

Lehar, A. (2005). Measuring systemic risk: A risk manage *Banking and Finance 29*, 2577–2603.

Li, J. and G. Zinna (2014). On bank credit risk: Systemic for the united states and united kingdom. *Journal of I Analysis 5/6*, 1403–1442.

Liang, N. (2013). Systemic risk monitoring and financial st *Credit and Banking 45*, 129–135.

Liu, S., C. Wu, C.-Y. Yeh, and W. Yoo (2015). What d evidence from the us state cds market. Working Paper,

Markose, S., S. Giansante, and A. Shaghaghi (2012). financial network of us cds market: Topological fragility *of Economic Behavior and Organization 83*, 627–646.

Merton, R. C. (1973). Theory of rational option pricing. *and Management Science 4*, 141–183.

Nier, E., J. Yang, T. Yorulmazer, and A. Alentorn (2007). Network models and financial stability. *Journal of Economic Dynamics and Control 31*, 2033–2060.

Nucera, F., B. Schwaab, S. J. Koopman, and A. Lucas (2016). The information in systemic risk rankings. *Journal of Empirical Finance 38*, 461–475.

Oh, D. H. and A. J. Patton (2016). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Journal of Business and Economic Statistics forthcoming.*

Pagano, M. S. and J. Sedunov (2016). A comprehensive approach to measuring the relation between systemic risk exposure and sovereign debt. *Journal of Financial Stability 23*, 62–78.

Perotti, E. and J. Suarez (2009). Liquidity risk charges as a macroprudential tool. CEPR Policy Insight.

Poledna, S., J. L. Molina-Borboad, S. Martínez-Jaramillod, M. van der Leije, and S. Thurner (2015). The multi-layer network nature of systemic risk and its implications for the costs of financial crises. *Journal of Financial Stability 20*, 70–81.

Saldías, M. (2013). Systemic risk analysis using forward-looking distance-to-default series. *Journal of Financial Stability 9*, 498–517.

Schwarcz, S. (2008). Systemic risk. *Georgetown Law Journal 97*, 193–249.

Sedunov, J. (2016). What is the systemic risk exposure of financial institutions? *Journal of Financial Stability 24*, 71–87.

Sensoya, A. (2017). Firm size, ownership structure, and systematic liquidity risk: The case of an emerging market. *Journal of Financial Stability forthcoming.*

Silva, W., H. Kimura, and A. Sobreiro (2017). An analysis of the literature on systemic financial risk: A survey. *Journal of Financial Stability 28*, 91–114.

Tasca, P., P. Mavrodiev, and F. Schweitzer (2014). Quantifying the impact of leveraging and diversification on systemic risk. *Journal of Financial Stability 15*, 43–52.

Billio, Getmansky, Lo, Pelizzon (2012)

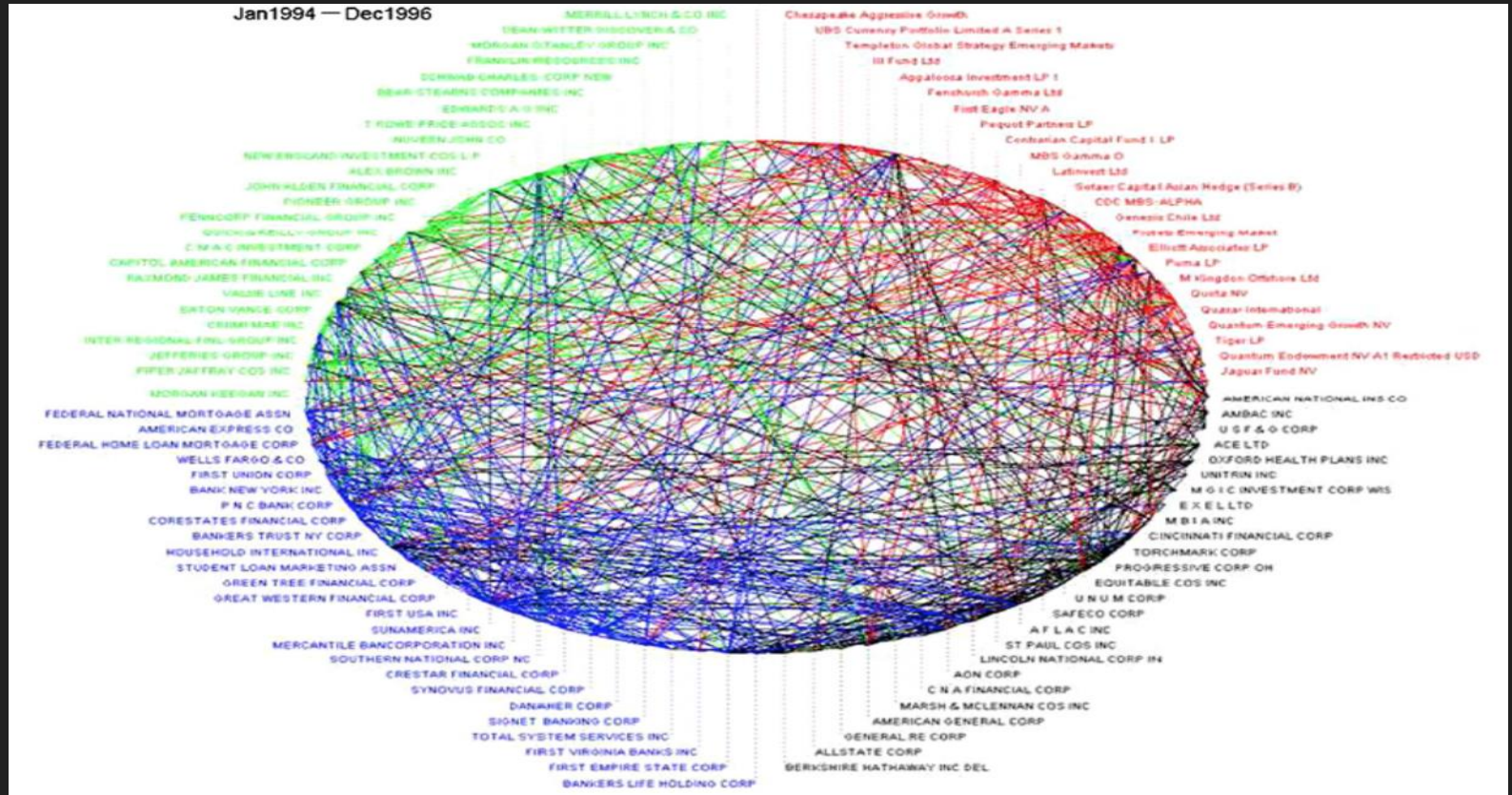
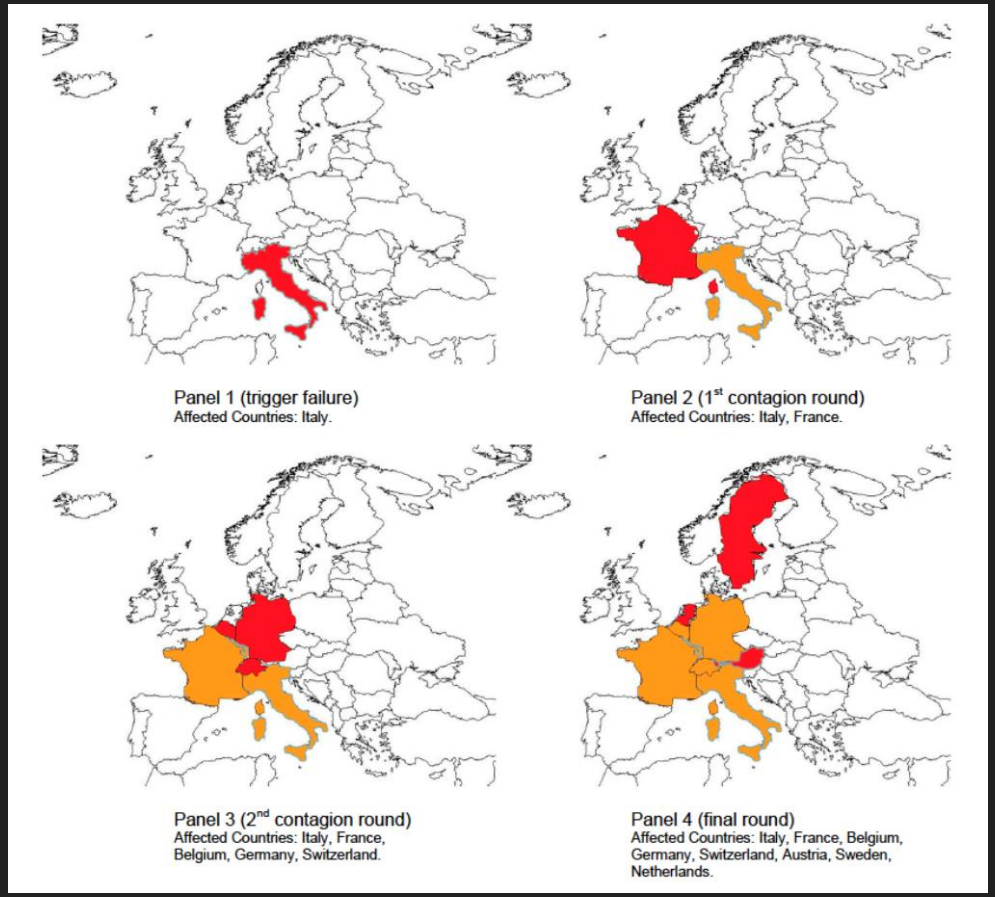


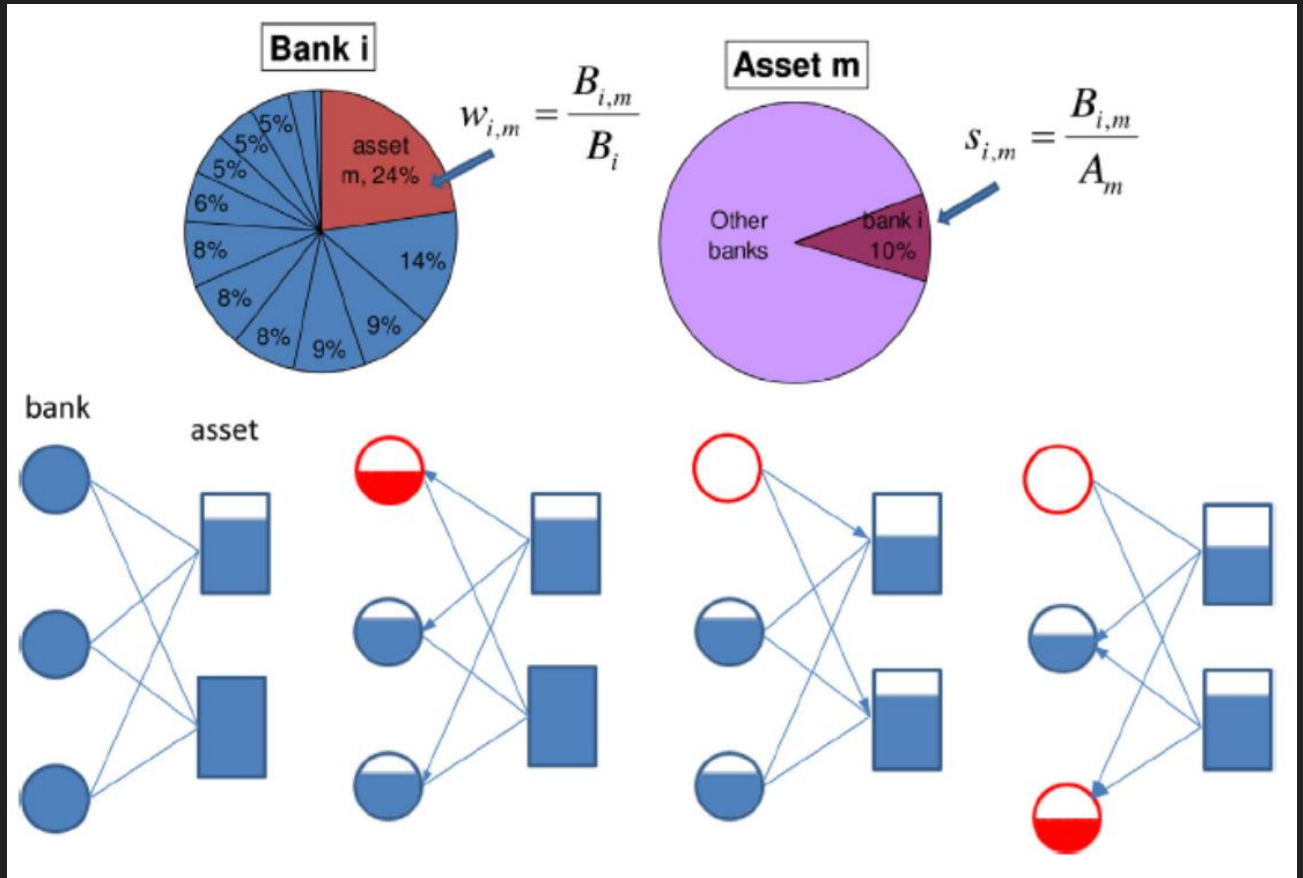
Fig. 2. Network diagram of linear Granger-causality relationships that are statistically significant at the 5% level among the monthly returns of the 25 largest (in terms of average market cap and AUM) banks, broker/dealers, insurers, and hedge funds over January 1994 to December 1996. The type of institution causing the relationship is indicated by color: green for broker/dealers, red for hedge funds, black for insurers, and blue for banks. Granger-causality relationships are estimated including autoregressive terms and filtering out heteroskedasticity with a GARCH(1,1) model.

Contagion Networks (Espinosa-Vega & Sole, IMF 2010)

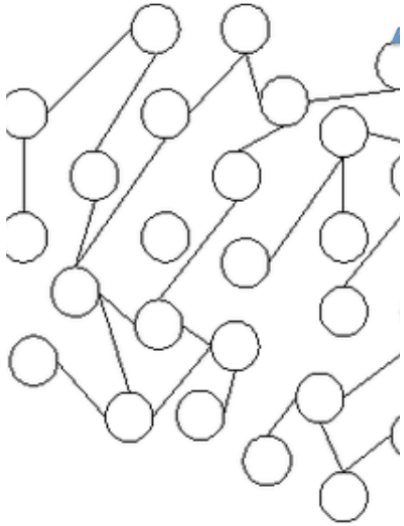


Bivalent Networks

Levy-Carciente, Kennet, Avakian, Stanley, Havlin, JBF 2015

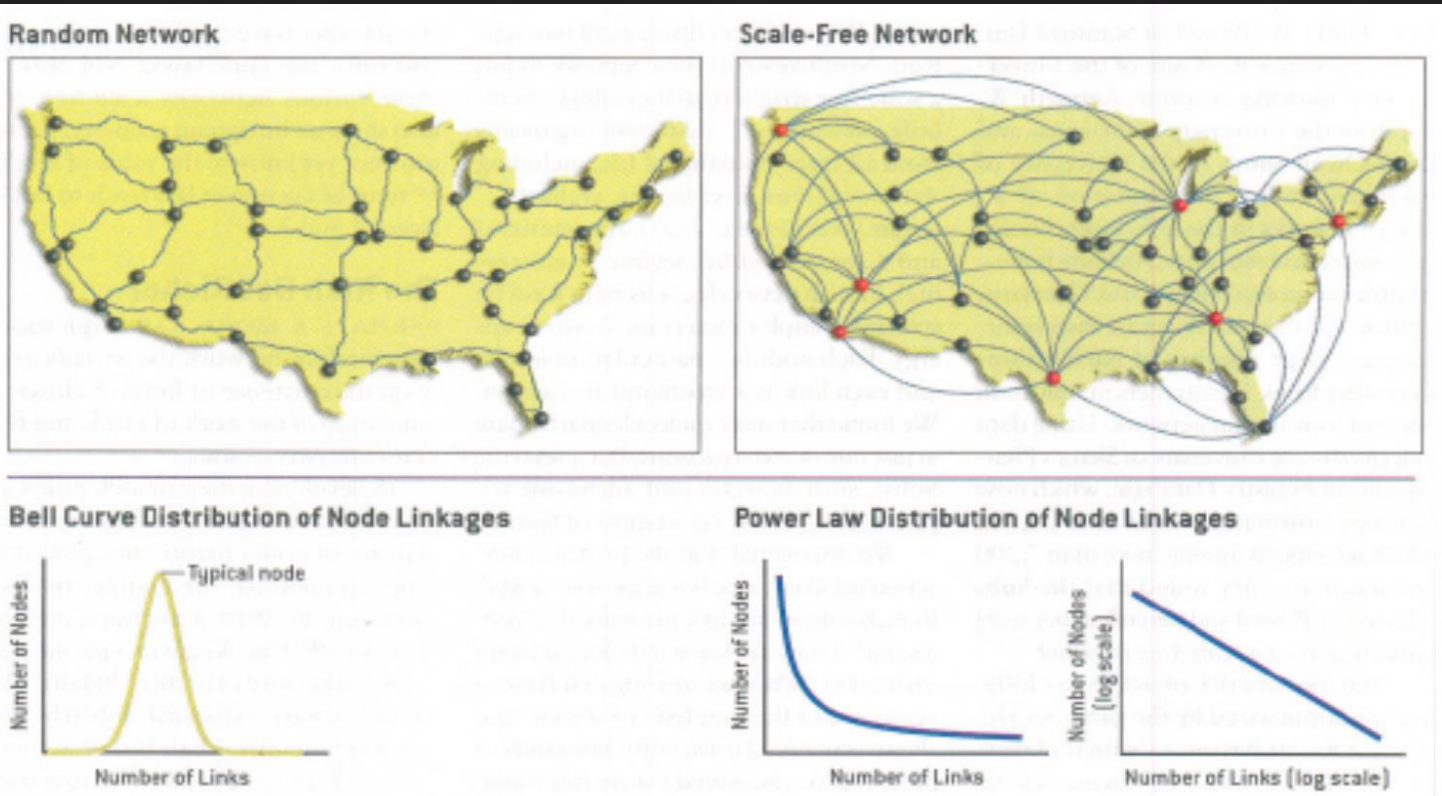


Graph Theory



(a) Random ne

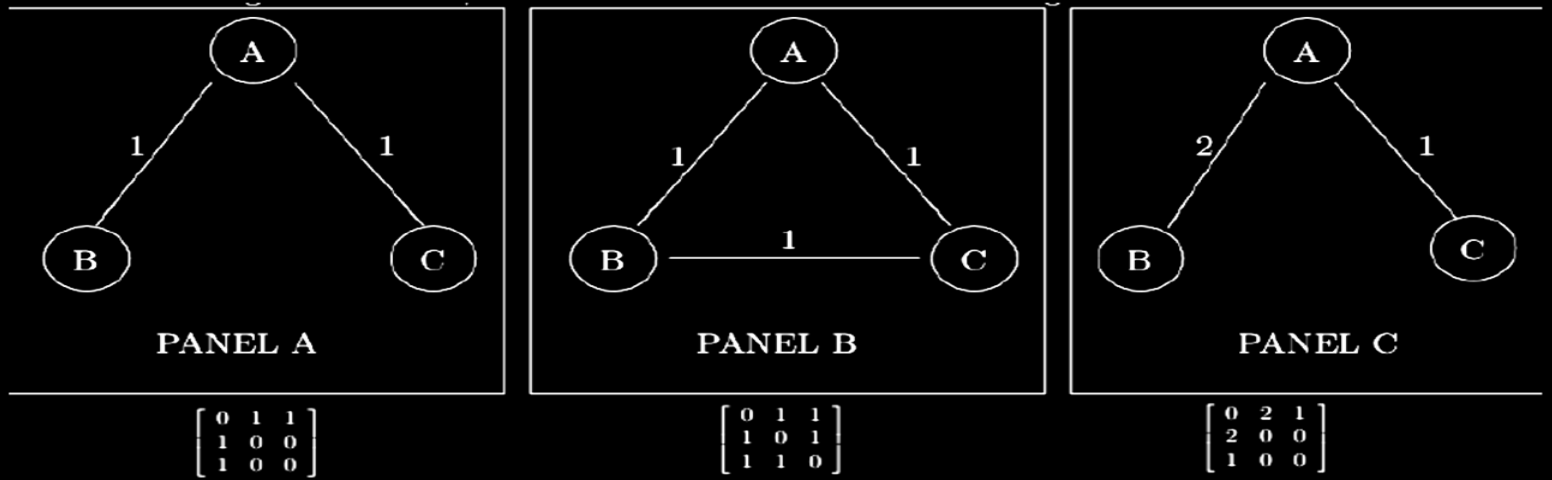
$f(d) \sim N(\mu, \sigma)$



Barabasi, Sciam, May 2003

Centrality (Bonacich 1987)

- Similar to PageRank by Google.
- Adjacency matrix: $A_{ij} \in \mathcal{R}^{N \times N}$
- Influence: $x_i = \sum_{j=1}^{N} A_{ij} x_j$
- $\lambda \mathbf{x} = \mathbf{A} \cdot \mathbf{x}$
- Centrality is the eigenvector $\mathbf{x}$ corresponding to the largest eigenvalue.



$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Centrality scores = {0.71, 0.50, 0.50}

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Centrality scores = {0.58, 0.58, 0.58}

$$\begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Centrality scores = {0.71, 0.63, 0.32}

Longest sh

Fragility

- Definition: how quickly will the failure of any one node trigger failures across the network? Is network malaise likely to spread or be locally contained?
- Metric:

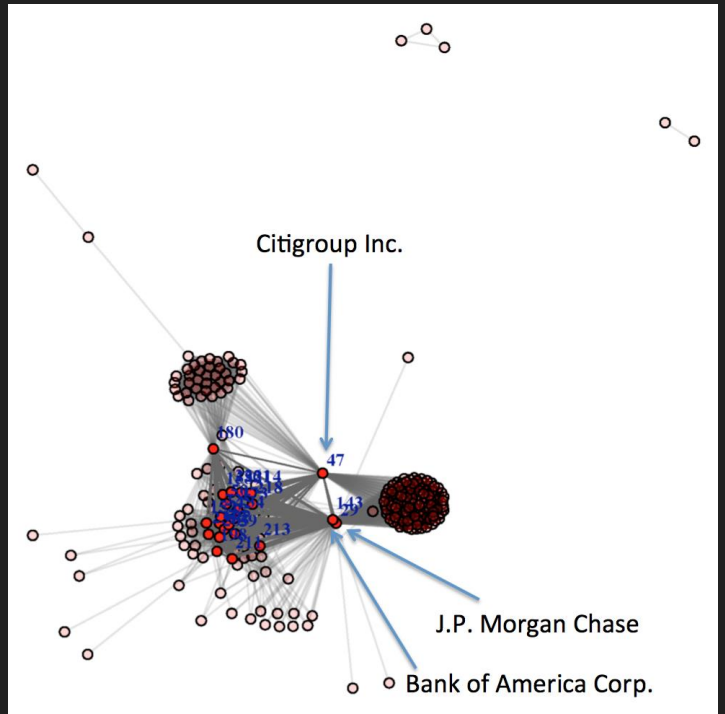
$$R = \frac{E(d^2)}{E(d)},$$

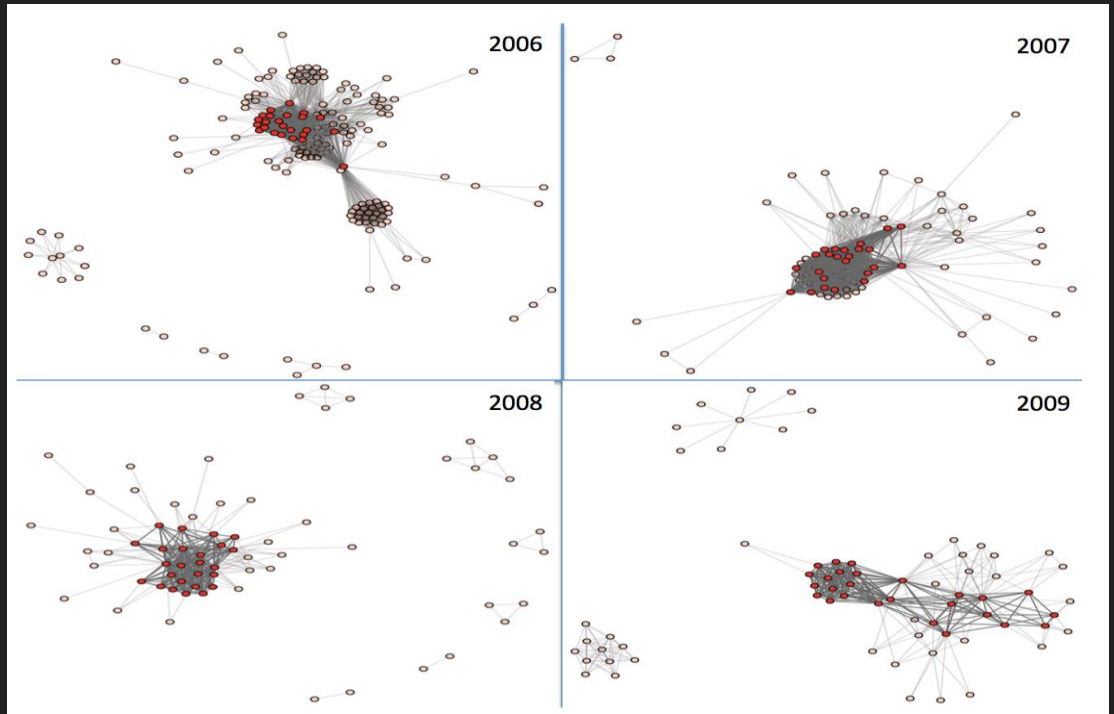
  where d is node degree.
- Similar to a normalized Herfindahl Index.
- Fragility of the sample network $= 20$

The diame

Interbank Loan Networks (U.S.)

"Extracting, Linking and Integrating Data from Public Sources: A Financial Case Study," (2011), (Douglas Burdick, Sanjiv Das, Mauricio A. Hernandez, Howard Ho, Georgia Koutrika, Rajasekar Krishnamurthy, Lucian Popa, Ioana Stanoi, Shivakumar Vaithyanathan), *IEEE Data Engineering Bulletin*, 34(3), 60-67.





Systemically Important Financial Institutions (SIFIs)

Year	#Colending banks	#Coloans	Colending pairs	$R = E(d^2)/E(d)$	Diam.
2005	241	75	10997	137.91	5
2006	171	95	4420	172.45	5
2007	85	49	1793	73.62	4
2008	69	84	681	68.14	4
2009	69	42	598	35.35	4

(Year = 2005)

Node #	Financial Institution	Normalized Centrality
143	J P Morgan Chase & Co.	1.000
29	Bank of America Corp.	0.926
47	Citigroup Inc.	0.639
85	Deutsche Bank Ag New York Branch	0.636
225	Wachovia Bank NA	0.617
235	The Bank of New York	0.573
134	Hsbc Bank USA	0.530
39	Barclays Bank Plc	0.530
152	Keycorp	0.524
241	The Royal Bank of Scotland Plc	0.523
6	Abn Amro Bank N.V.	0.448
173	Merrill Lynch Bank USA	0.374
198	PNC Financial Services Group Inc	0.372
180	Morgan Stanley	0.362
42	Bnp Paribas	0.337
205	Royal Bank of Canada	0.289
236	The Bank of Nova Scotia	0.289
218	U.S. Bank NA	0.284
50	Calyon New York Branch	0.273
158	Lehman Brothers Bank Fsb	0.270
213	Sumitomo Mitsui Banking	0.236
214	Suntrust Banks Inc	0.232
221	UBS Loan Finance Llc	0.221
211	State Street Corp	0.210
228	Wells Fargo Bank NA	0.198

One Score for Systemic Risk

$$S = \frac{1}{n}\sqrt{C^{\top} \cdot A \cdot C} \geq 0$$

banks (normalization across time)

Adjacency matrix

A(i,j) in (0,1)
A(i,i) = 1

Vector of credit risk scores {PD, rating, etc}. Higher = more risk

C(i) > 0

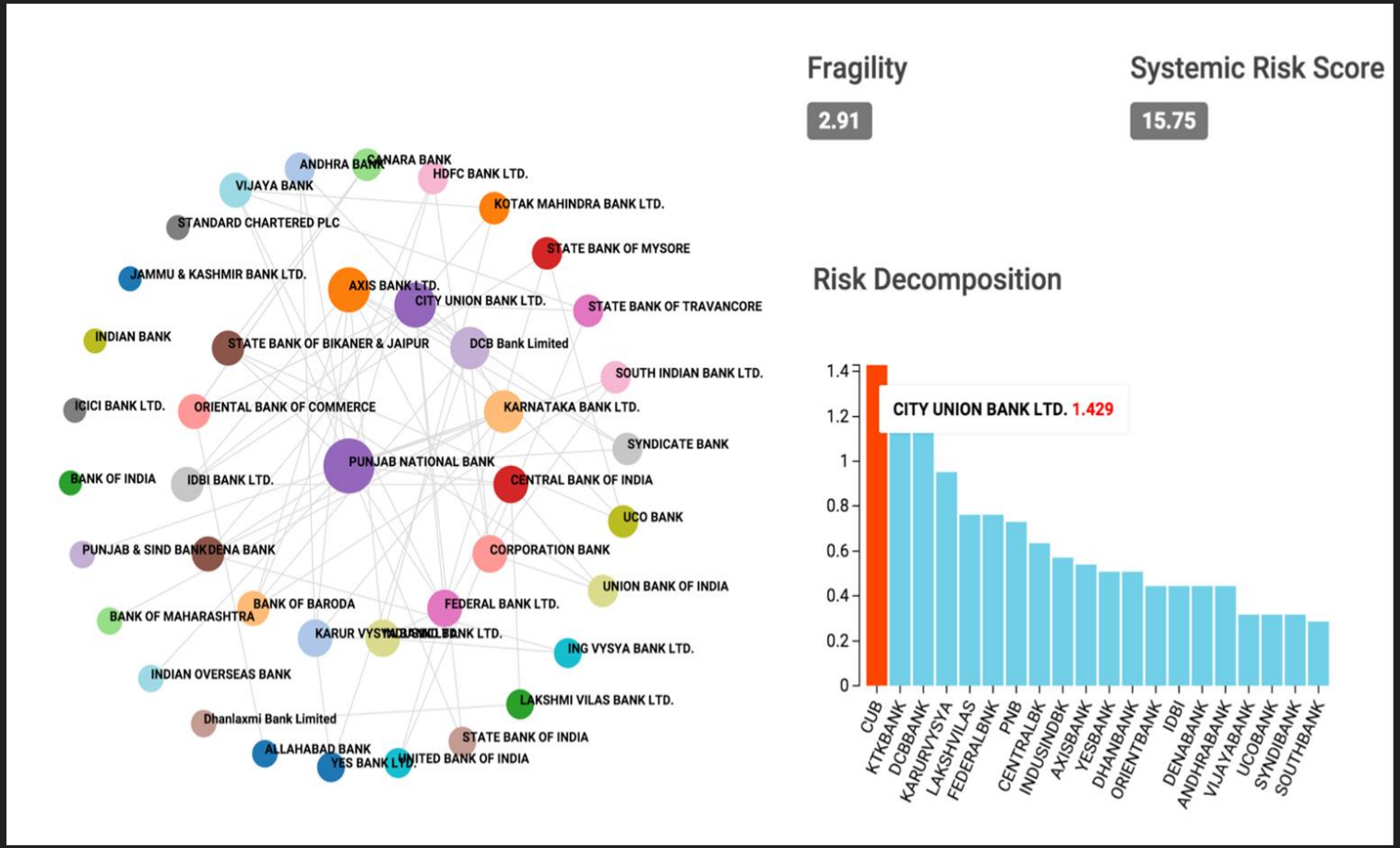
S(C,A) is linear homogenous in *C*

Apply Euler's Formula

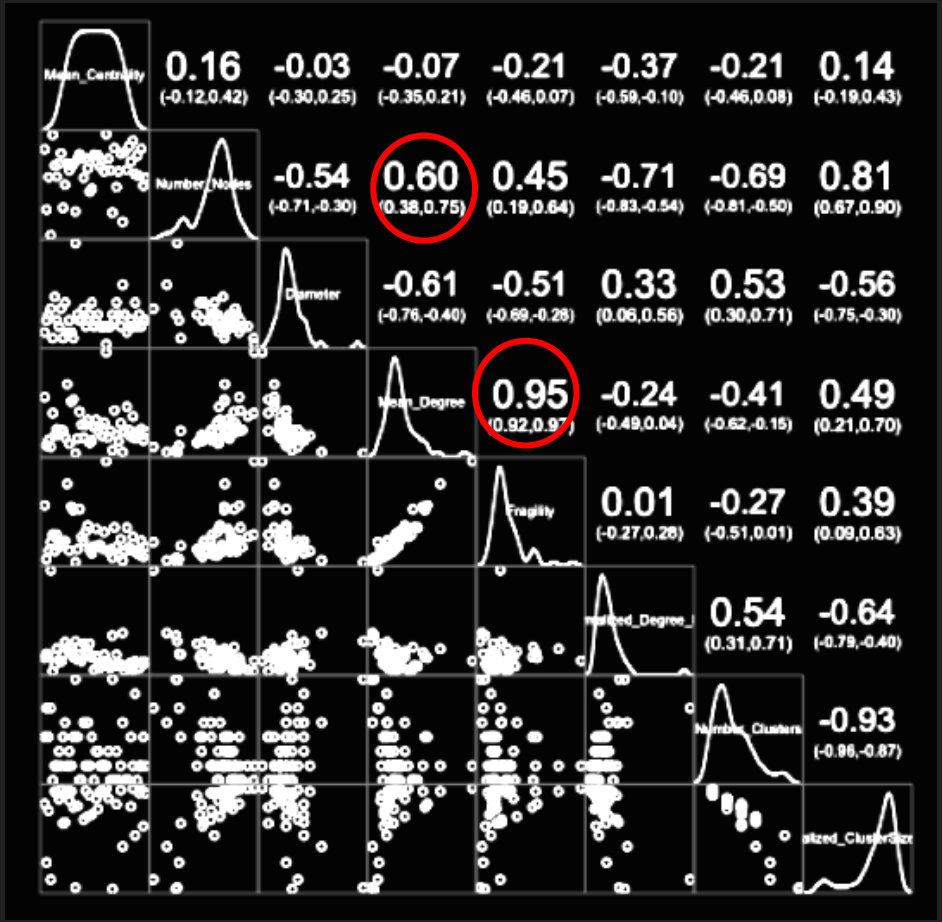
$$S = \frac{\partial S}{\partial C_1} C_1 + \frac{\partial S}{\partial C_2} C_2 + \ldots + \frac{\partial S}{\partial C_n} C_n = \sum_{i=1}^{n} \frac{\partial S}{\partial C_i} C_i$$

Risk Contribution

First iteration : India



Correlations



Mean Centrality

Number of Nodes

Diameter

Mean Degree

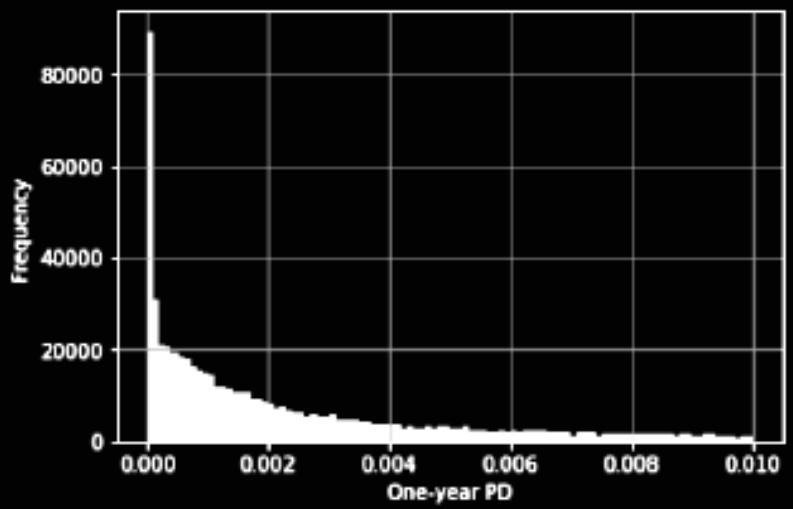
Fragility

Normalized degree Herfindahl Index

Number of Clusters

Normalized cluster size Herfindahl

Probabilities of Default (PDs)



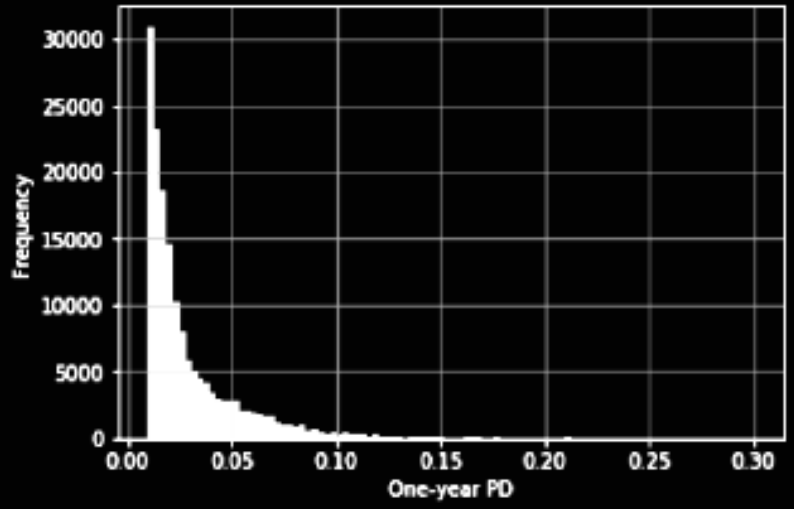


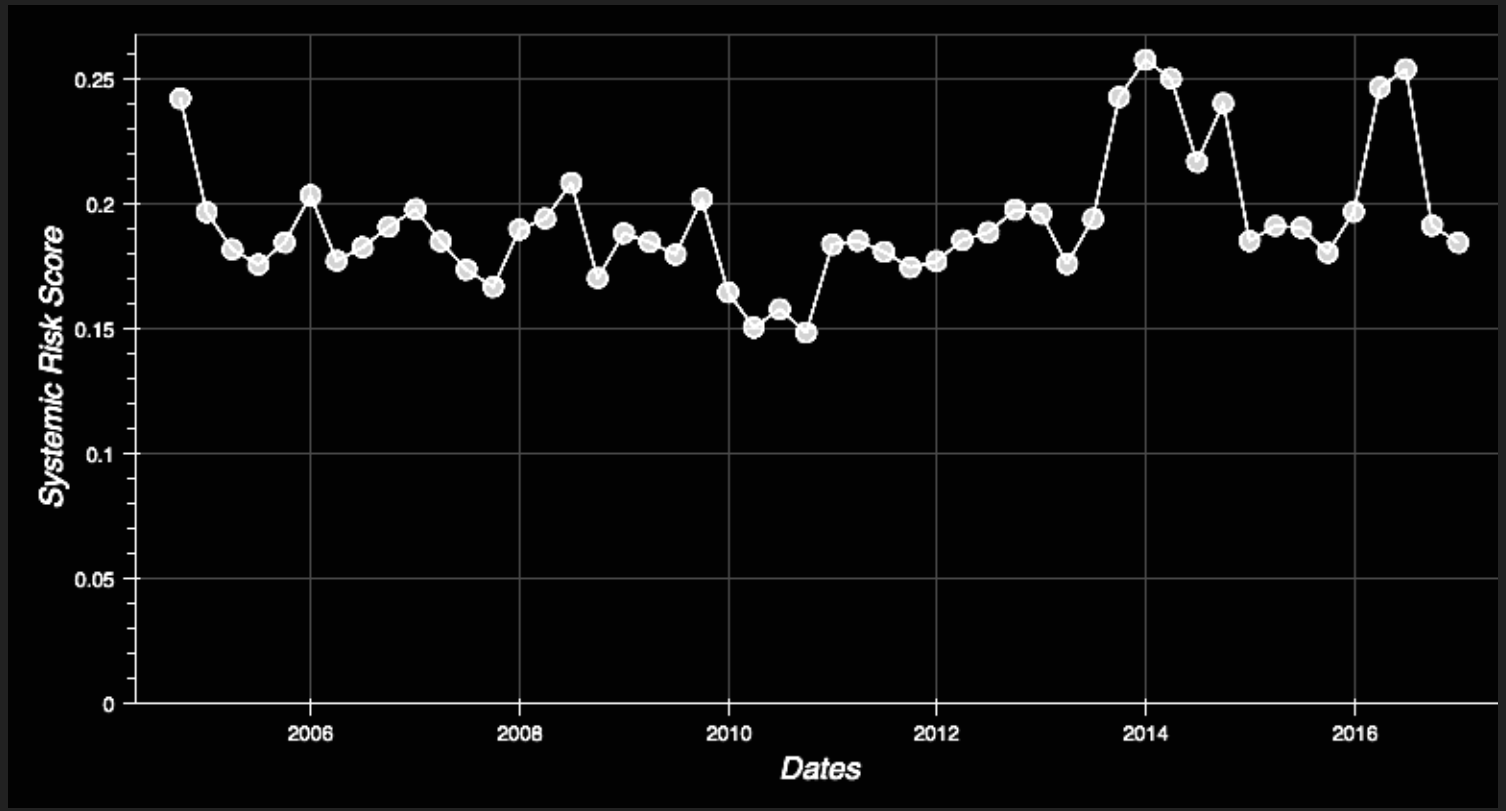
Figure 11: Distribution of PDs of all Indian FIs from 2004 to 2016. The first plot is the histogram of PDs that lie in the interval $(0, 0.01)$, and the second in the interval $(0.01, 0.30)$.

Highest PD = 26.36%

C = 1 + 30 PD

Since PD < 0.30, C lies in (0,10)

Systemic Risk Score (*S*)



Correlation of PDs and S = 69.7%

Risk Contributions of top 20 banks

	2005-Q1		2016-Q1	
	Bank Name	Risk Decomp	Bank Name	Risk Decomp
1	PRIME SECURITIES	2.705139	BANK OF MAHARASHTRA	2.222866
2	STATE BANK OF INDIA	2.476634	UCO BANK	1.698109
3	UCO BANK	2.438924	POWER FINANCE	1.437113
4	CORPORATION BANK	1.882045	UNITED BANK OF INDIA	1.410672
5	GIC HOUSING FINANCE	1.771204	STATE BK.OF BIN.& JAIPUR SUSP - SUSP.15/03/17	1.388539
6	I N G VYSYA BANK SUSP - SUSP.15/04/15	1.696898	DENA BANK	1.343904
7	UNION BANK OF INDIA	1.607279	STATE BANK OF INDIA	1.335314
8	IFCI	1.597618	INDIAN OVERSEAS BANK	1.331388
9	SUNDARAM FINANCE	1.569000	BANK OF TRAVANCORE SUSP - SUSP.15/03/17	1.309907
10	P N B GILTS	1.492469	CIL SECURITIES	1.282169
11	DHANLAXMI BANK	1.328556	COMFORT COMMOTRADE	1.137495
12	JAMMU & KASHMIR BANK	1.322932	BANK OF BARODA	1.093183
13	INDIABULLS FINL.SVS. SUSP - SUSP.18/03/13	1.215547	ANDHRA BANK	1.066791
14	DEWAN HOUSING FINANCE	1.198211	DEWAN HOUSING FINANCE	0.994385
15	ALMOND GLOBAL SECURITIES	1.195593	ORIENTAL BK.OF COMMERCE	0.917884
16	DENA BANK	1.194755	JAGSONPAL FIN.& LSG.	0.917517
17	ANDHRA BANK	1.193921	ELIXIR CAPITAL	0.873306
18	INDUSIND BANK	1.163923	MAHA.& MAHA.FINL.SVS.	0.871946
19	MARGO FINANCE	1.163827	CUBICAL FINANCIAL SVS.	0.855089
20	UNITED CREDIT	1.148539	VAX HOUSING FINANCE	0.852056
	TOTAL	31.36301	TOTAL	24.33963

Explaining quarterly systemic risk

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.1580***	0.1498***	0.2685	-0.0112	-0.0112	0.2730	0.2730
	(26.93)	(9.57)	(1.41)	(-0.04)	(-0.04)	(1.45)	(1.45)
Mean PD	3.8253***		5.2279***	5.0884***	5.0884***	5.2666***	5.2666***
	(6.73)		(18.85)	(9.50)	(9.50)	(10.45)	(10.45)
Mean Degree		0.0041*	0.0134***	0.0065	0.0065	0.0130**	0.0130**
		(2.30)	(3.58)	(1.58)	(1.58)	(2.76)	(2.76)
Degree HHI		6.4870*	5.7260*	4.3454	4.3454	6.2504**	6.2504**
		(2.42)	(2.55)	(2.01)	(2.01)	(3.00)	(3.00)
Mean Bet. Centrality			-0.0001***	-0.0001**	-0.0001**	-0.0001*	-0.0001*
			(-4.65)	(-3.05)	(-3.05)	(-2.67)	(-2.67)
Diameter			0.0003	0.0002	0.0002	-0.0002	-0.0002
			(0.50)	(0.31)	(0.31)	(-0.26)	(-0.26)
Fragility			-0.0039	0.0004	0.0004	-0.0034	-0.0034
			(-1.72)	(0.15)	(0.15)	(-1.26)	(-1.26)
Num. Clusters			-0.0030	-0.0014	-0.0014	-0.0033	-0.0033
			(-1.17)	(-0.52)	(-0.52)	(-1.35)	(-1.35)
Cluster HHI			-0.1429	-0.0110	-0.0110	-0.2230	-0.2230
			(-0.76)	(-0.05)	(-0.05)	(-1.20)	(-1.20)
Median Log(Assets)				0.0046	0.0046		
				(1.04)	(1.04)		
Median Log(Market Cap)						0.0040*	0.0040*
						(2.65)	(2.65)
Median Loans/Assets				0.0001	0.0001	0.0122	0.0122
				(0.01)	(0.01)	(0.97)	(0.97)
Median Loans/Deposits				0.0564	0.0564	-0.0035	-0.0035
				(0.59)	(0.59)	(-0.05)	(-0.05)
Median Debt/Assets				0.1058	0.1058		
				(1.29)	(1.29)		
Median Debt/Equity						0.1224*	0.1224*
						(2.16)	(2.16)
Median Debt/Capital				-0.0000	-0.0000	0.0002	0.0002
				(-0.04)	(-0.04)	(0.89)	(0.89)
Median ROA				0.0015	0.0015		
				(1.64)	(1.64)		
Median ROE						0.0001	0.0001
						(0.06)	(0.06)
Median Market/Book				0.0096	0.0096	-0.0020	-0.0020
				(1.23)	(1.23)	(-0.22)	(-0.22)
Observations	50	50	50	50	50	50	50
R^2	0.485	0.160	0.923	0.948	0.948	0.955	0.955
Adjusted R^2	0.475	0.124	0.908	0.925	0.925	0.935	0.935

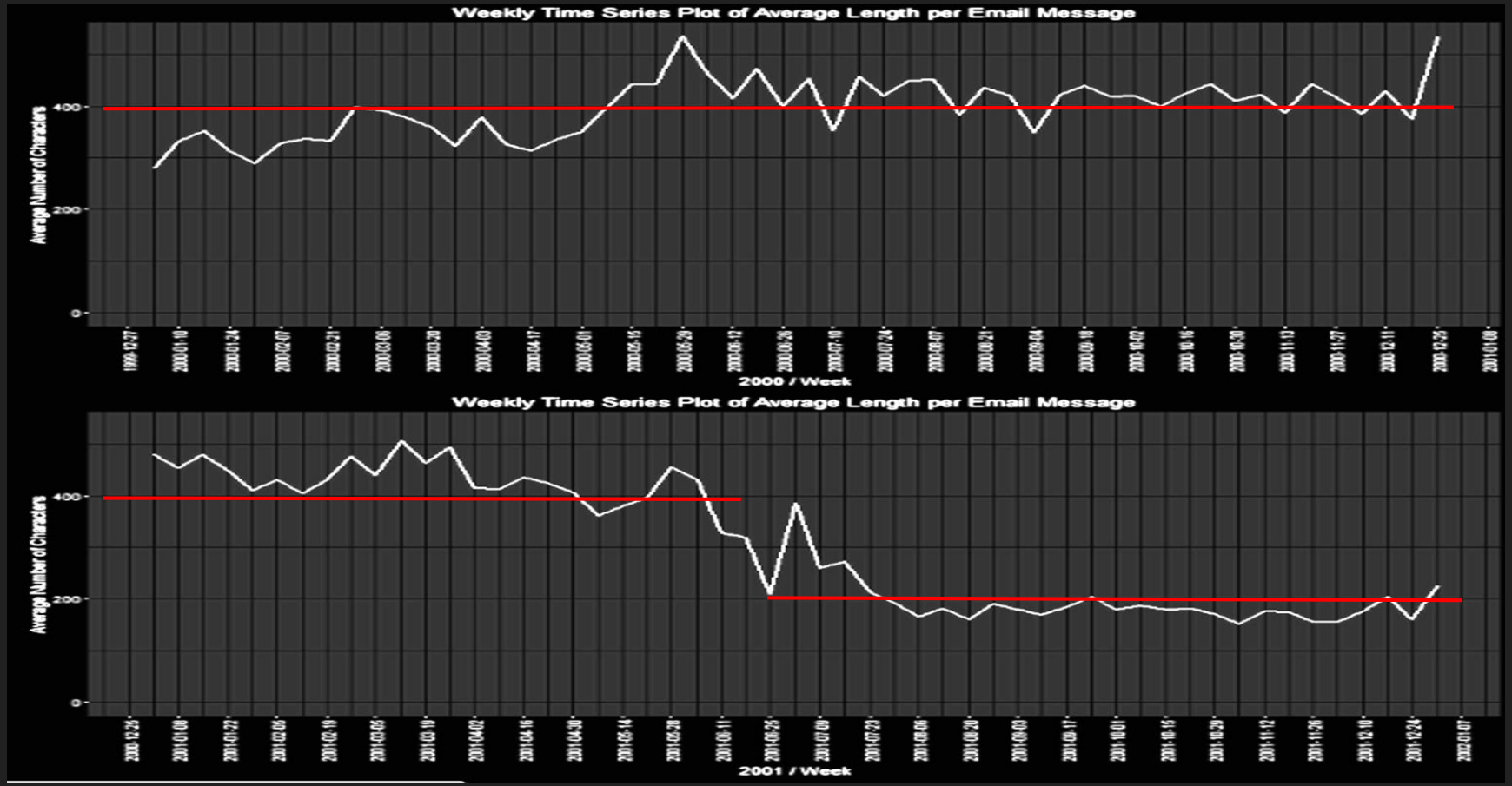
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Text, Sentiment, and RegTech

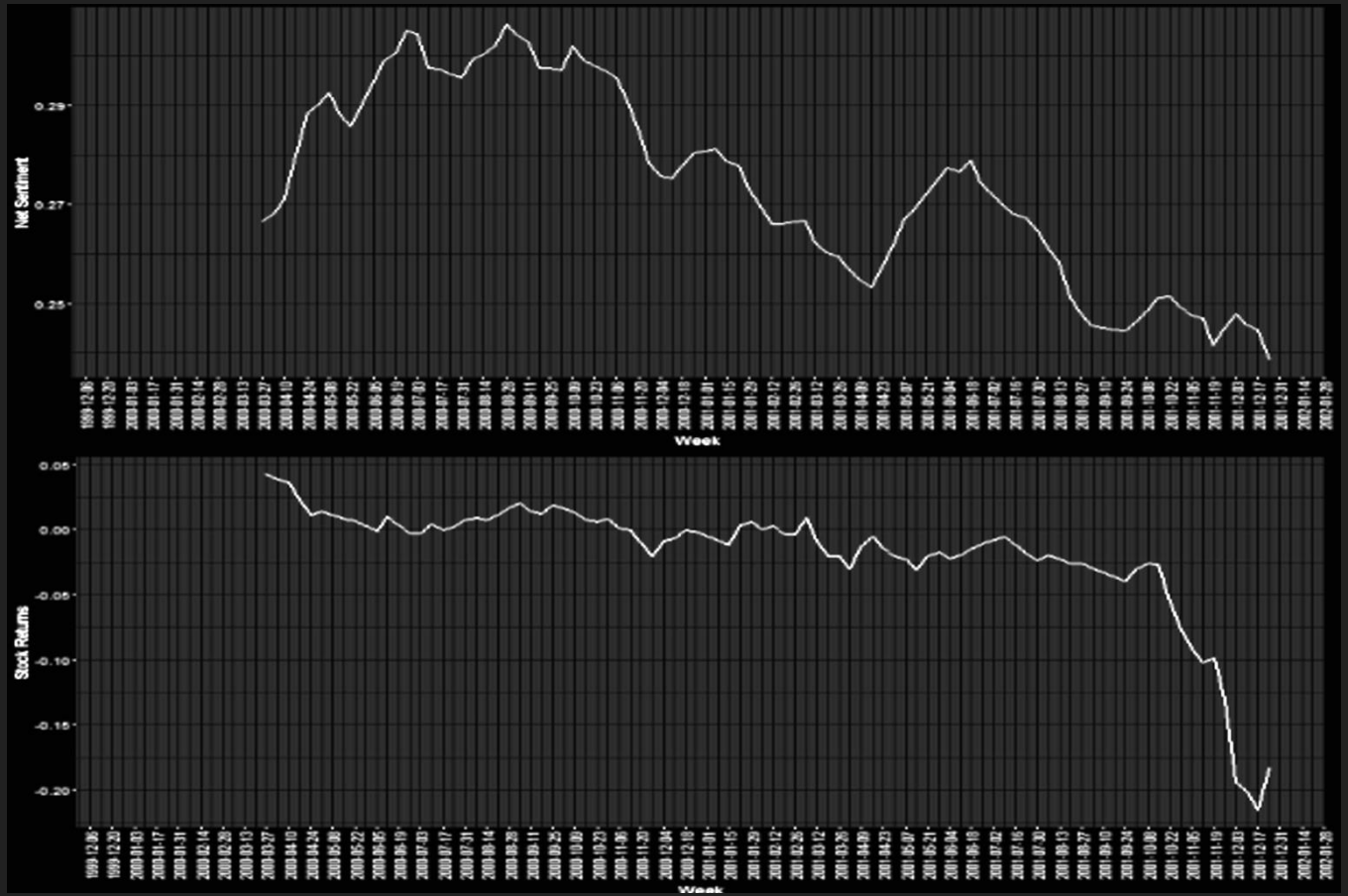
Zero-Revelation Linguistic Regulation: Detecting Risk Through Corporate Emails and News (Das, Kim, Kothari 2016)

- Financials are often delayed indicators of corporate quality.
- Internal discussion may be used as an early warning system for upcoming corporate malaise.
- Emails have the potential to predict such events.
- Software can analyze vast quantities of textual data not amenable to human processing.
- Corporate senior management may also use these analyses to better predict and manage impending crisis for their firms.
- The approach requires zero revelation of emails.

Enron: Email Length



Enron: Sentiment and Returns



Enron: Returns and Characteristics

Variable	Coefficient Estimate (*t*-statistic)			
	(1)	(2)	(3)	(4)
MA Net Sentiment $_t$	XXX***	0.575	2.330***	-1.397
	(XXX)	(0.63)	(3.14)	(-1.25)
MA Email Length $_t$		0.584***		1.046***
		(2.97)		(4.19)
MA Total Emails $_t$			-0.004	-0.131***
			(-0.10)	(-2.83)
Intercept		-0.406*	-0.671***	0.117
		(-1.93)	(-3.08)	(0.43)
Adjusted *R*-squared	XXX		0.09	0.24
Number of observations	88	88	88	88

Enron: Returns and Characteristics

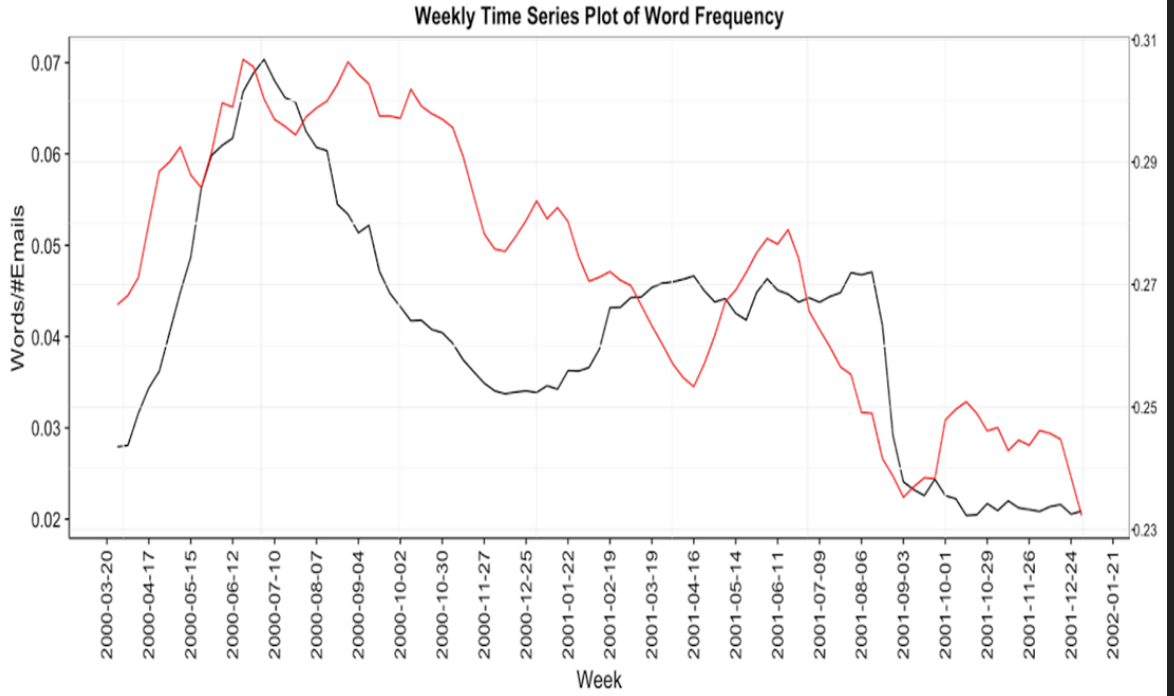
Variable	Coefficient Estimate (*t*-statistic)			
	(1)	(2)	(3)	(4)
MA Net Sentiment $_t$	XXX***	0.575	2.330***	-1.397
	(XXX)	(0.63)	(3.14)	(-1.25)
MA Email Length $_t$		0.584***		1.046***
		(2.97)		(4.19)
MA Total Emails $_t$			-0.004	-0.131***
			(-0.10)	(-2.83)
Intercept		-0.406*	-0.671***	0.117
		(-1.93)	(-3.08)	(0.43)
Adjusted *R*-squared	XXX		0.09	0.24
Number of observations	88	88	88	88

Enron: WordPlay

Select Word

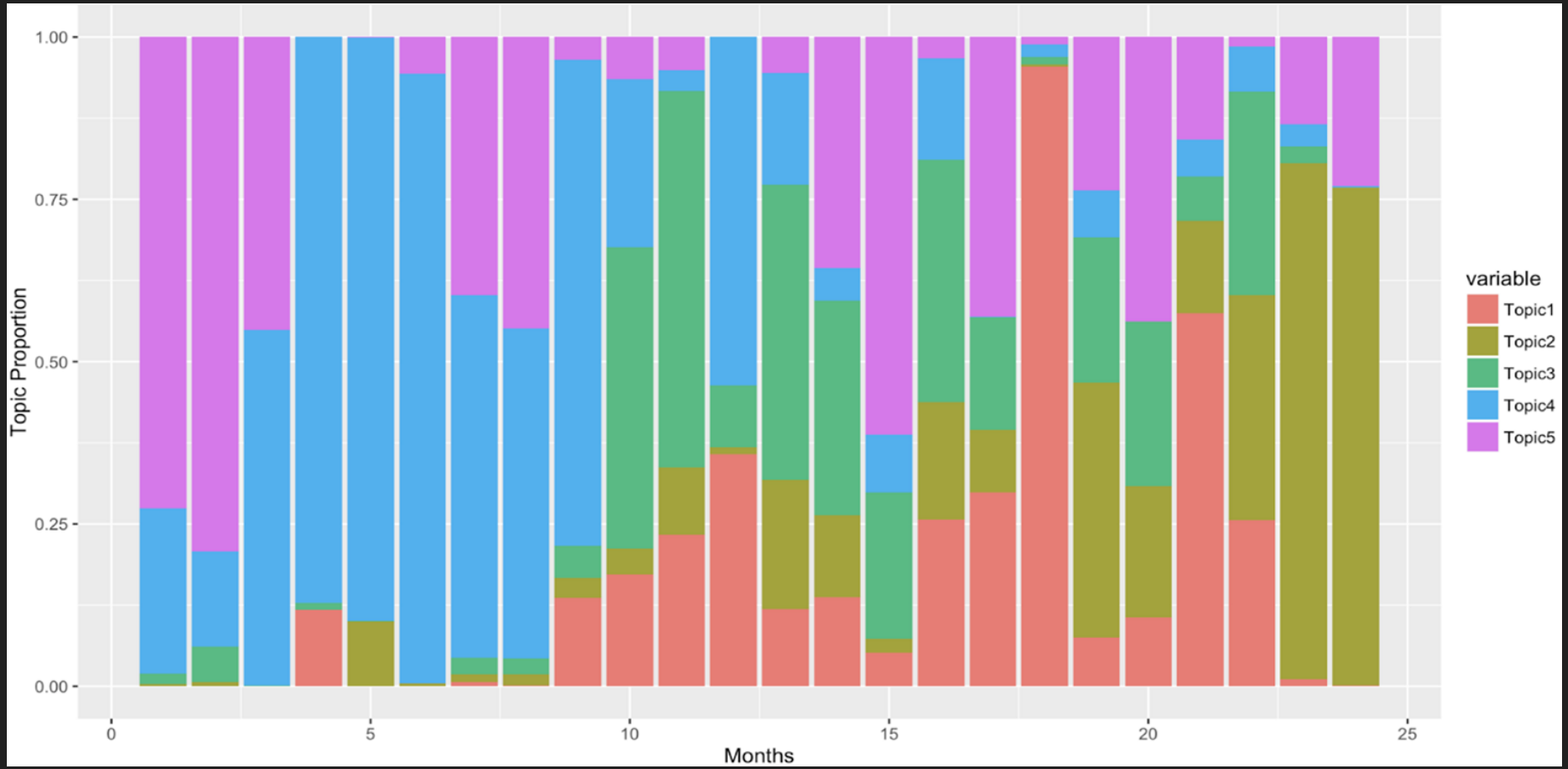
credit

Word over time | Word_vs_Sentiment | Word_vs_Return | Word_Sentiment_Correlation | Word_Return_Correlation



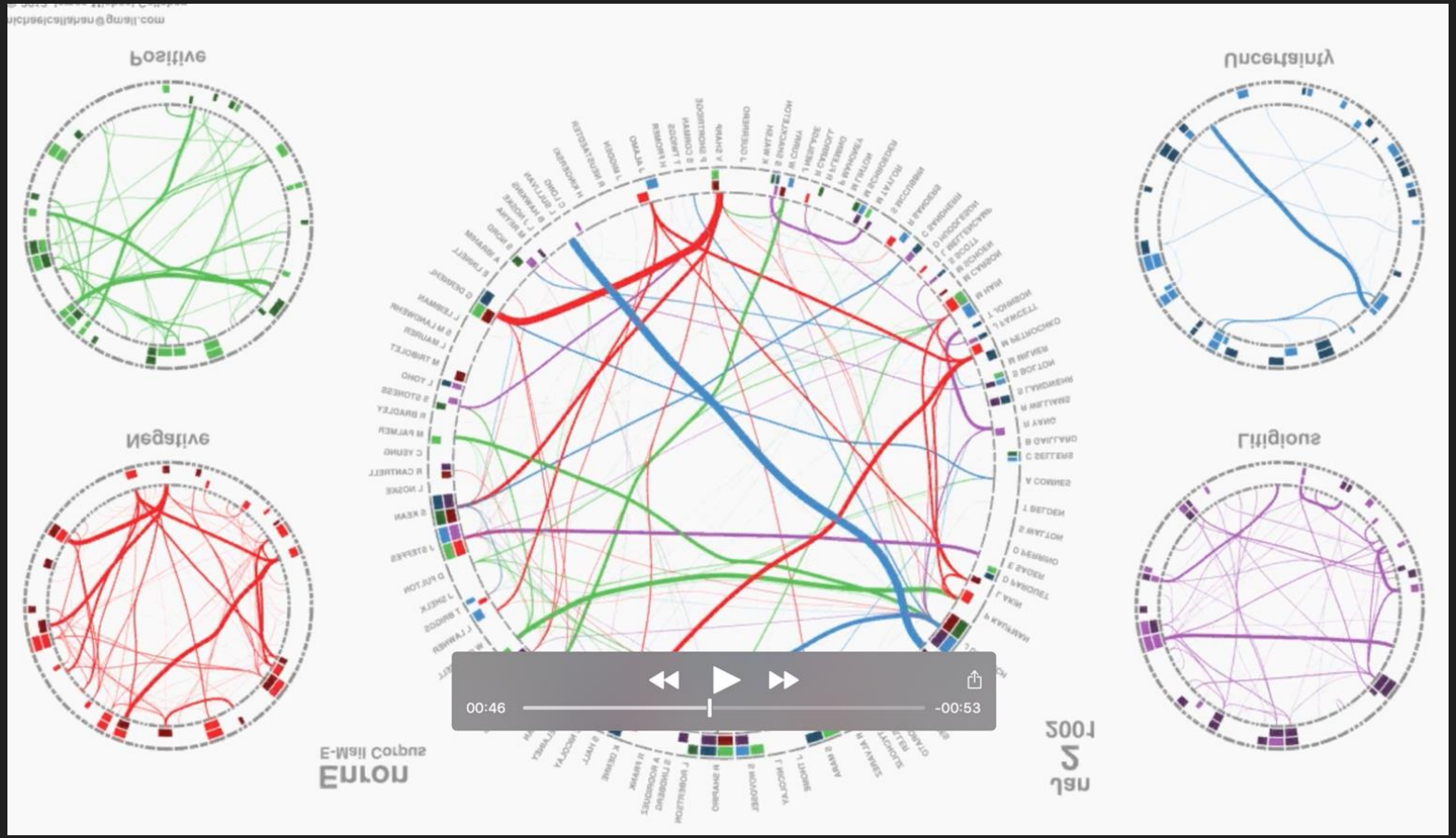
Black: words/#emails , Red: sentiment

Enron: Topic Analysis



Enron Movie (by Jim Callahan)

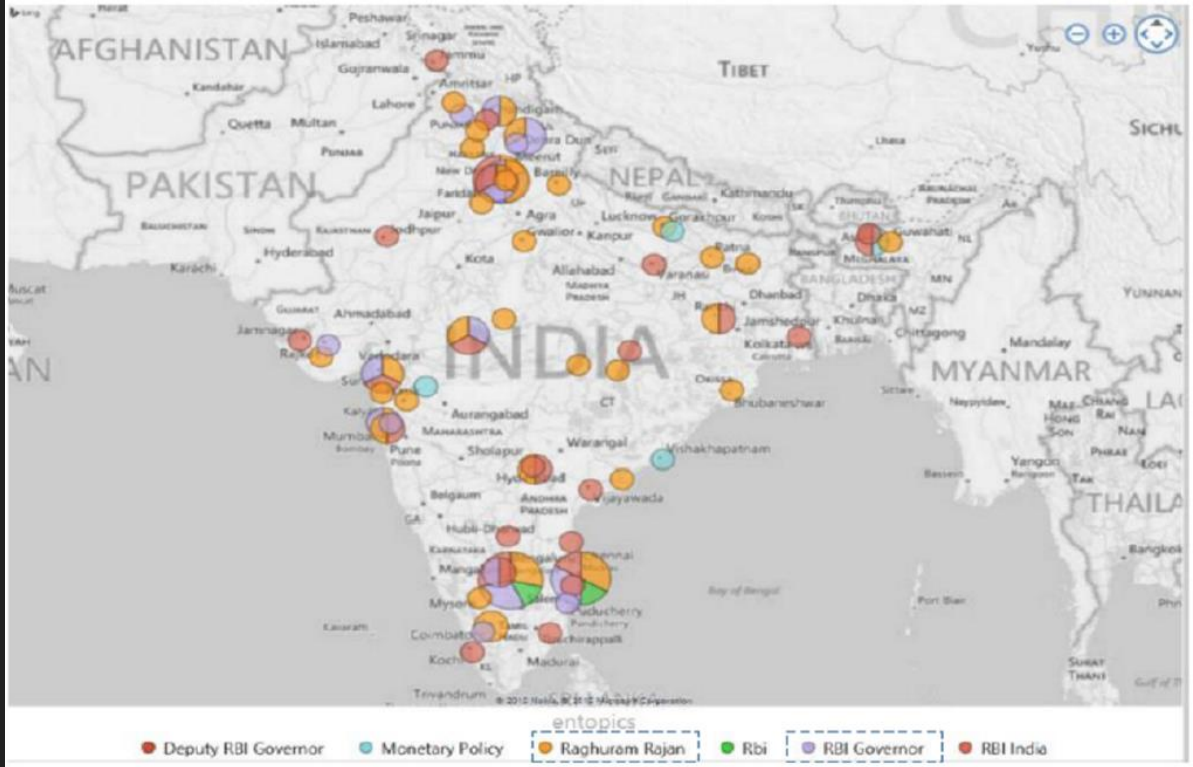
http://srdas.github.io/Presentations/JimCallahan_enron-sm.mov



India: Topic Analysis

Conversations across India and around RBI

topycs



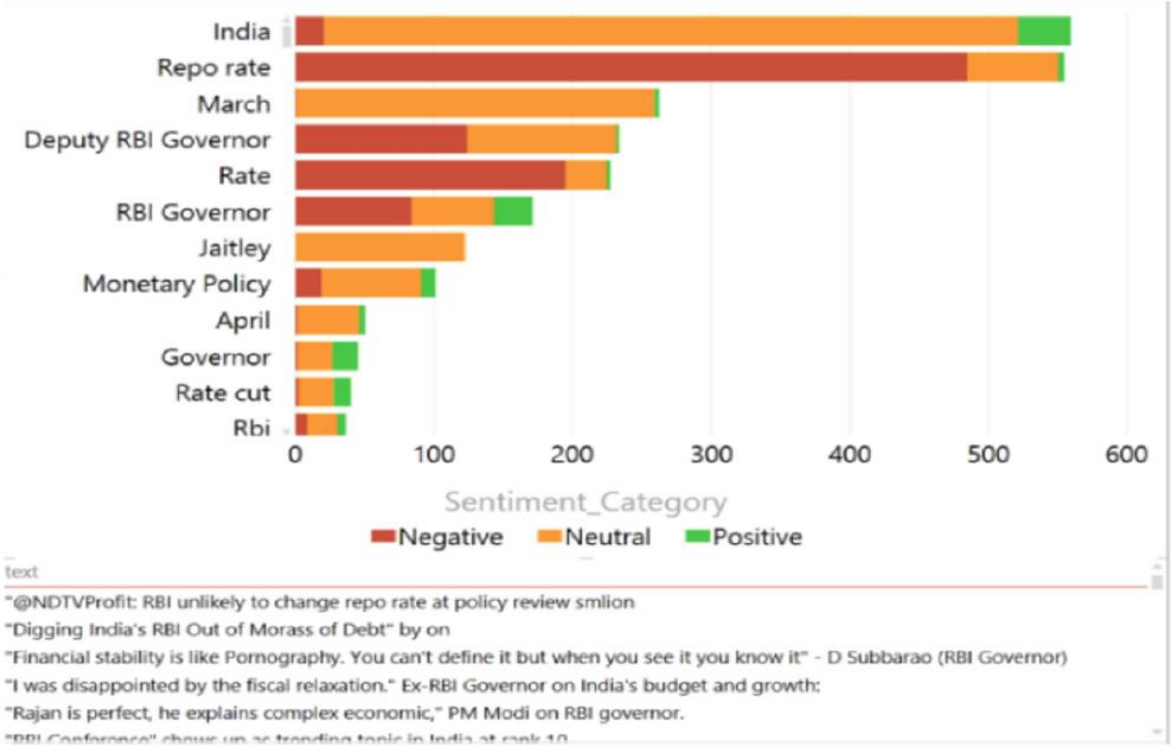
- Conversations across India on RBI, its people and the monetary policy
- Governor features in many conversations across both rural and urban areas
- Some conversations specifically around monetary policy

- Bubbles show split of conversations around Deputy RBI Governor, Monetary Policy, Raghuram Rajan, RBI and RBI Governor.
- Based on count of unique conversations
- Date Range: 1st – 14th April, 2015

India: Topic Analysis

Top Topics along with RBI

topycs



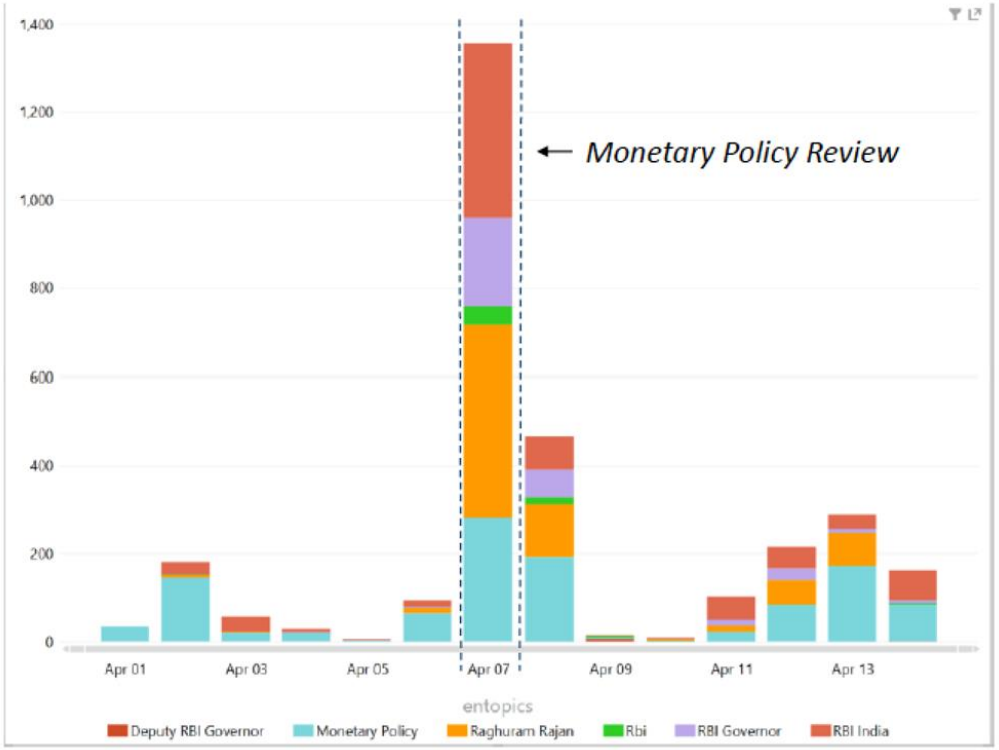
- Repo rate evokes negative sentiment as people don't expect it to be changed
- Repo rate, rate cut and monetary policy are discussed frequently with RBI

- Vertical Axis – Topics of Discussion
- Horizontal Axis – Count of Unique Conversations
- Date – 25th March - 14th of April
- Colors represent sentiment for conversation, Negative – Red, Neutral – Orange, Positive - Green

Monetary Policy (India)

Largest number of conversations around the date of the policy review

topycs



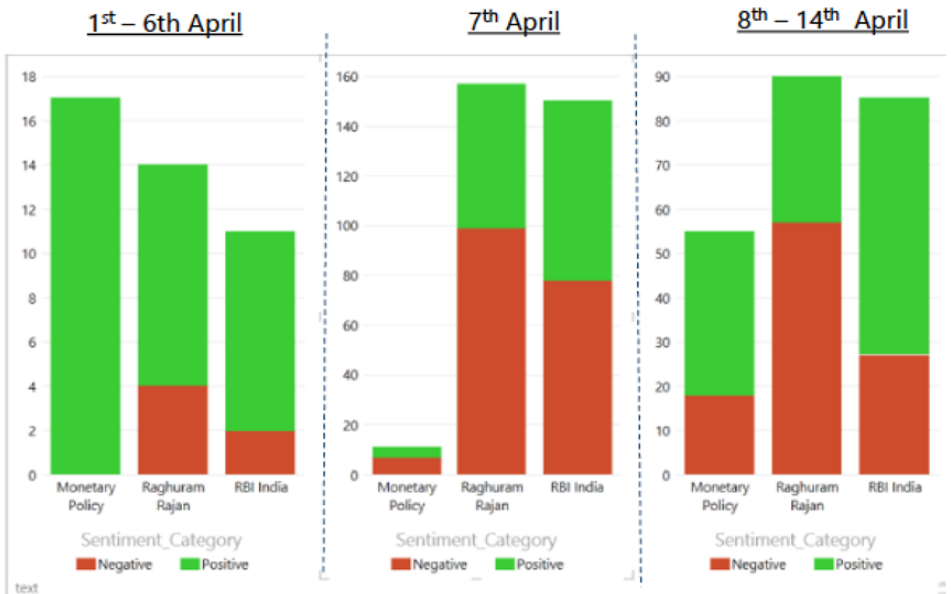
- Largest number of conversations on 7th April and the day after
- People are talking about the monetary policy prior and after the review
- RBI and Governor show up only around the review

- Vertical Axis – Count of Unique Conversations
- Horizontal Axis – Dates between 1st – 14th April
- Colors represent topics of conversation

India

Sentiment around key topics before, during and after the review

topycs



text

"@IndiaToday: SBI, ICICI, HDFC Bank pay heed to Raghuram Rajan's call, cut base rate trending
"@timesofindia: Top 3 banks cut rates after RBIs tough talk" what about you?
"Effectively, monetary policy transmission may move from market forces to fiat which w ... - NewsInTweetsIndia
"Effectively, monetary policy transmission may move from market forces to fiat which would be regressive."
"India Inc disappointed with RBI's move on policy rates - SME Times"
"Making monetary policy more potent"
"RBI Conference" shows up as trending topic in India at rank 10

- Conversations positive around monetary policy prior to review, but show more negativity after the review
- More conversations around Raghuram Rajan and RBI India on day of and after the review than prior to the review

- Vertical Axis – Count of Unique Conversations
- Horizontal Axis – Topics of Conversation
- Date – 1st – 14th April
- Colors represent sentiment for conversation, Negative – Red, Positive - Green

Key Principles in Using Big Data for Finance

- Using Theory to develop models to apply to Big Data.
- Questions/problems are primary, data is secondary, in the success of FinTech ventures.
- Simplicity, transparency of models fosters implementability.
- Analytics per se is multidisciplinary.
- Disparate data is the norm.
- Significant investment in hardware and talent.

Thank You!!

http://srdas.github.io/Papers/India.pdf
http://srdas.github.io/Papers/JAI_Das_issue.pdf
http://srdas.github.io/Papers/EnronZeroRev.pdf
http://srdas.github.io/Papers/dyn_syst.pdf

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Exploiting big data for sharpening financial sector risk assessment[1]

Kimmo Soramäki,

Financial Network Analytics

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

FNA

Exploiting Big Data for Sharpening Financial Sector Risk Assessment

Dr. Kimmo Soramäki
Founder & CEO, FNA

www.fna.fi

Agenda: Three Examples

Advanced Analytics	Monitoring	Simulation
1	2	3
Measuring interconnectedness of global CCPs	Identify Liquidity and Solvency problems from payments	Operational failure of an FMI member bank
Identify Risks Concentrations	Provide Early Warning	Predict Outcomes

Motivation

The **New Systemic Risk**

Three CCP failures in the past (Paris, Kuala Lumpur and Hong Kong)

Interest by regulators, CCPs and members.

Especially with tie in to Cyber, IT and other operational risks.

"*They [CCPs] are not equipped, however, to test the impact of their failure on the financial system as a whole nor are they equipped to assess the potential contagion effect on other members of a given member's failure.*"

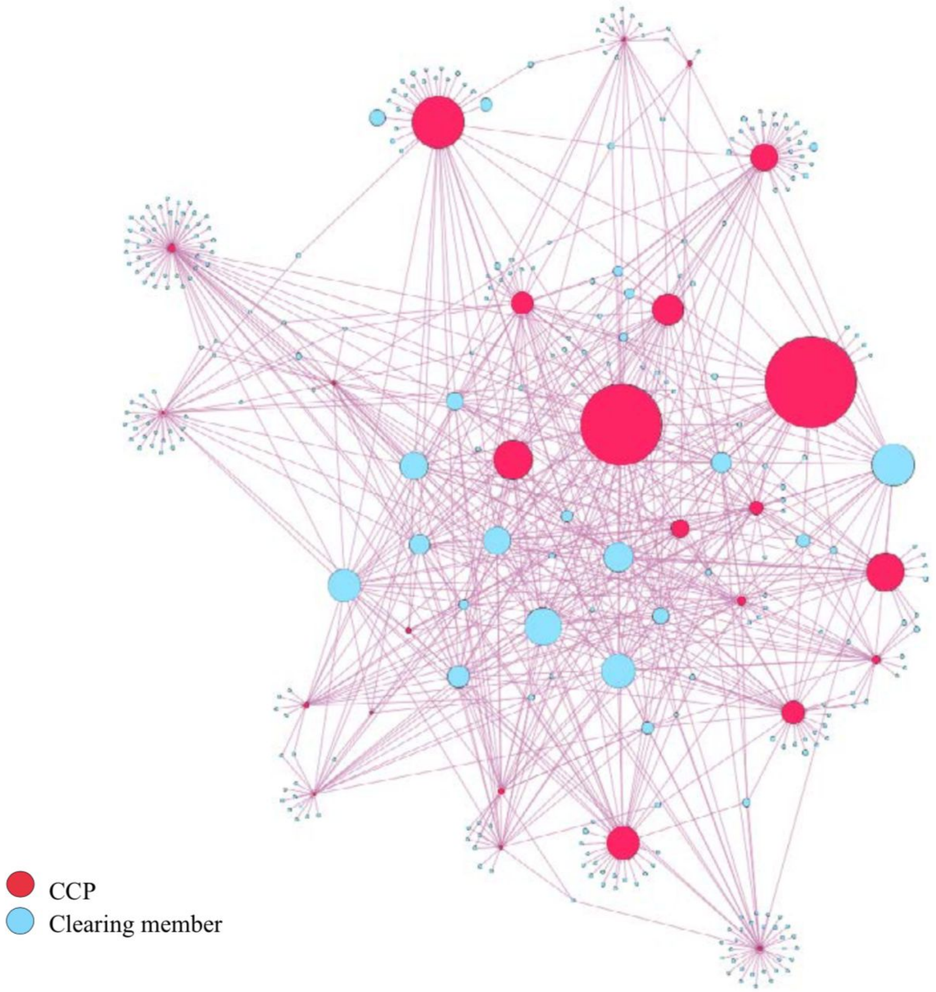
Cox & Steigerwald (2018)

Scope of Analysis

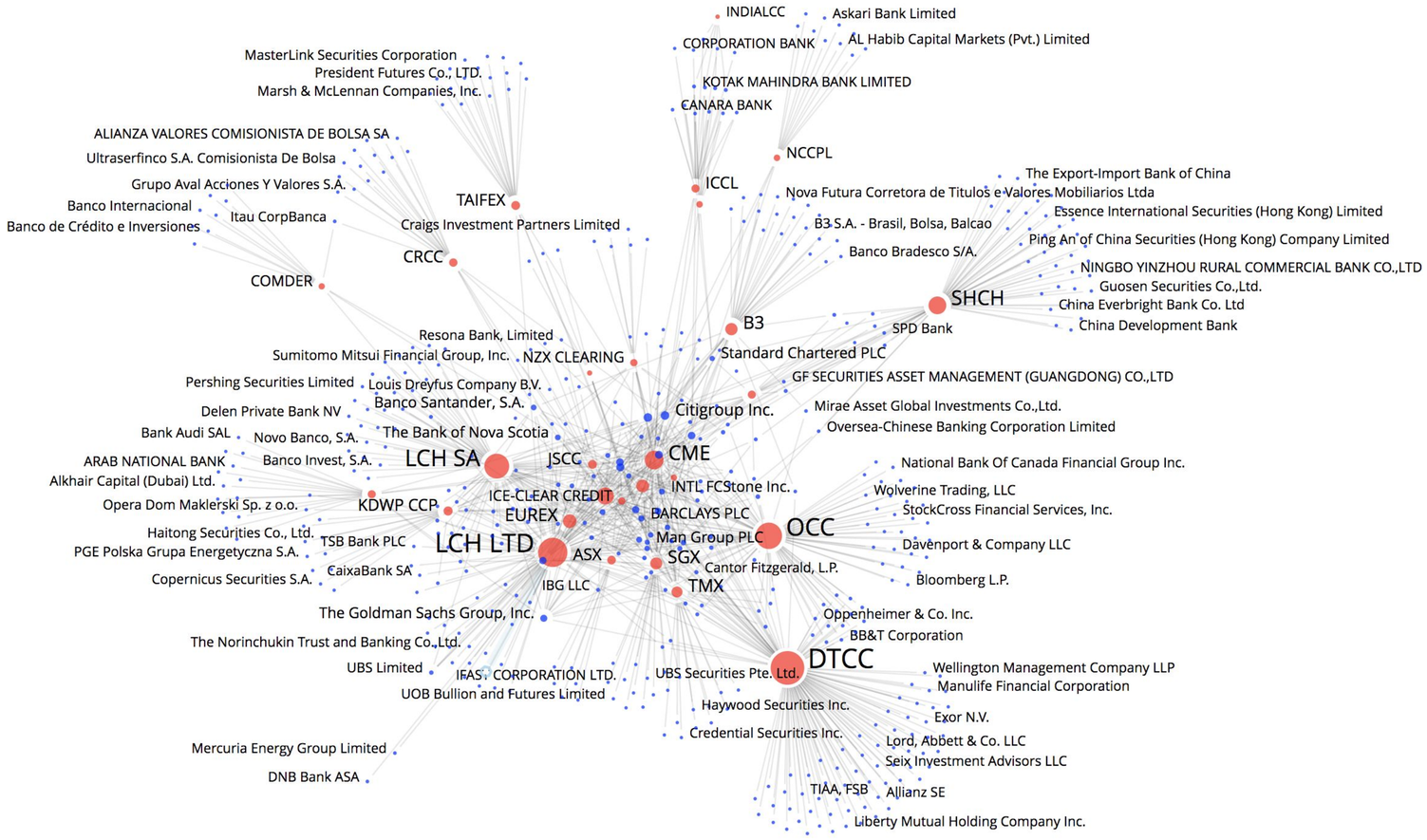
Comparison with BIS "Analysis of Central Clearing Interdependencies" (2017)

	BIS (2017)	**FNA (2018)**
CCPs	26	29
Jurisdictions	20	25
Clearing Members	n/a	811
Parents Organizations	307	563
Roles	5 (member, settlement, LOC, ...)	1 (member)

Private vs Public Data



BIS (2017)



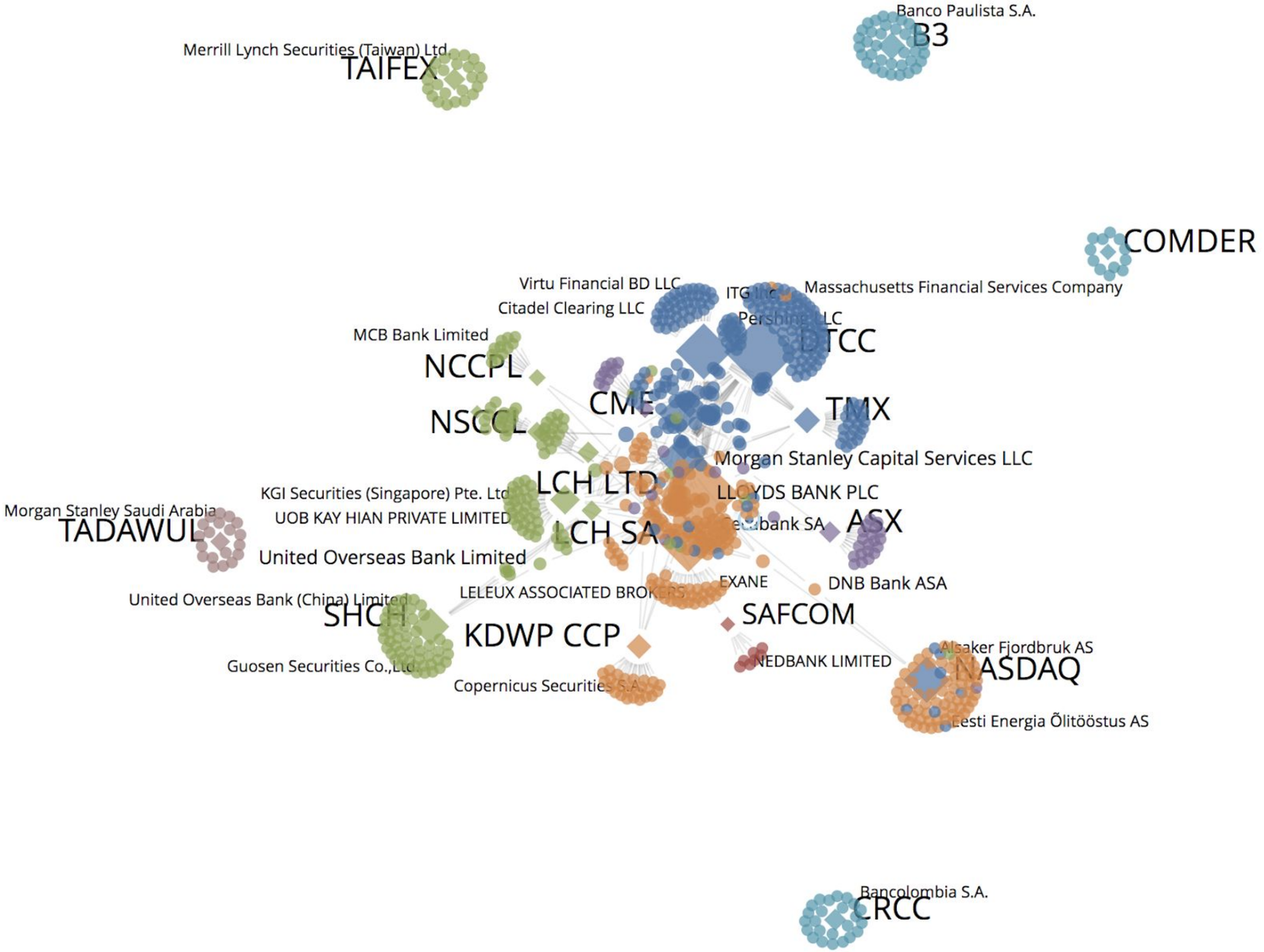
FNA (2018)

CCP Interconnectedness - Subsidiary Level

We see CCPs (diamonds) and their members (circles) from different regions:

- North America (blue)
- Europe (Yellow)
- Asia (green)
- Middle East (brown)
- Latin America (blue)
- Australia & Oceania (purple)

On subsidiary level, we see a tight core with peripheral CCPs and a number of completely disconnected CCPs from Latin America and Middle East.



Banking Groups

210 Banking Groups

Largest (# of entities):

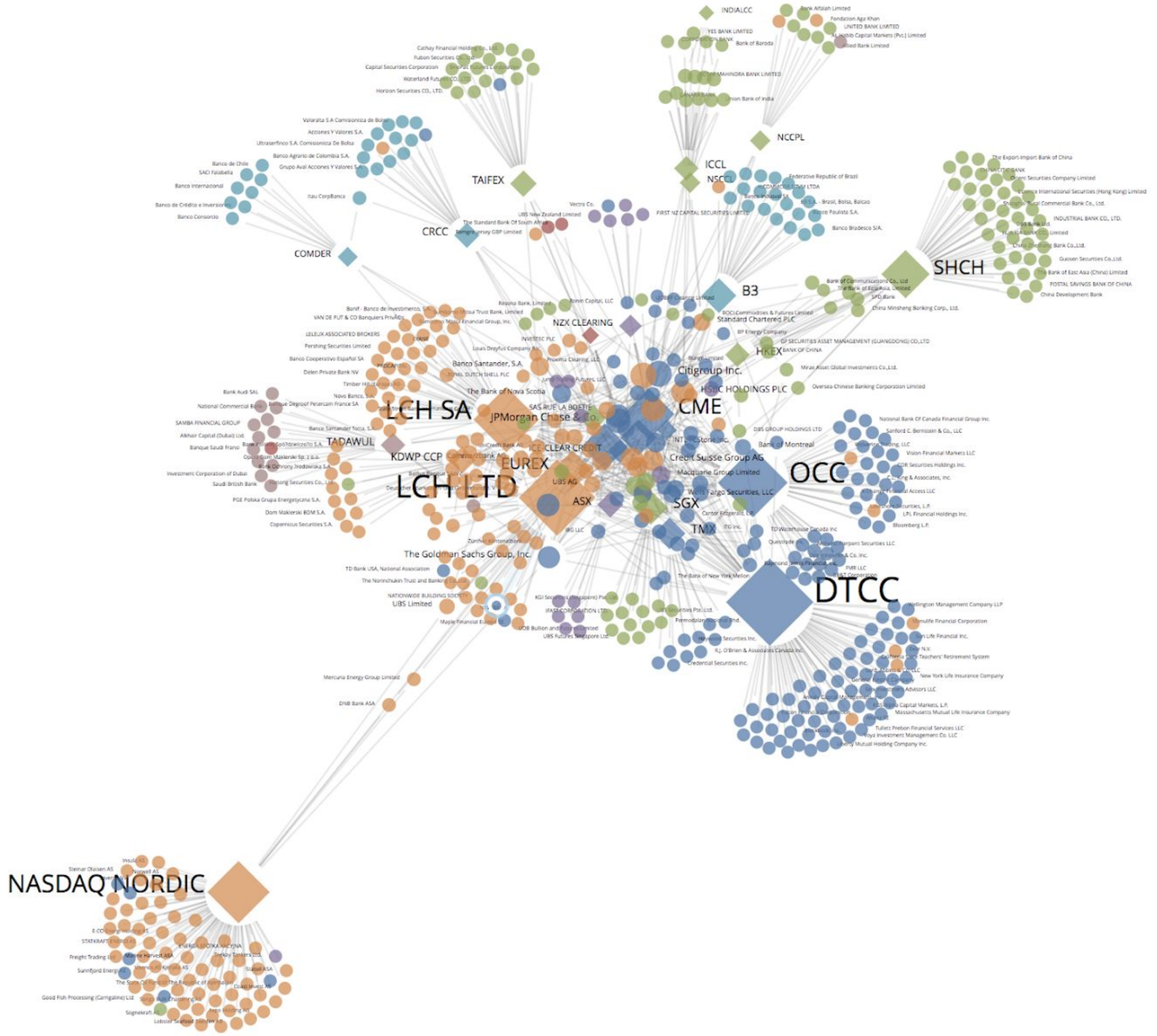
1. Citigroup (19)
2. Morgan Stanley (13)
3. Goldman Sachs (12)
4. JPMorgan Chase (12)
5. Bank of America (12)
6. HSBC (11)
7. Credit Suisse (10)
8. Deutsche Bank (10)
9. Nomura (9)
10. Banco Santander (8)

CCP Interconnectedness on Parent Level

We see CCPs (diamonds) and their members (circles) from different regions:

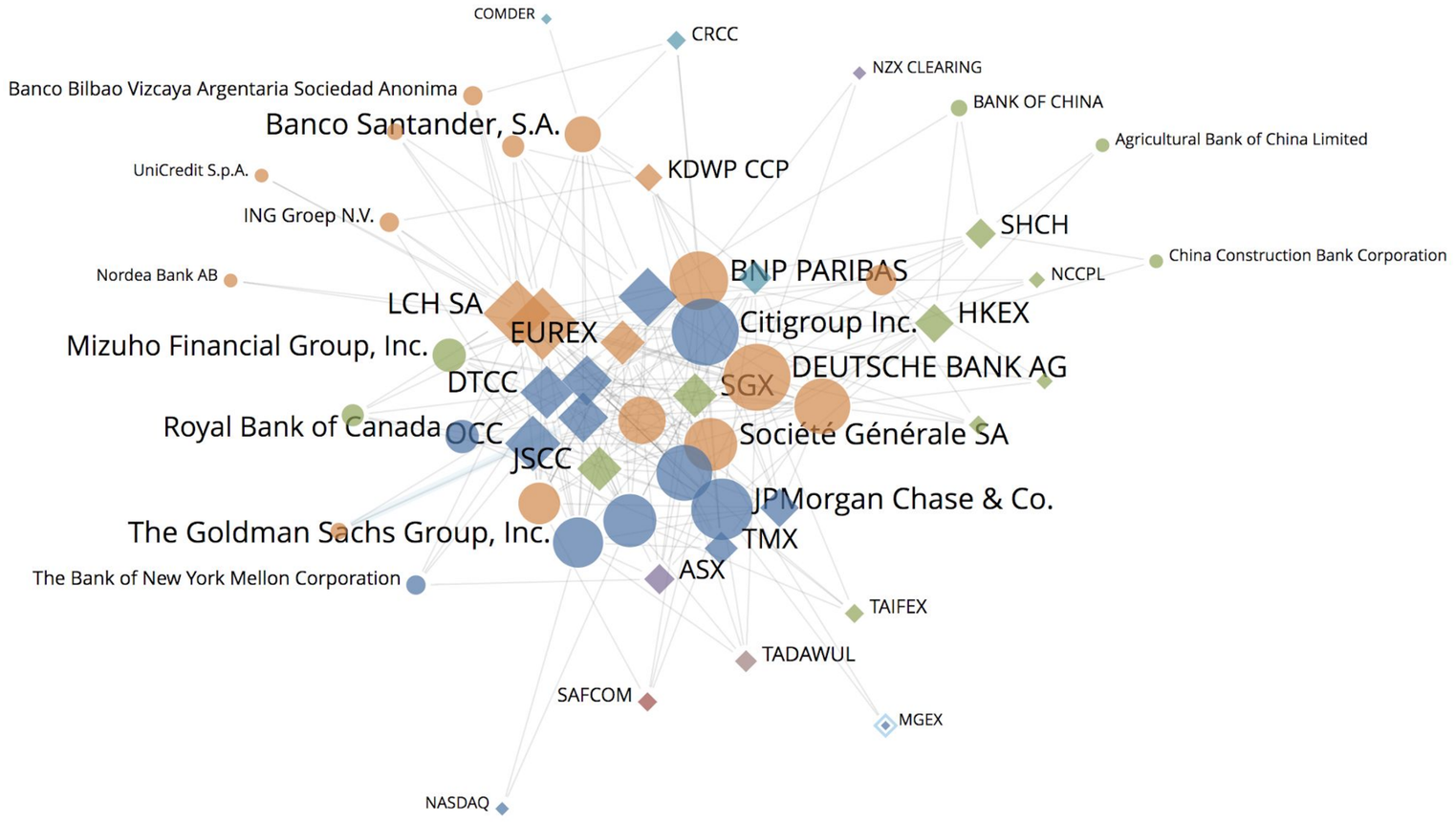
- North America (blue)
- Europe (Yellow)
- Asia (green)
- Middle East (brown)
- Latin America (blue)
- Australia & Oceania (purple)

On parent level we see a completely connected network dominated by a core consisting of CCPs from North America and Europe and global banks.



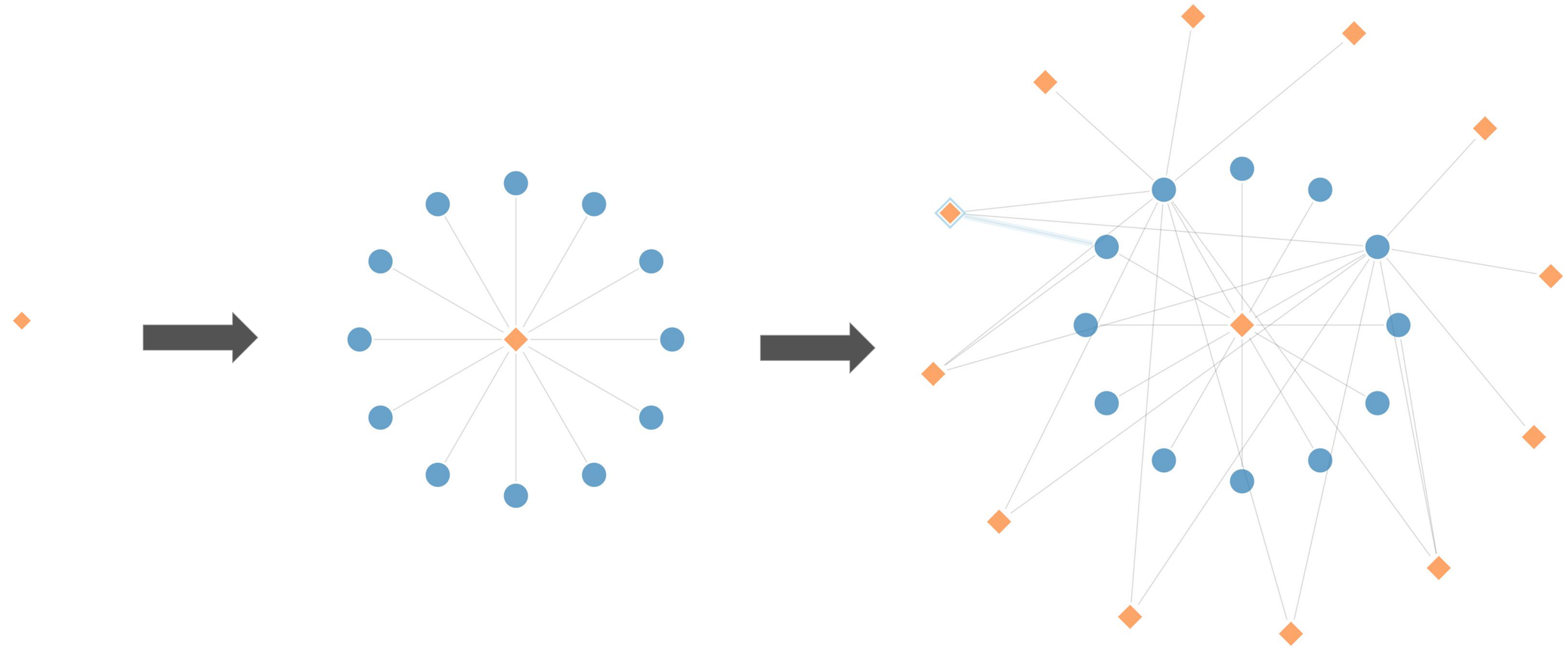
CCP Interconnectedness on GSIB Level

Bank (Parent)	# of FMIs
Citigroup	21
DEUTSCHE BANK	21
JPMorgan Chase & Co.	19
BNP PARIBAS	18
Bank of America	17
HSBC	17
Morgan Stanley	16
Societe Generale	16
The Goldman Sachs	15
Credit Suisse	14



Contagion - CCP Disruption

A disruption in a CCP would affect all of that CCP's clearing members, thereby affecting the other CCP's to which the affected CCP's members belong, possibly creating a cascading cycle as disruption is propagated across members and CCPs



CCP disruption

Propagates to all members

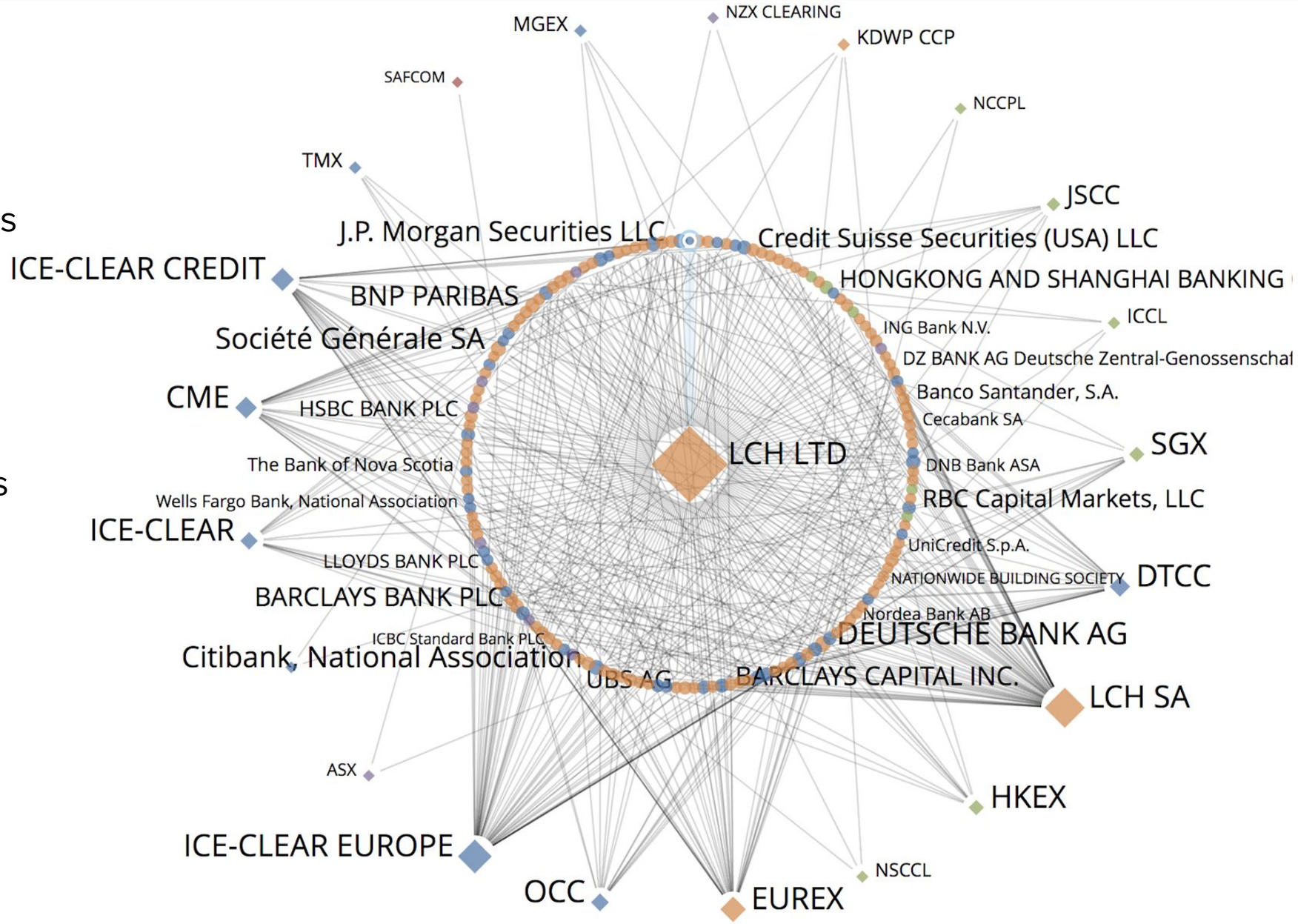
Propagates to other CCPs

Footprint of CCPs - LCH Ltd

LCH Ltd 100 members are connected to 27 other CCPs

The membership is mostly European with a significant US base.

The most connected CCP is LCH SA and ICE-CLEAR EUROPE.



Contagion – Member Disruption

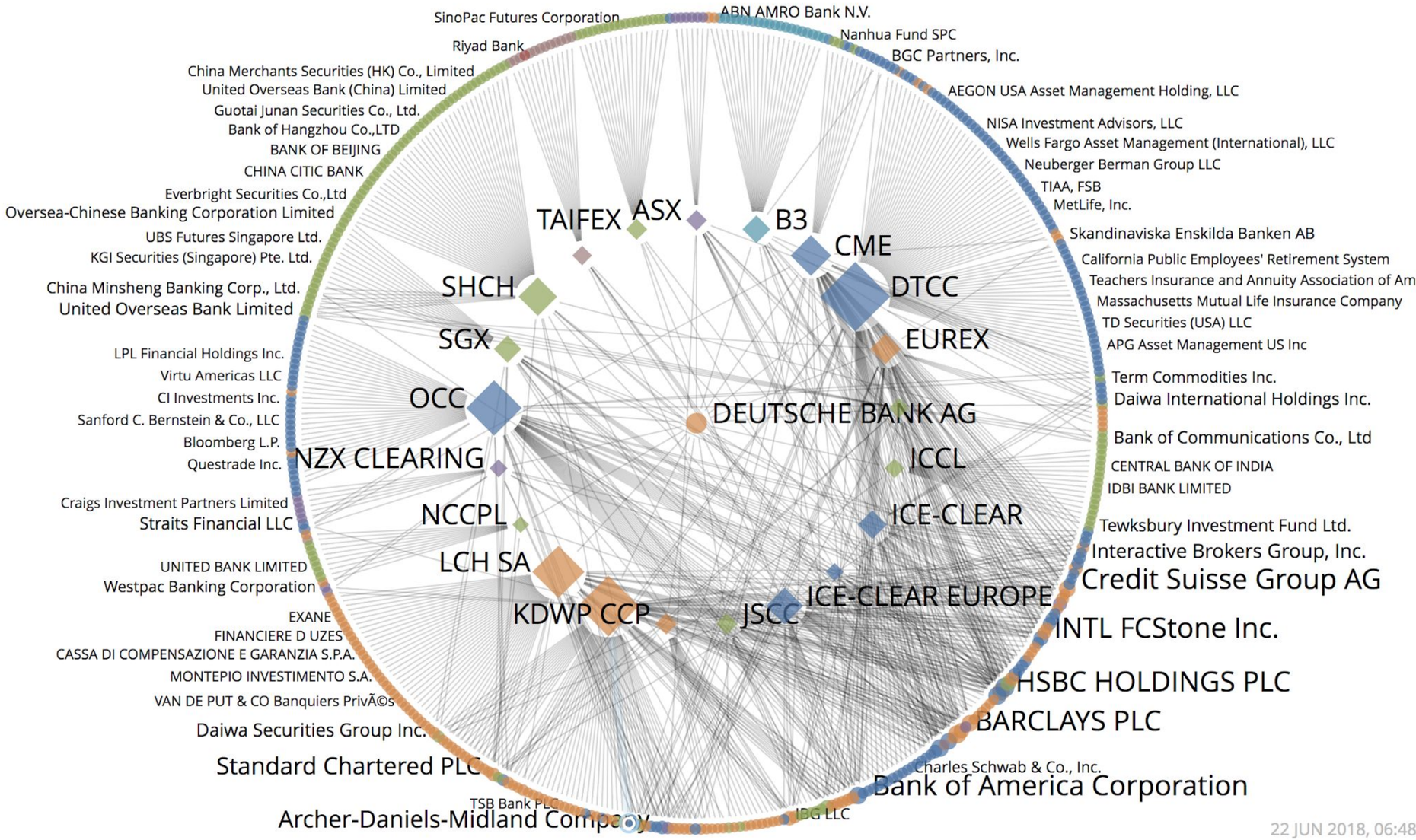
A member disruption can be felt by up to **458** banking groups or banks (of total of 563, or 80%) that are members of the same CCP as the stricken group.

Banking Group	**# banking groups connected via a CCP**
Deutsche Bank	458
Citigroup	446
Morgan Stanley	442
BNP Paribas	423
Goldman Sachs	412
HSBC Holdings	402
JPMorgan Chase	388
Bank of America	382
Credit Suisse	348
Société Générale	340

Contagion – Member Disruption

Deutsche Bank Group participates in 21 CCPs (of 29 mapped).

458 other banking groups or banks are members of these CCPs.



Journey

Advanced Analytics

1

Understand interconnectedness, data visualization

Identify Risks Concentrations

Monitoring

2

Detect risks and anomalies in real-time

Provide Early Warning

Simulation

3

What-if analysis, failure simulations & remediation scenarios

Predict Outcomes

Use Case: Monitoring Liquidity and Solvency of FIs

Central Bank of Colombia identifies early warning on liquidity and solvency of financial institutions with FNA

Background

The Central Bank of Colombia has been using balance sheet and regulatory reporting data to understand the liquidity and solvency of participants in the Colombian financial system. However, the analysis is time consuming and the data comes months late.

Objective

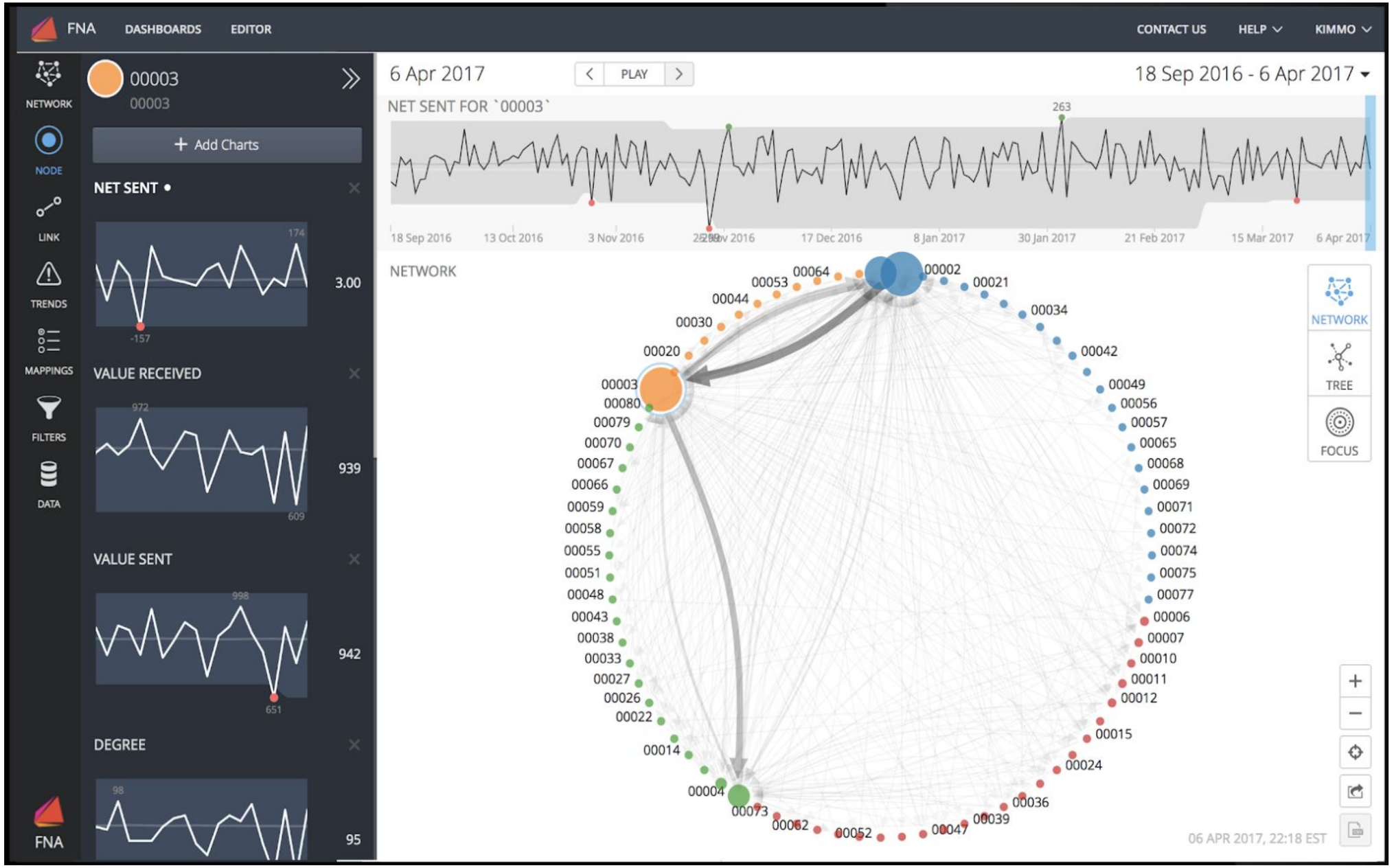
Using network analysis of data from the interbank payment system would allow the Bank to get early warning about risks substantially faster.

Outcomes

Using the FNA Platform, the Bank is now able to monitor its banking system in near real time. Automatic alerts notify the bank of any abnormal behavior in the network. Furthermore, automated stress tests where they fail the two largest participants in the network help to understand the riskiness of the system.

Carlos León et al (2015). Assessing Systemic Importance With a Fuzzy Logic Inference System.

Use Case: Monitoring Liquidity and Solvency of FIs



Journey

Advanced Analytics	Monitoring	Simulation
1	2	3
Understand interconnectedness, data visualization	Detect risks and anomalies in real-time	What-if analysis, failure simulations & remediation scenarios
Identify Risks Concentrations	Provide Early Warning	Predict Outcomes

Concept: Operational Failure of a Settlement Member

Mapping
This network shows settlement relationships between the:

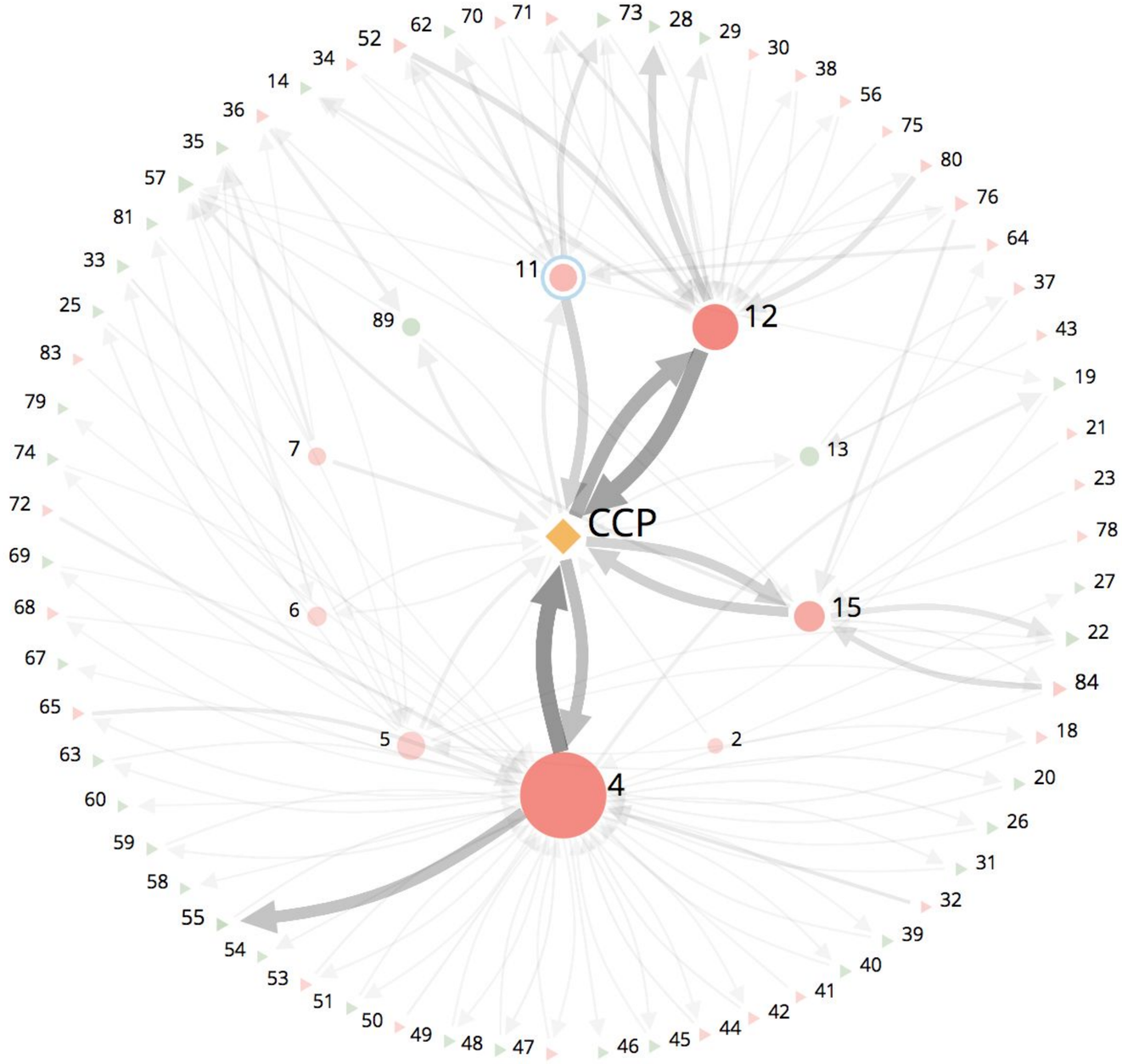
- CCP (center)
- Settlement members (inner circle) and
- Clearing members (outer circle)

Note: Data is representative, not real

Size of node shows value of multilateral position

Width of lines shows value of bilateral positions

Question
What would happen if member 4 had an operational failure?



Backup Relationships

Map

Shows Clearing Members on the left, and Settlement Members on the right.

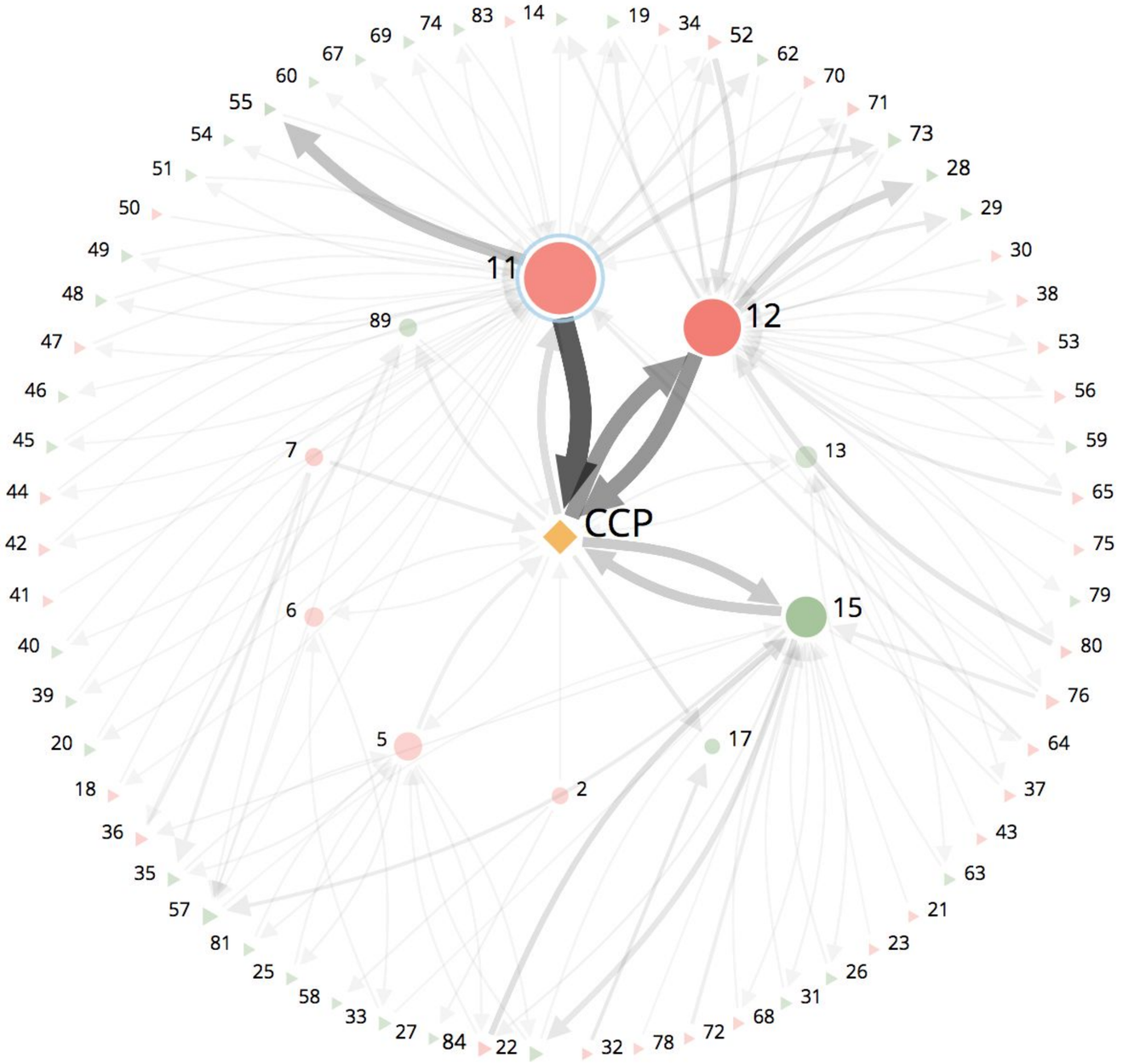
The lines denote which settlement member the clearing member can use for settlement (ie its main and its backups)

Rewiring for Maximum Concentration

Each clearing member using Bank 4 must now effect settlement through one of its backup relationships.

Findings
Simulation shows that settlement flows could be concentrated on a few participants, e.g. causing operational challenges for Bank 11.

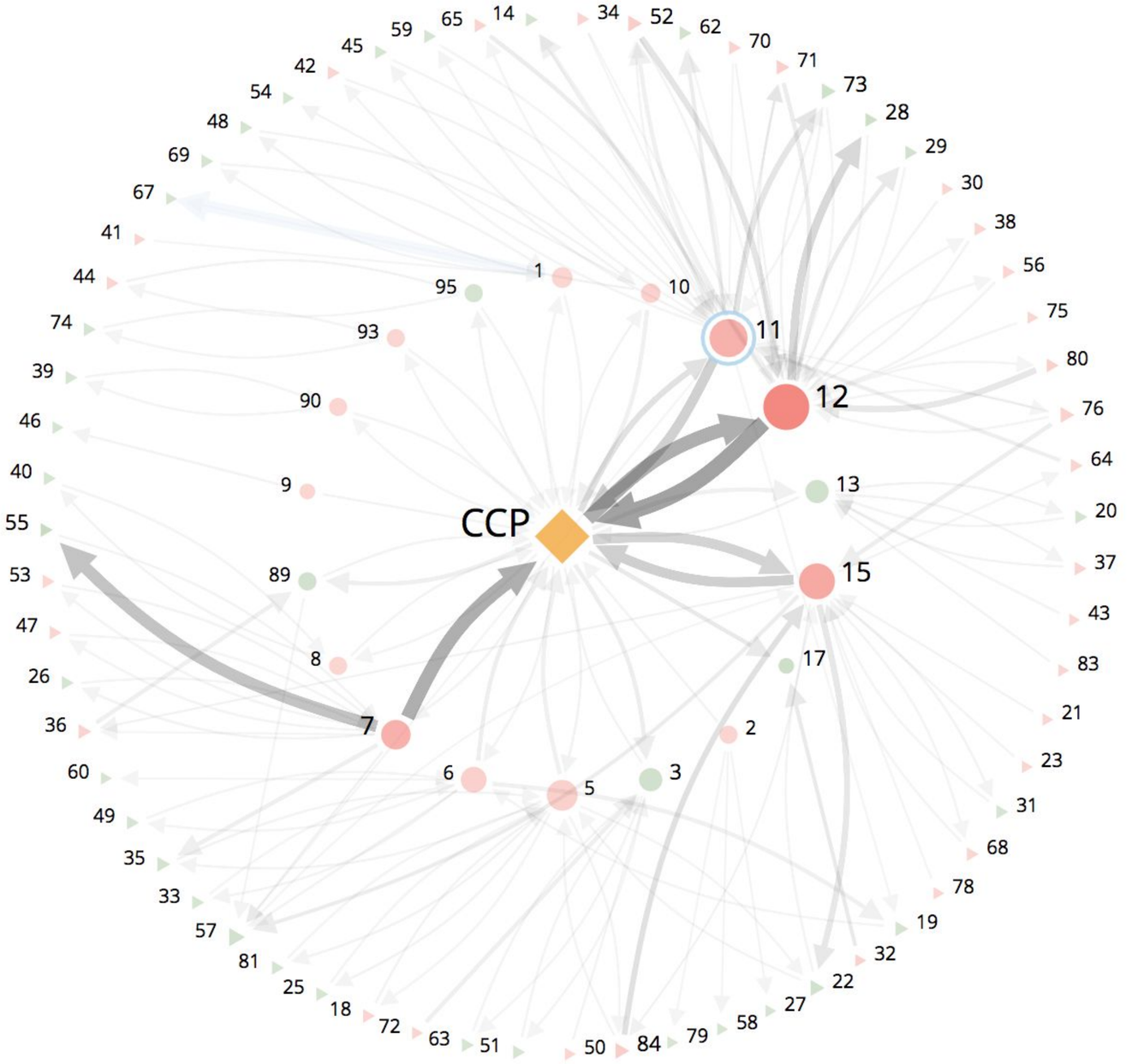
Insight
Bank 11 was not among the most active settlement members on a normal day, but might need to build operational capacity to cover for rare failure days.



Rewiring for Minimum Concentration

Findings
… or clearing members might use different settlement members resulting in a much higher number (18 instead of 10) of settlement members for the day.

Insight
The CCP may need to build operational capacity to be able to complete settlement.



FNA

Dr. Kimmo Soramäki
Founder & CEO
FNA - Financial Network Analytics Ltd.

kimmo@fna.fi
tel. +44 20 3286 1111

Address
4-8 Crown Place
London EC2A 4BT
United Kingdom

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

How do central banks use Big Data to craft policy?[1]

Per Nymand-Andersen,

European Central Bank

[1] This presentation was prepared for the meeting. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Per Nymand-Andersen
Adviser, European Central Bank

How do central banks use big data to craft policy?

Bank Indonesia - IFC/BIS seminar "building pathways for policy making with big data"

July 26, 2018

Disclaimer: The opinions expressed in this presentation are not necessarily those of the European Central Bank (ECB) or the European System of Central Banks (ESCB)

Data never sleeps

Which preparations are needed today to have the capacity and functionality required in 3 years time?

- From experimenting to central banking tool kits?
- Linking current and past data
- Querying variety of formats
- Analytical techniques & tools
- Technical independent
- Skill sets

Source: 2018 Update: What Happens In An Internet Minute, Lori Lewis[2]

Data mania versus phobia – a paradigm of borderless records

Digital transformation in finance and economics

E- trade

Settlement systems

Credit cards

Mobile transaction

Lending & financing

Big data

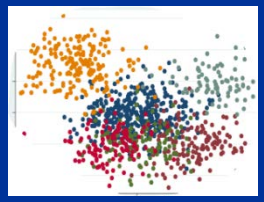
S-media

Price scans

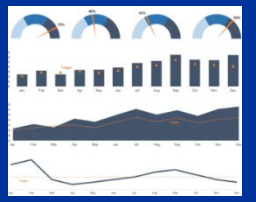
- Fintech
- Virtual tokens
- DLT
- S-contracts

Data lab

Systematic acquire, Structure, Process,



Statistical algorithm and data explorations

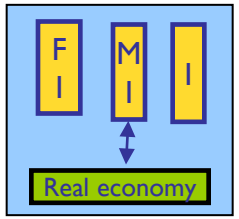


Packaging data for Insights & business

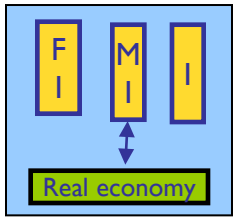
Impact of digital transformation on Monetary Policy & Financial Stability

Other systems ⟷ **Financial system** ⟷ **Real economy**

Financial systems



Financial systems



Other sources

Policy events,
fiscal policies,
Global economy
External chocks
Sector failures

Transmission of central banking policies, managing expectations, feed-back loops

Regulatory and oversight framework

Financial Intermediaries (FI)

Banks
Insurance, pension
hedge funds,
Rating agents, others

Markets and instruments (MI)

Securities,
Derivatives
and other products

Financial infrastructures (I)

Pre-trading,
Trading,
post-trading
Infrastructures
Clearing,
settlements

Raw sources:

Pre-trading, trading and post trading activities, actors, instruments, platforms, volumes, prices, frequencies, maturities, regulators and overseers alike

Derived indicators:

Yield curves, density, liquidity, price measures,
Functioning of markets, instruments and post trading activities
Risks measures and contagions indicators
Imbalances and concentrations measures
Effects on performance (Collective behaviour, Interconnections networks, transmission of instability measures

CORPORATE SECTOR

HOUSEHOLD SECTOR

GOVERNMENT SECTOR

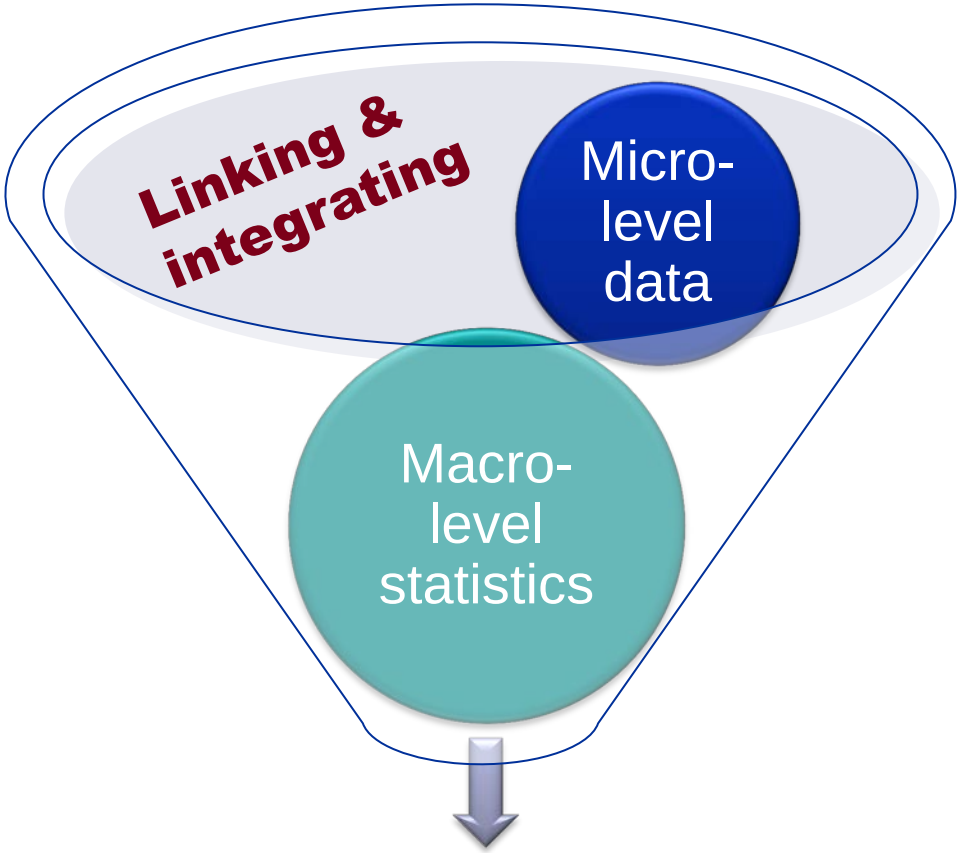
FOREIGN SECTOR

Reflections for central banking policy purposes

Source: **"Big data: The hunt for timely insights and decision certainty - Central banking reflections on the use of big data for policy purposes**, IFC working Paper No 14, 2016, Per Nymand-Andersen

Moving to micro level and granular data

Central bankers are collecting structured and standardised “big data”

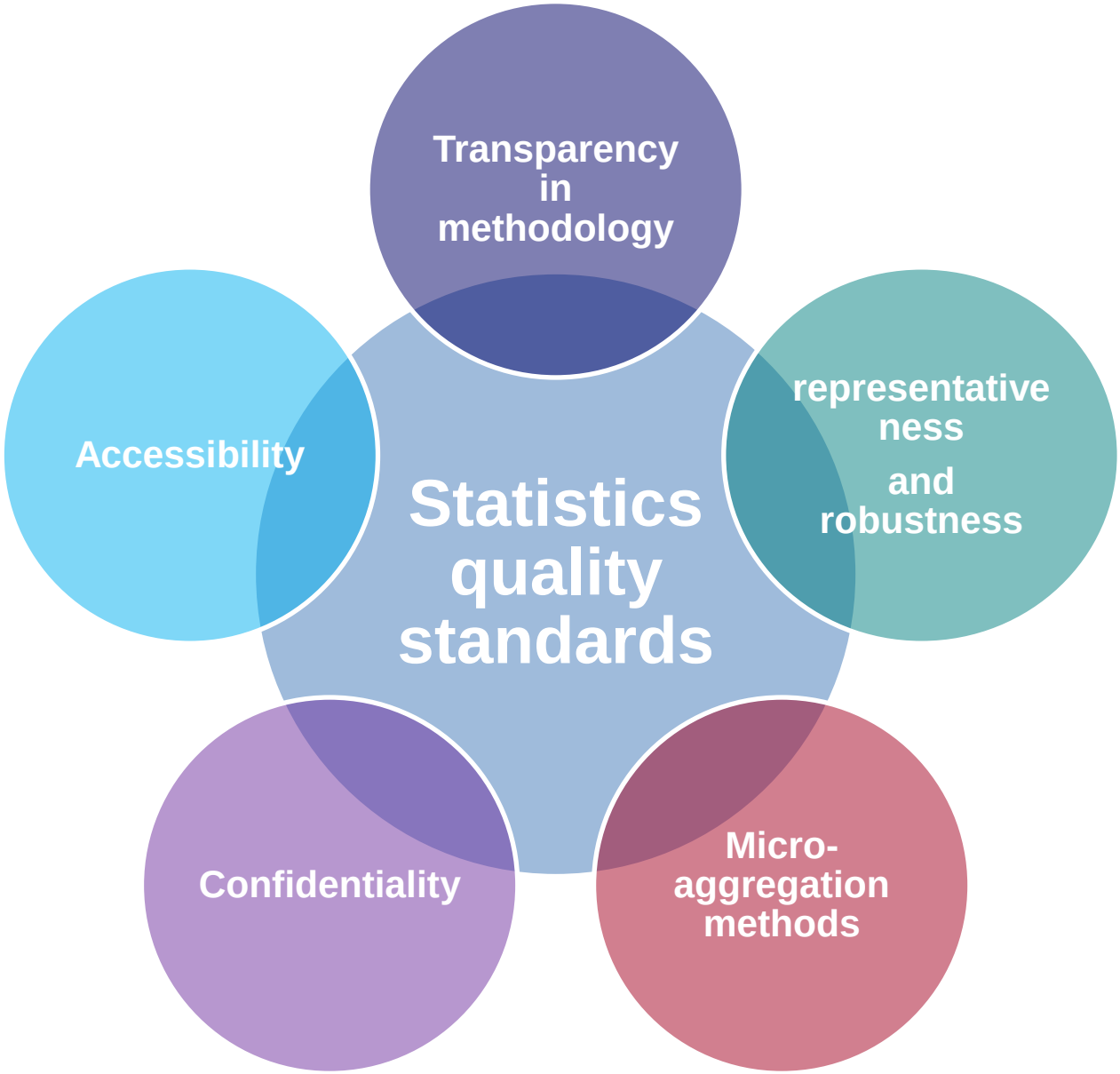


New micro-level statistics

- Security-by-security statistics
- Holdings of individual securities
- Money market transactions
- Loans by loans transactions (Ana Credit)
- Register of Financial Institutions
- Individual bank supervisory data
- Financial markets price &, volumes
- Digital records of operations
- Private data sources

New insights for crafting policies

Quality requirements for using big data as a policy source



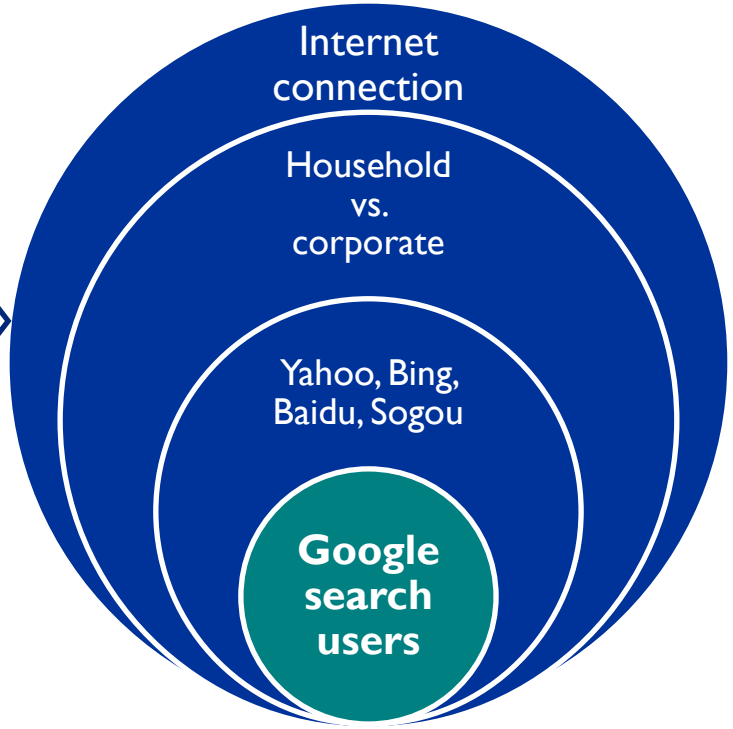
Big Data analytics – Quality assessment

One **misperception** of big data is that we **do not need** to worry about **sample bias and representativeness**, as large volumes of information supersede standard sampling theory, since big data provide census-type information

1

Digital recording of operations

Census?



2 **Unit Measurement** → **Not people - but searches**

3 **Event driven** → **Volume changes may not refer to changes in demand (e.g. emission checks)**

Google logo is downloaded from *Wikipedia*.

How do central banks use big data to craft policy?

Three takeaways

Significant digital transformations in finance and economics

experimenting for supplementary insights

Anticipate its impact and move from experimenting to central banking tool-kits

Collaborations - build effective partnerships for excellence

Any question?

EUROPEAN CENTRAL BANK
EUROSYSTEM

ECB STATISTICS PAPER SERIES

Gaining insights - Growing understanding – Spreading knowledge

FACTS COUNT

The ***ECB Statistics Paper Series (SPS)*** is a channel for statisticians, economists, researchers and other professionals to publish innovative work undertaken in the area of statistics and related methodologies of interest to central banks.

Fact-check your talk before you walk

WHAT ABOUT YOU WRITING?

Bank Indonesia / IFC-BIS international Seminar on big data,

"*Building pathways for policy making with big data*",

Session 3: Panel discussions "How do central banks use big data to craft policy?

Per Nymand-Andersen[1], Adviser, European Central Bank

26 July 2018

- [Slide 1 – cover page]

Selamat Siang
Saya senang berada di Bali
Nama saya Per Nymand-Andersen

I would like to thank

- Bank Indonesia
- Irving Fisher Committee and its secretariat; and
- Mrs Yati Kurniati, Executive Director and Mr Erwin Rijanto, Deputy Governor of Bank Indonesia,

for the kind personal invitation to contribute to this exclusive seminar.

Now let's me start with Charles Darwin who published his "theory of evolution" within his book "**On the origin of species**" in 1859. In this book, he explains the process of natural selection, whereby successful species adapt to changing environments and those who fail to change and reproduce, die off.

[1] Contact: Per.nymand@ecb.int

- [Slide 2 – One minute of internet]

Likewise, the wide spread application of innovations and inventions influence our lives; the way we operate and interact both as individuals, as market operators and regulators. The ***ability to adapt*** to changes seem to be fundamental for survival, being it species, cultures or established institutions, such as central banks operating, regulating, overseeing financial markets.

- [Slide 3 – Data mania vs phobia]

While digital transformation is a relatively new phenomena, it differs from past changes, at least in terms of the "speed to market" and the "short life cycles" of digital applications. For instance, in March 2018, there were 2.1 million apps available for downloading from the Apple store up from 800 apps 10 years ago (July 2008). Source: lifewire.com.

The speed of the digital transformation kicks-in following the mass production and widespread of technologies, such as

- **The internet** has changed the way people obtain and share information. Citizens spend more time online, perform more tasks and are more social activate – belonging to several niche communities, exchanging videos, playing games and are themselves digital creators and spreading digital footprints.

- **Smart devices –** the widespread use of tablets and smart phone devices enable citizens to be on-line, shop online and access information anywhere. These devices have facilitated the way people operate in real time staying connected 24-7.

- **Digital social communities**, affecting the way citizens socialize using digital media, how they engage and meet friends, dating and become members of multiple sub-cultures sharing common interests. Digital social media allow people to join and switch among social platforms and contribute collaboratively to social projects, such as, Wikipedia and/or benefits from "the shared economy".

- [Slide 4 – impact of digital transformation]

The impact is widespread and has transformed economic behaviors, how financial markets interact, which then impacts central bank's monetary policy, the transmission mechanism and Financial Stability. Just think of digital trading platforms, consumer online spending, crowdfunding, mobile payments and algorithm trading. The latter - the hunt of executing first - have been reduced to mille-seconds or even nano-seconds – contributing to efficiency gains, market transparencies and the pass-through of monetary policy to the real economy.

- [Slide 5 – reflections for central banking]

Given the impact of the widespread use of technology within our society, central banks naturally should continue to react prudently and in a similar forward-looking way as we have implemented policies. Central bankers do not need to be front-surfers on the digital wave – though needs to understand, analyse and assess the dynamics and impact of the digital transformation not only within the financial system but also in the real economy, including changes in consumer patterns, spending and saving behavior of households.

Big data may overcome some of the shortcomings of traditional Marco-economic data – by timely and richer details, allowing to test new theories and behavioral economics, hoarding effects and psychology.

It will eventually provide supplementary insights and early warning indicators as part of the central banking tool kits.

- [slide 6 - Moving to micro-level/granular data]

The recent financial crisis has given a boost to obtain more granular and timelier data supplementing the Macro-level statistics: Central bankers are becoming producers of big data. We have started collecting daily euro money market transactions which refer to approximately 10.000 daily unsecured money market transactions with the value of round EUR 100 billion/day and around 30.000 secured short term loans of the value of around EUR500 billion;

This will be shortly supplemented with individual bank loans to corporates (AnaCredit) with significant larger daily volumes of transactions. We are expecting approximately 70 million loan exposures per month granted to approximately 15 million counterparties.

These statistics will provide meaningful complement to existing official data – for instance to assess the impact of our Assets Purchasing Programme (APP) on market functioning and for calculating a new overnight unsecured interest rate for the euro area (ESTER). All structured and standardised data.

From private sources, we are experiencing with

1) The use of google search terms for nowcasting macro-economic indicators such as unemployment (lagging indicator) and car sales (leading indicator) providing evidence that the search terms may correlate and may reduce the forecasting errors;

2) On-line price data and barcode scanner data can improve short term forecasts for checking robustness and reliability of current price indexes and factors determining prices and price dispersion.

3) Electronic payment data from credit cards and ATMs has providing insights on forecasting private consumptions and so far as GDP growth;

4) Textual exploration from financial news reports and central banking communications are used to gauge public sentiments

- [slide 7 – Statistics quality requirements]

Central banks are ***significantly experienced*** in handling large and structured datasets – which requires using machine learning techniques for instance for quality checking daily transactions and operations.

Though the quality challenges of micro-data remains very similar as for the regular provision of macro-economic statistics

The challenges remain the same for complying with the statistics quality requirements, as part of using a source for regular policy purpose, such as

transparency in methodology, micro-aggregation methods, preserving confidentiality while facilitating access.

- [slide 8 – Assessment of quality requirements]

Let me give you one example relating to representativeness. While one may have access to the full census of all available internet searches, it may not fully represent the household sector as not all households have access to the internet and corporates a likewise using search engines so statistical adjustments are required also for big data sources. Furthermore, for unit measurement, an internet search may not relate to a person, as one person can make several searches and may likewise be driven by events rather than increases in demands. Other challenges relate to that Fintech is borderless and challenges the concept of national territories and statistics.

- [slide 9 – three take away(s)]

Allow me though to conclude.

As central banker, we need to take a holistic view as part of ensuring trust in markets, statistics and systems and assess the viability of new digital transformation in finance economics.

1 - Technological progress and Digital transformation are integral drivers of economic and financial development and have a profound impact on financial markets. They are borderless and impacts the structure and functioning of our economies and societies.

Big data is part of this technological service evolution. New sources for explorations, new methods, new software and hardware combined with open sources technology

and cloud services enable central banks to experiment using big data sources for central banking purposes, without necessary spending large IT investments and maintenance costs.
We have started this path of experimenting with google search data, on-line price and scanning data, electronic payment data and textual from financial news.

2 - Progress lies in experimenting with these now digital outlets for laying the grounds for moving **from experimental to policy tool kit** – more efforts are required for applying standards, obtaining structured data and the inter-operability of datasets. It is of paramount importance to ensure that data remains of sufficient quality and reliability to systematically inform policy makers.

We need to explore and experiment with the new data and technologies as not to fall too far behind the curve and to *anticipate* its impact on central bank's monetary policy and its transmission throughout the economy, financial stability and for banking supervision.

3 – Collaborative efforts among central banks has been initiated by the Irving Fisher Committee, bringing central banks together in showcasing few pilot projects on the use of big data for central banking purposes. Collaborative efforts are needed to build effective partnerships for central banking excellence.

We are on a journey which will continue to drastically rework the financial ecosystem. Central banks need to monitor closely, assess - and now I will come back to the thinking of Charles Darwin – adapt - to the new market place.

Thank you very much for your attention.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

The Bank of France datalake[1]

Renaud Lacroix,

Bank of France

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANQUE DE FRANCE
EUROSYSTÈME

Directorate general Statistics

The Banque de France Datalake

Renaud Lacroix

Director, Statistical and IT engineering division

« Building Pathways for Policy Making with BigData »

BI – IFC/BIS Seminar
Bali, 26 July 2018

BANQUE DE FRANCE
EUROSYSTÈME

A FEW DRIVERS

1. New age of statistics

- ✓ Growing appetite for statistics…while
- ✓ General public more skeptic with regard to numbers (including official ones!)
- ✓ Legitimate requests for granular data
- ✓ New and very powerful competitors (GAFA)

2. Strategic challenge for Central banks

- ✓ Central banks must be able to **deliverer rapidly reliable intelligence at both micro and macro levels.**

3. Large consequences for data management

- ✓ Functional silos are not adapted anymore
- ✓ A clear guidance and an orderly process is a key to manage wide volumes of diverse data
- ✓ An innovative and scalable technology is crucial

BANQUE DE FRANCE
EUROSYSTÈME

A POSSIBLE WAY FORWARD: BUILDING A DATALAKE

1. Building a coherent and unique set-up: for data

- ✓collection,
- ✓quality management,
- ✓pooling,
- ✓analysis,
- ✓dissemination

2. Integrating the Big Data techniques

3. Delivering both economies and better work : it is possible to do more with less spending

BANQUE DE FRANCE
EUROSYSTÈME

OBJECTIVES OF THE DATALAKE PROJECT

- **The Datalake project consists in the building of a multidisciplinary granular data platform supplying flexible and innovative services to internal users.**

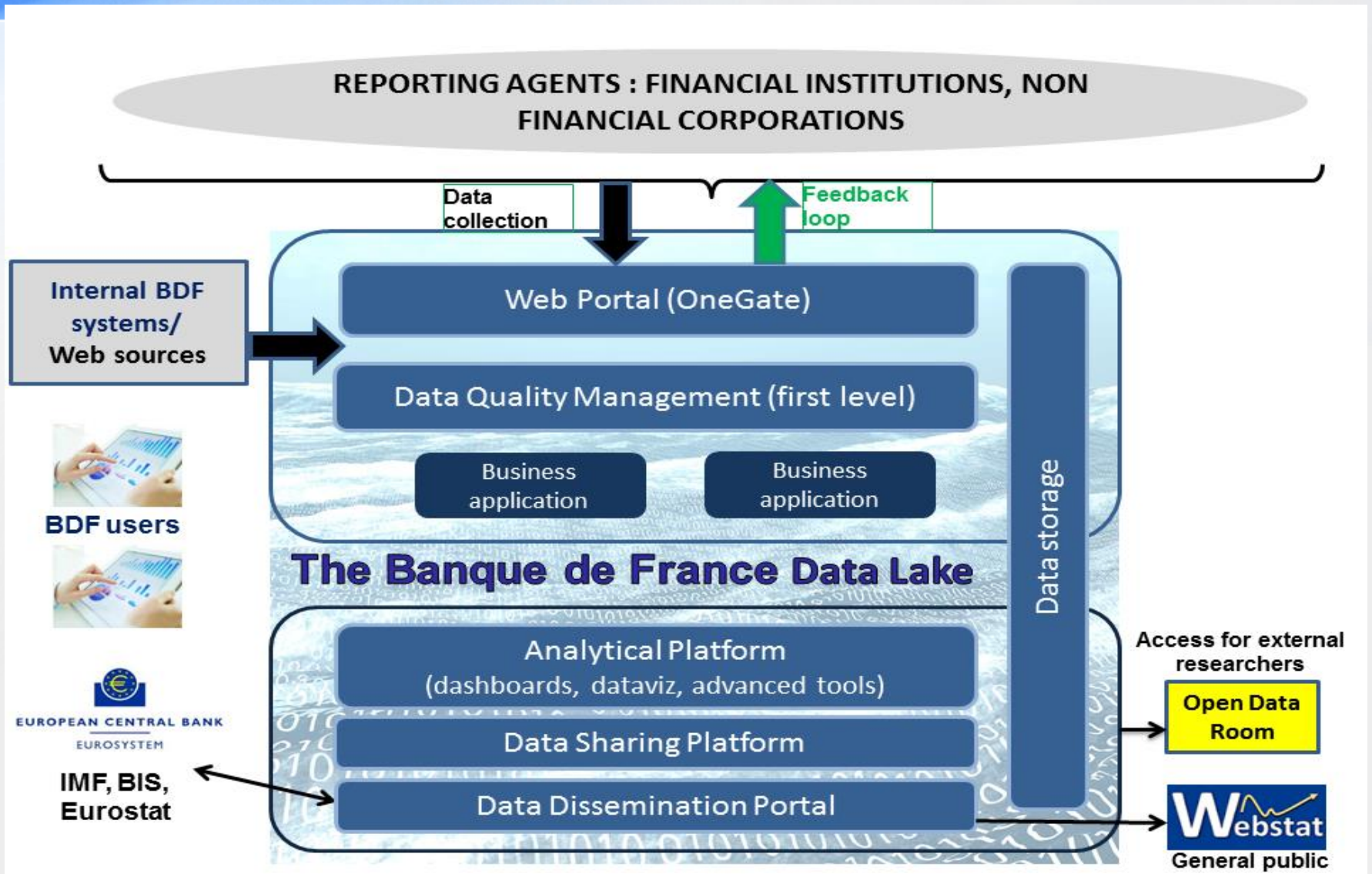
- The platform provides key services pivotal to operational activities of the Banque de France: data collection, data supply, data quality management, data storage, data sharing, analytics, and dissemination (external data exchange platform).

- 2 new components :
  - data storage and **automated data quality management** for first-level quality checks
  - analytical platform (dashboarding, statistical treatments covering a wide range of user needs)

- **The implementation phase of the project will be completed by end2018**

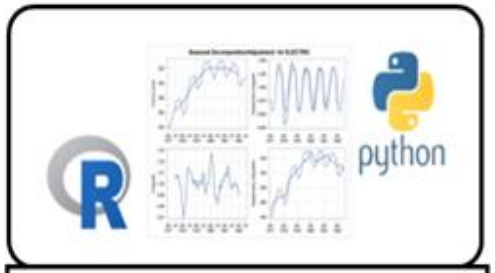
BANQUE DE FRANCE
EUROSYSTÈME

AN OVERVIEW OF THE DATALAKE ECOSYSTEM



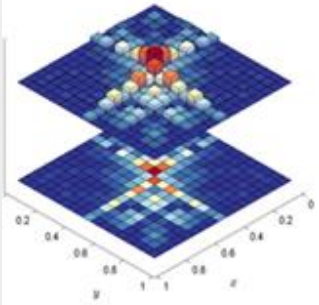
Implementation of the Datalake core structure: an analytical platform for all users

Interactive query tools



Libraries of statistical functions, programming tools

Data science: machine learning, neural networks, deep learning



From prototyping ....

.... to large-scale production

- Specifications - Samples - Large-scale production - Final product

Final product designed by end-users

« Industrialized » final product with the support of the project team

BANQUE DE FRANCE
EUROSYSTÈME

Early BDF customers of Datalake services

❑ Banking Services Directorate

- Gathering payments for private and corporate customers monitored by the BDF in a unique decision-making system.
- Provide the analytical tools to contribute to the fight against money laundering and terrorism financing.

❑ DATAGAPS project

- G20 recommendations on gathering information reported by systemic banks on financial risks by type of exposures

❑ Anacredit project (implementation of the ECB regulation)

- Collection of detailed information on both credit lines and credit risks from credit institutions

❑ Reengineering of the French NSA's information system

BANQUE DE FRANCE
EUROSYSTÈME

A FEW FIRST CASE STUDIES IN MACHINE LEARNING

➢ Nowcasting / forecasting French GDP

- **Use of a rich database (more than 200 explanatory factors)** in the framework of adaptative LASSO with automatic selection of relevant variables for forecasting purpose
- **Valuable complement to more traditional approaches for forecasting**

➢ Improving tourism statistics

- Web scraping of accommodation platform (Airbnb, Booking.com,..) and machine learning to anticipate future demand relying on meteorological data and future events
- Comprehensive Use of credit card data for the estimation of both the spending of French residents abroad and the spending of foreign residents in France
- First attempts to use mobile phone data

KEY CHALLENGES OF THE PROJECT 1/2

- **From an organizational perspective** :

  - **Top-down strategy, but bottom-up implementation :**
  - **Strong involvement** of the management at all levels

  - **Adapt to users requests** : business lines contribute to the definition of Datalake services while the Datalake team defines standards for mutualized services

  - Move **step by step** in order to deliver the best services to the customers

  - **Reshape the organization of work to benefit from automation and new techniques**

  - **Need for data-skills** in big data technologies (computer science, data science, IT experts) : develop internal training

  - **Leave no-one by the wayside** : everyone should benefit from the Datalake services and understand what we are trying to achieve and why

BANQUE DE FRANCE
EUROSYSTÈME

KEY CHALLENGES OF THE PROJECT 2/2

- **On the IT side** :
  - large changes on IT infrastructures
  - a large number of components (softwares) are required, including some brand new
  - Do not seek innovation at any price !
- **On the statistical side** :
  - Learn to work with new data sets from Google, social media, wesites
  - Learn to work with new data analysis algorithms and enrich the traditional toolbox
  - Develop (interactive) visualizations tools and techniques for data dimensionality reduction
  -

BANQUE DE FRANCE
EUROSYSTÈME

BIG CHALLENGES AHEAD FOR THE PUBLIC SECTOR

- **Public authorities are now in direct competition with the private sector in the sphere of economic information**
  - the appetite for real time intelligence can be inflated by the Big Data Revolution
  - GAFA and other global players are not at all regulated
  - technological progress in Big Data is permanent, rapid and difficult to anticipate

- **Risks therefore that "bad data chases good ones"**

- **Central banks must be more innovative and user friendly**
  - Make the publications more readable and visible
  - Supplement Bigdata strategy by open data initiatives

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

The framework of big data: a microdata strategy[1]

Robert Kirchner,

Deutsche Bundesbank

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

DEUTSCHE BUNDESBANK
EUROSYSTEM

The framework of Big Data: A microdata strategy

BIS/IFC International Seminar on Big Data: Building Pathways for Policy Making with Big Data

Robert Kirchner, Deputy Director General Statistics, Deutsche Bundesbank

Outline

1) On the way from yesterday to tomorrow: the statistical value added chain

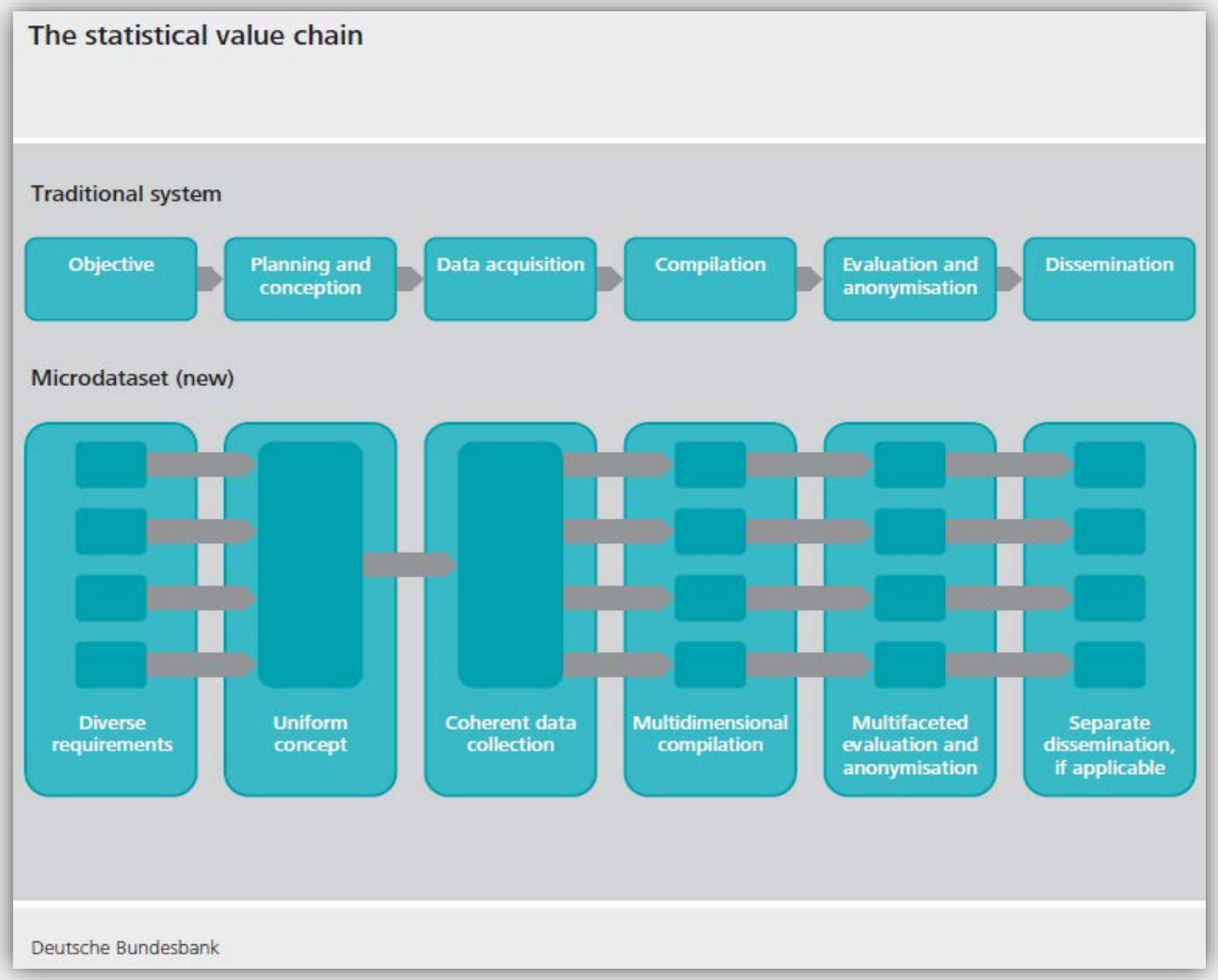
2) The Bundesbank‘s Integrated Microdata based Information and Analysis System (IMIDIAS)

3) Enhancing knowledge sharing: INEXDA

4) Big Data Projects

5) Conclusion

On the way from yesterday to tomorrow: the statistical value added chain



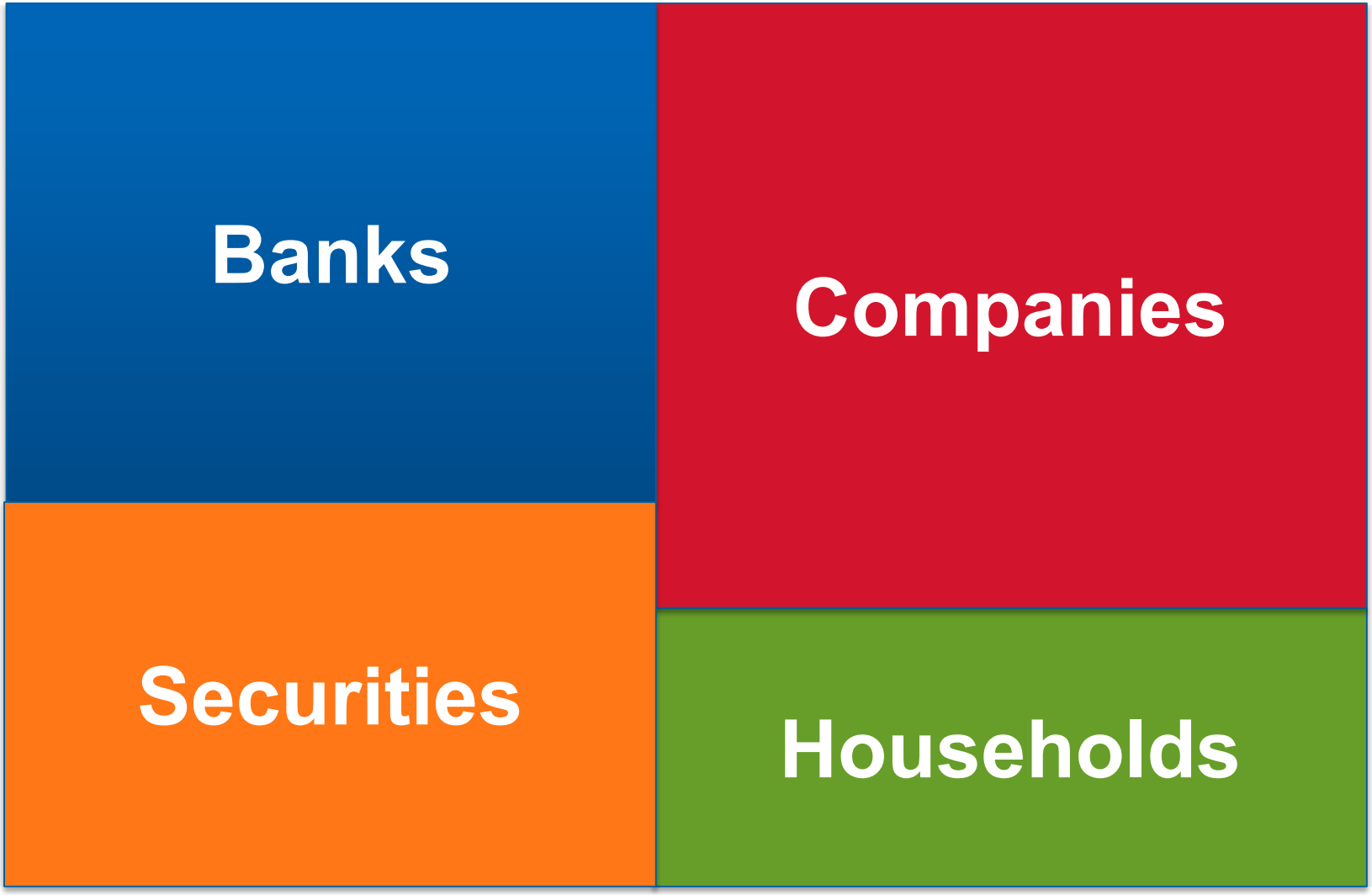
On the way from yesterday to tomorrow: the statistical value added chain

- The **four fundamental advantages of microdata**, which supplement the established macroeconomic analyses, can be summarized as:

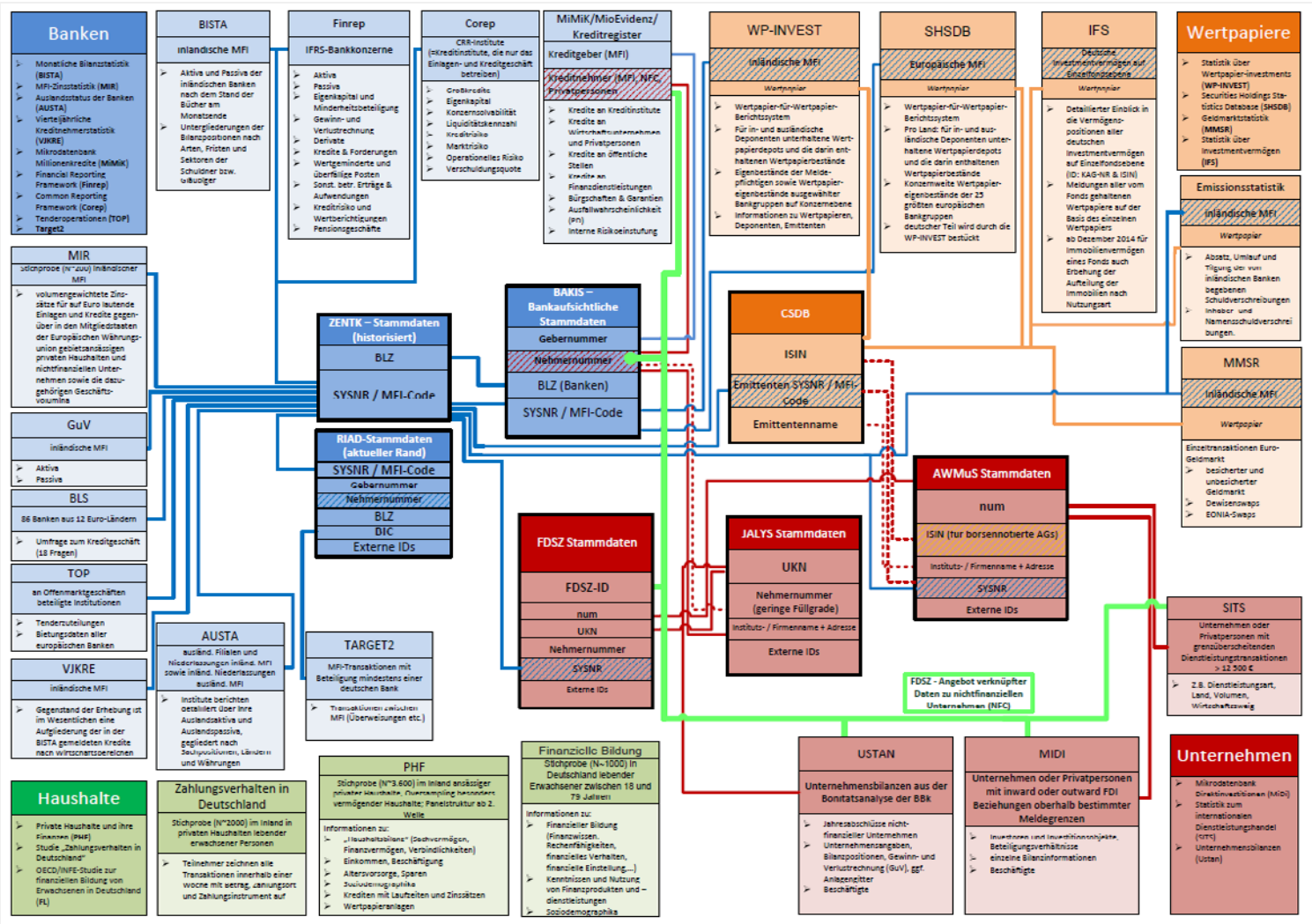
  - **Distribution**
  - **Interconnectedness**
  - **Flexibility**
  - **Policy evaluation**

- **First lesson: Make better use of existing data**

On the way from yesterday to tomorrow: the statistical value added chain



On the way from yesterday to tomorrow: the statistical value added chain



The Bundesbank‘s Integrated Microdata based Information and Analysis System (IMIDIAS)

- **Goals:**
  - Enhance internal accessibility of micro data
  - Support evidence-based policy-making
  - Encourage cooperation with (external) researchers

- **Structure:**
  - Steering Committee
  - House of Microdata: Warehouse for internal policy work
  - Research Data and Service Centre (RDSC)

- **The RDSC supports internal and external research.**
  - 12 working places for guest researchers
  - > 300 active projects, >160 institutions involved

Enhancing knowledge sharing: INEXDA

INEXDA
The Granular Data Network

- On 6th January 2017,

BANCA D'ITALIA · BANQUE DE FRANCE EUROSYSTÈME · BANK OF ENGLAND · BANCO DE PORTUGAL EUROSISTEMA · DEUTSCHE BUNDESBANK EUROSYSTEM

- have launched the **I**nternational **N**etwork of **Ex**changing Experiences on Statistical Handling of Granular **Da**ta (INEXDA), an international cooperative project to declare their willingness to further strengthen their cooperation.

- Since its foundation, the following institutions have joined INEXDA as a member:

BANCO DE ESPAÑA Eurosistema · EUROPEAN CENTRAL BANK EUROSYSTEM

Enhancing knowledge sharing: INEXDA

- **The INEXDA framework comprises several workstreams:**
  - Metadata and database
  - Administrative Data Research Facility (ADRF)
  - Modes of Accreditation
  - Contracts for research projects/bodies
  - Models of data provision
  - Dissemination
  - Procedures on output control
  - Procedures of risk management for results published out of data access

Big Data Projects

1) Internal stocktaking

2) Selected projects:

- Götz, Thomas B., Knetsch, Thomas A. (2017). **Google data in bridge equation models for German GDP**. Bundesbank Discussion Paper No 18/2017.
- Fecht, F., Thum, S. and Weber, P. (2018). **Fear, Deposit Insurance Schemes, and Deposit Reallocation in the German Banking System**. Available at SSRN: https://ssrn.com/abstract=3180107.
- Oehler, S. (Forthcoming). **Developments in the residential mortgage market in Germany** – What can Google data tell us?

3) General Issues:

- Inference problems occuring from large but biased samples
- See: **Meng, X.L. (2014)**, "A Trio of inference problems that could win you a nobel price in statistics […]"

Conclusion

- **Fundamental paradigm shift from the traditional system to a microdata approach**
- **Big Data projects are part of the new system**
- **Traditional samples remain important**

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Exploring big data to sharpen financial sector risk assessment[1]

David Roi Hardoon,

Monetary Authority of Singapore

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

EXPLORING BIG DATA TO SHARPEN FINANCIAL SECTOR RISK ASSESSMENT

"Building Pathways for Policy Making with Big Data"
BI-IFC / BIS International Seminar on Big Data, 26 July 2018

David R. Hardoon
Monetary Authority of Singapore

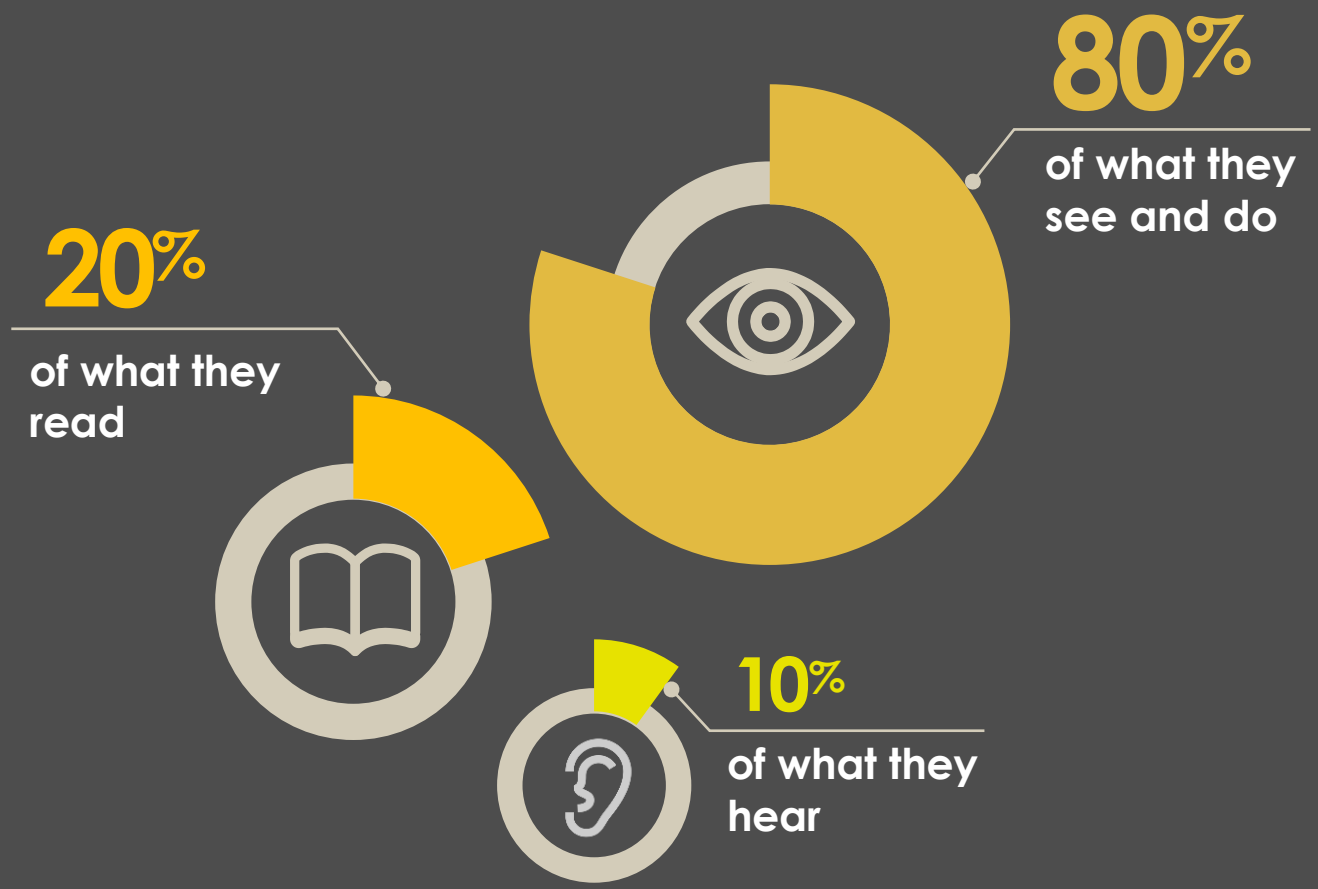
1 THE VISION

DOING BUSINESS AS UNUSUAL

https://www.denverpost.com

2 | VISUALISATION

People remember….



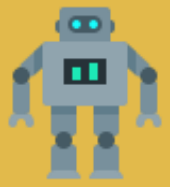
3 TRAINING PLANS

Interactive Data Analytics Course Catalogue

- In-house
- External - Online
- External - Classroom

Data Science	Collection, Preparation and Exploration	Visualisation	Machine Learning	Others
Advanced				
Data Science Specialisation			Deep Learning Specialisation	Building A Data Science Platform with Kubernetes, Anaconda & Jupyterhub
Reproducible Data Science			Advanced Data Science: Machine Learning	
Intermediate				
Intermediate Data Science (Python)	Intermediate Data Management & Analysis	Data Analyst - Visualisation II	Introduction to Machine Learning	
Introduction to Python for Data Science		Introduction to Coded Visualisation		
Introduction to Python Programming	Data Preparation & Exploration Using Python			Stata for Public Policy Workshop
Getting Started with Data Science (Python)				Statistics for Bureaucrats - Stats 101 for Policymakers
Foundational Python for Data Science		Tableau Test Drive (for Power Users)		
Master Class on Advanced Analytics	Data Analytics and Techniques with Excel and VBA	Tableau Townhall (for Normal Users)		Workshop on Randomised Controlled Trials in Public Policy
Master Class in Decision Making with Data Analytics	Intermediate Data Collection & Analysis	Data Analyst - Visualisation I		
Basic				
		The Art and Science of Data Visualisation		
Data Analytics: Core Concepts & Techniques	Customer Research through Surveys and Analytics	Data Analytics & Visualisation for Managers	Introduction to Predictive Analytics - Tools & Techniques	Stats Alive - A Foundation for Analytics
Data Analytics for Managers	Introduction to Data Management & Analysis	Business Intelligence & Dashboarding		Master Class on Building An Analytics Culture
Data Science for Managers: Crash Course in Data Science	Survey Questionaire - Design & Implementation	Introduction to Data Visualisation and Communication		Basic Analytics for HR Professionals
Starting a Data Science Project	Introduction to Data Collection & Analysis	Data Visualisation for Effective Communication	Data Science & Artificial Intelligence for Senior Executives	Driving Communications with a Data Analytics Strategy

4 CURRENT TECHNOLOGY & PROJECTS



MACHINE LEARNING

TEXT ANALYTICS & NATURAL LANGUAGE PROCESSING

VISUALISATION

AUTOMATION

5 TEXT ANALYTICS & NATURAL LANGUAGE PROCESSING

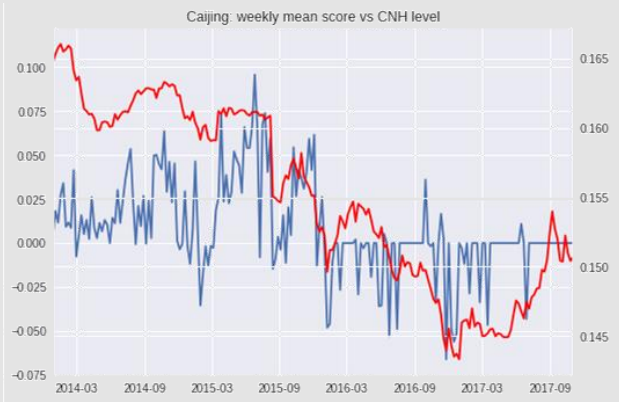
Examples of Text Analytics and NLP

SENTIMENT ANALYSIS

1. Scrape Chinese news websites
2. Clean data

3. Calculate sentiment scores

4. Visualise results and compute correlation



Correlation coeff: 0.56

TOPIC MODELLING

6 TEXT ANALYTICS & NATURAL LANGUAGE PROCESSING

Examples of Text Analytics and NLP

SENTIMENT ANALYSIS

TOPIC MODELLING

DATA

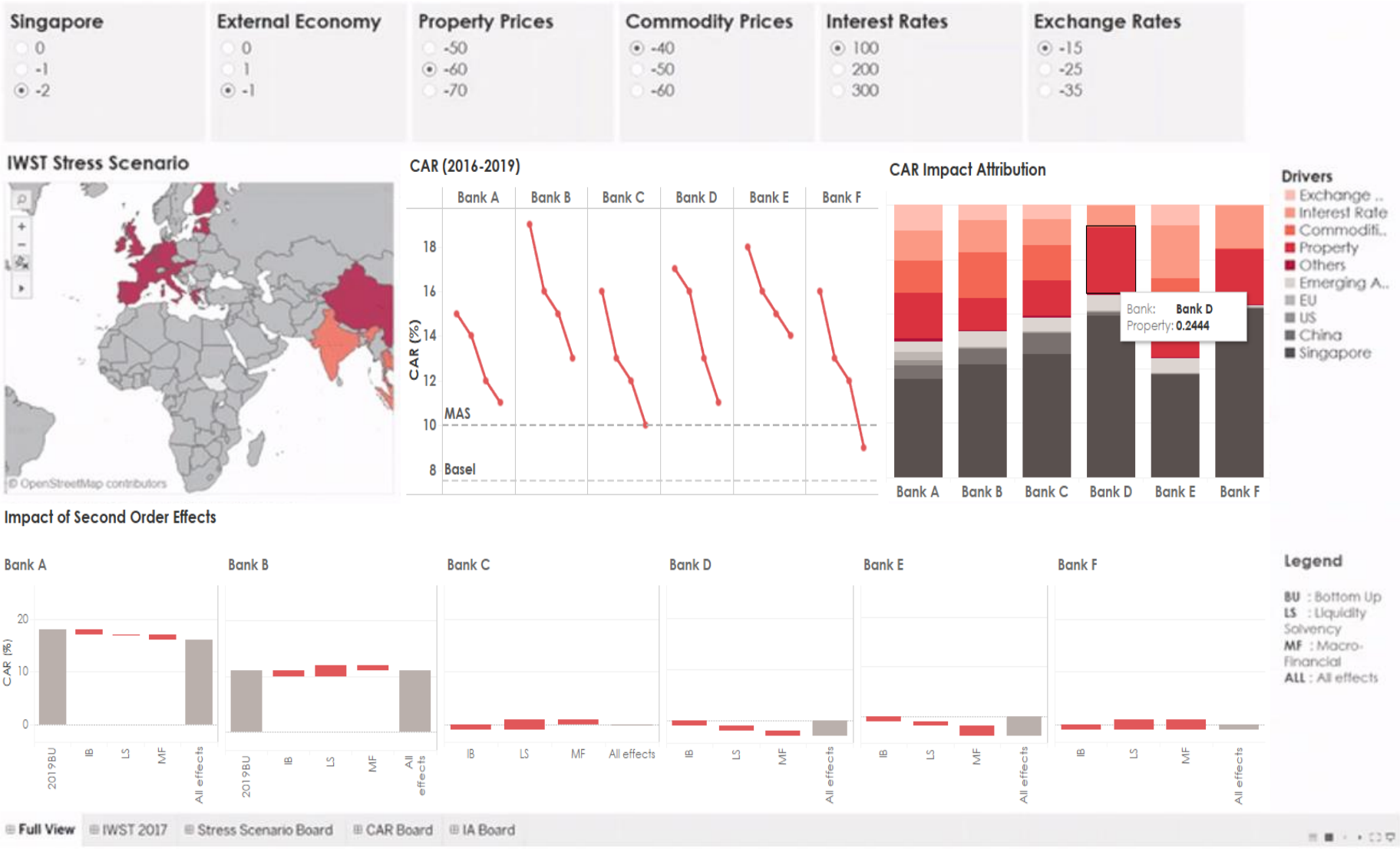
TOPIC MODELLING
(e.g. NMF, LDA)

TOPICS

Topic 1	work processes, streamlining, prioritisation
Topic 2	workload, headcount, resources
Topic 3	productivity, efficiency, systems and technology

7 VISUALISATION

Industry Wide Stress Test (IWST) Dashboard



8 AUTOMATION

Filtering News Articles

Automating news surveillance

PROBLEM

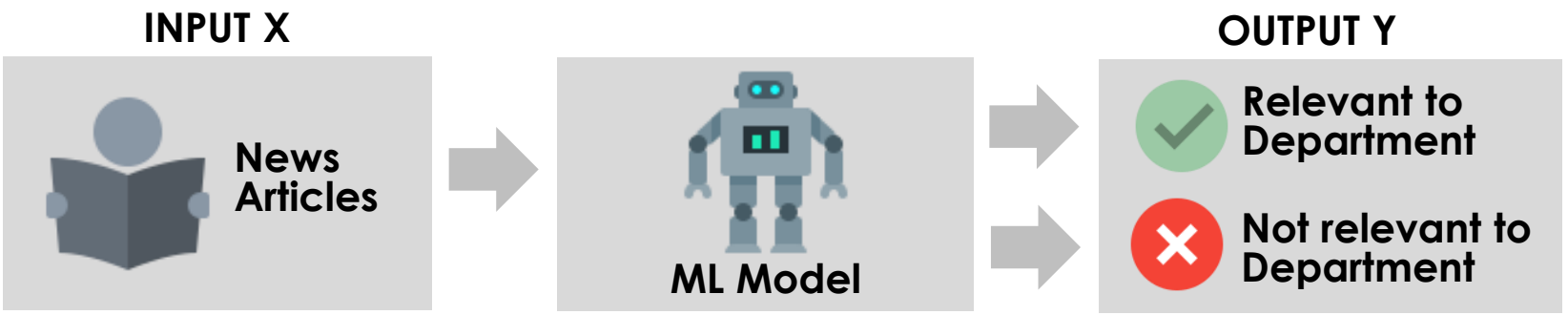
Many departments in MAS currently filter news alerts by keywords, and rely on daily checks by support officers to assess if there are specific events that require attention. This process is time-consuming.

AUTOMATION

To help automate the process, we can train a model based on labelled articles.

PRODUCTIVITY GAINS

The model predicts whether a new article is relevant to a department. A similar model is currently used by Comms in filtering articles relevant to MAS, saving them a lot of time from the manual filtering.



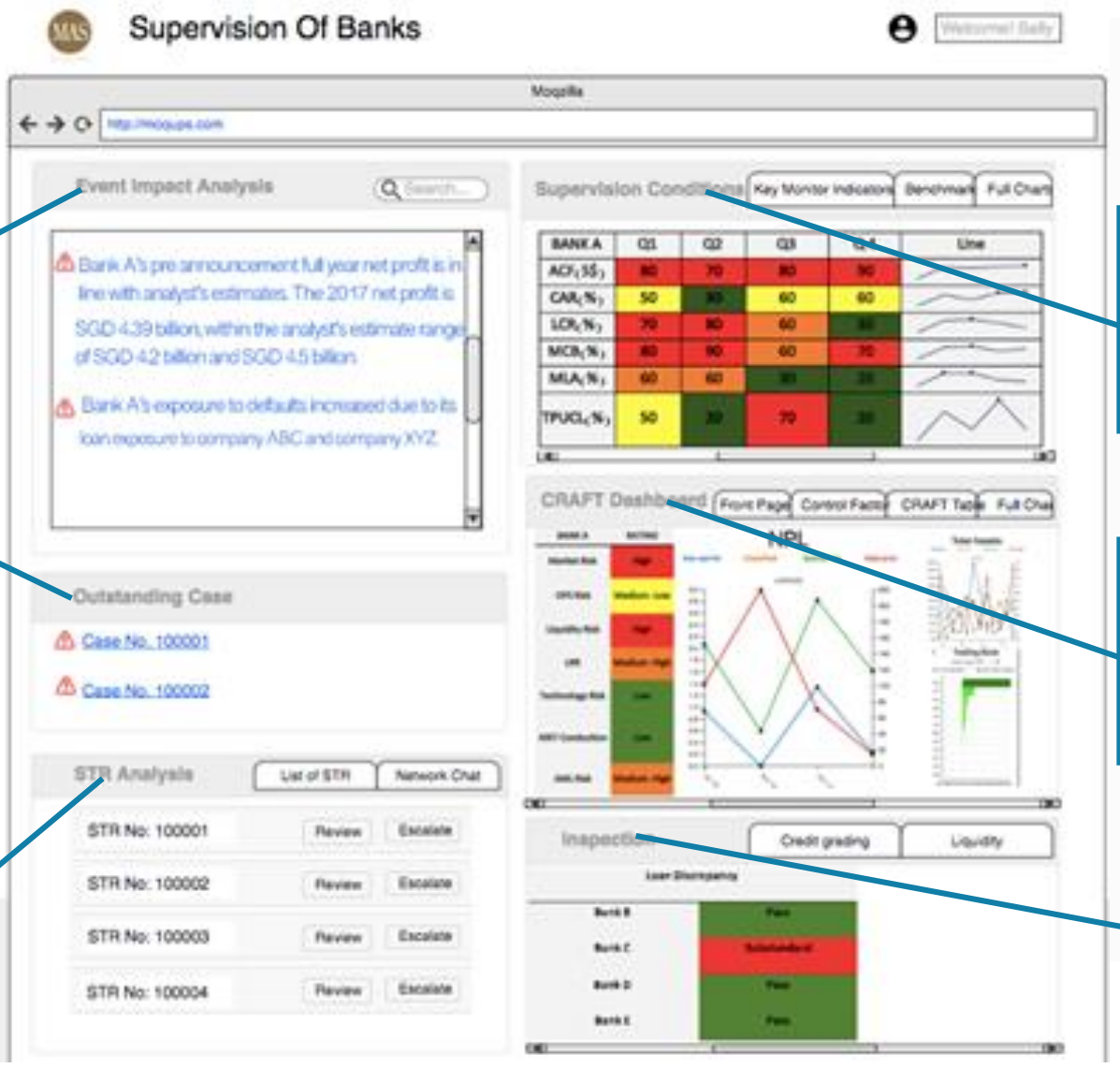
9 FUTURE OF MAS DIGITAL SUPERVISION

Example of 'Bank A' Portal View

Latest news directly relevant to Bank A

Ongoing communication btw MAS & Bank A

Suspicious transactions reported by Bank A



MAS' assessment of Bank A's risk over time

Supervisor's rating of Bank A on different indicators

Inspection assessment of Bank A

10 ALERT PRIORITIZATION

The Pain & The Cure

130 Alerts per week

11.6% are important

4 Analysts

2 days per week

Pain

Cure

Machine to learn and classify alerts by importance

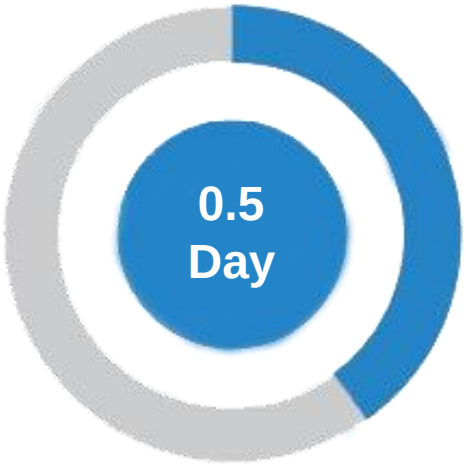
75%

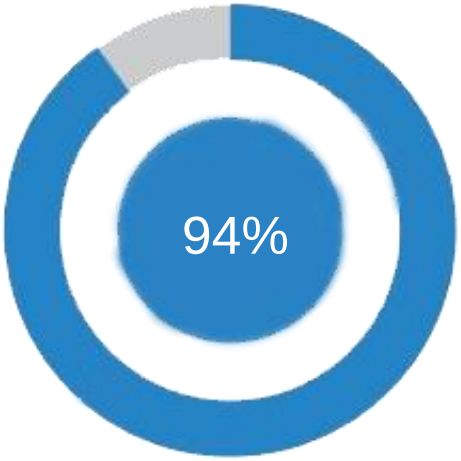
To capture as many important alerts as possible

25%

To reduce workload

11 | ALERT PRIORITIZATION - THE RESULTS





Workload Efficiency Gain

2 days per week reduced to 1.5 days. Monthly gain of 4 days

Alert Capture Recall

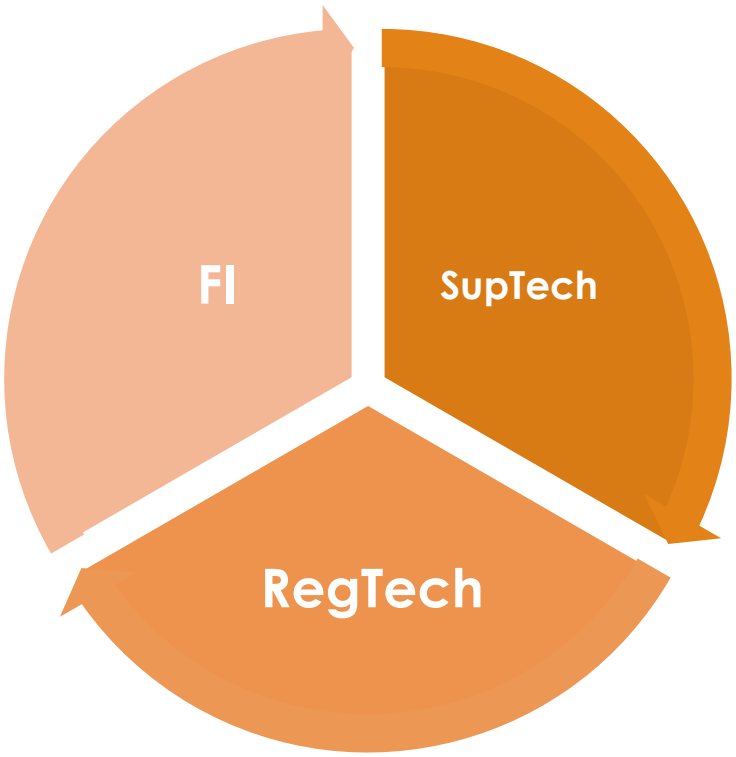
Objective is to capture as many positive alerts as possible

Note: The results are based on the optimal model and 10-fold cross validation

12 FI-REGTECH DIALOGUE

Enhance regulatory compliance with RegTech and latest technology

1. RegTechs to better understand the problem statements from FIs
2. FIs to better understand solutions from RegTechs
3. Encourage RegTech/FI interaction and potential POCs



13 ENABLING CROSS BORDER INSIGHTS

THANK YOU!

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

Data science at the Netherlands Bank[1]

Iman van Lelyveld,

Netherlands Bank

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

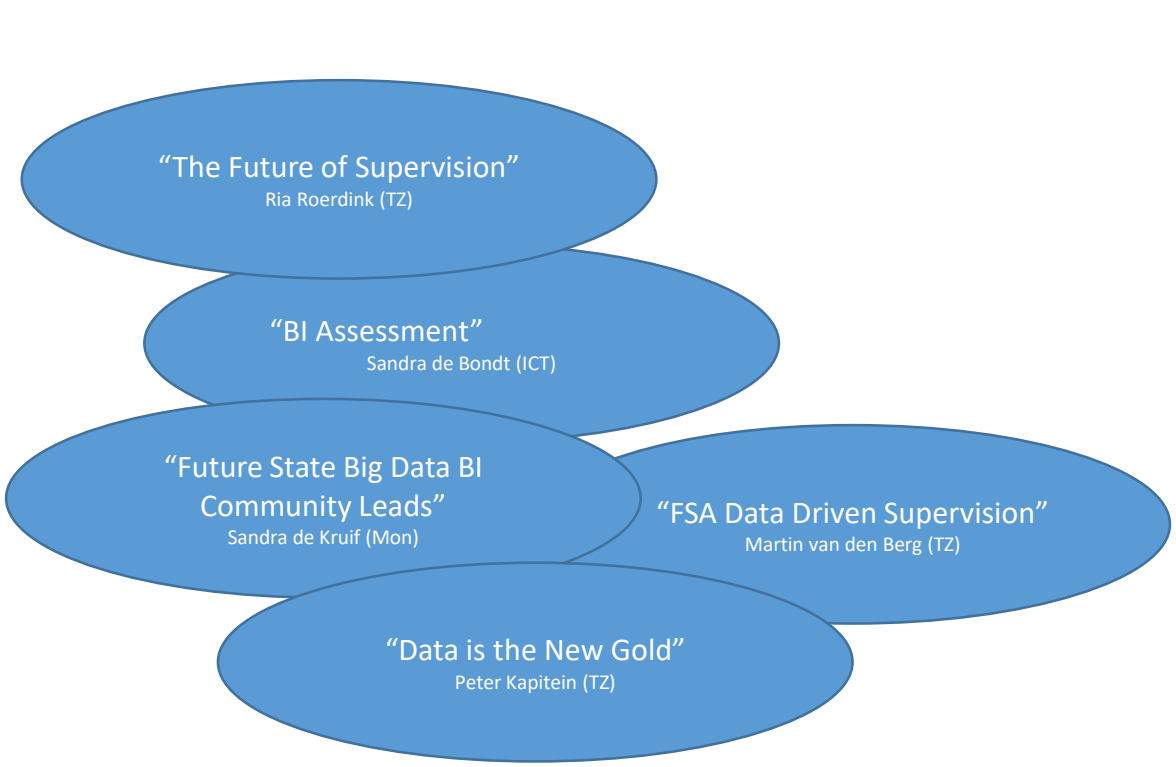
DATA SCIENCE AT THE NETHERLANDS BANK

Building Pathways for Policy Making with Big Data
BI/IFC-BIS Seminar

IMAN VAN LELYVELD

JULY 2018

Many Data Science Related Initiatives…



Bank wide programme:
In Charge of Data!

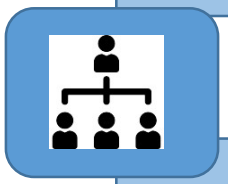
Which elements do we need?

Tooling

- Research Area Network (RAN), Data Platform + Analytical Workspaces/Datalabs/Data science Toolkit, memory/cpu/storage
- Cloud deployment; Data(platform) connectivity, other connectivity (open data, etc..), quick scaling of datalabs
- Open source tooling (e.g. R,python, Git, Neo4J, MongoDB, SQLlight, MySQL, …..)

People

- Appreciation of the scientific method
- Knowledge of statistics (descriptive, explorative, predictive, causal, ...)
- Knowledge of coding in 'interpreter' languages (Python, R, Julia, ...) and support (Anaconda, Jupyter Notebooks, Git, ...)



Organisation

- Decentral vs. Central
- Governance (!!!) – data protection, deployment of analysis (KIV→KII)
- Agile, pilots, data science as a brand
- FTE's

Culture

- Informal: knowledge networks, lunches, seminars
- Creating a community, many already do 'something' with datascience: Get-togethers, what do people need?, datascience 101 sessions, seminar with externals, deep-dive sessies (R, Python, Git, LAMP stack, Neo4J, MySQL, MongoDB, etc..), show preliminary - results

Current tooling is inadequate

- Data not sufficiently standardized
- Data not sufficiently accessible
  - Application Programming Interfaces (API-layer)
- Research Area Network (RAN) too successful
  - Tedious to move information back and forth
  - Is limited in size and computational speed
  - Fails at a basic level (even simple code loops crash for no reason)

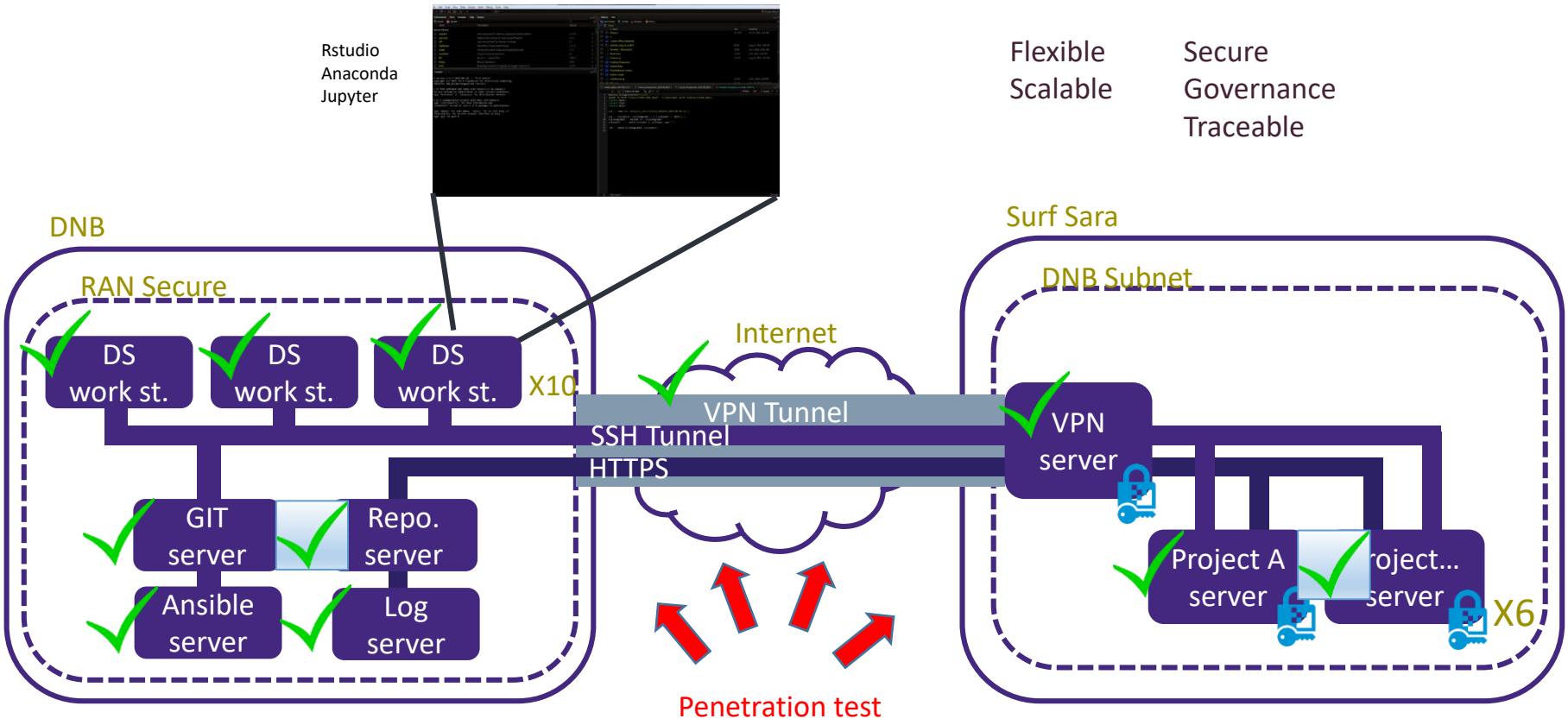
We need a propper analytical workbench … but how ??

Let's just find out!!!

Today’s agenda

1. Bring confidential data to the cloud
2. Deliver 4 Proofs of Concepts
3. Build a Data Science Community

Bringing confidential data to the cloud



Proofs of concepts I

PoC 1 Credit risk — Pieter van den Berg

Importance

- Credit risk is the largest contributor to overall financial risk and required capital for all of the large banks. Large banks in the NL have permission to calculate minimum capital requirements using internal models.
- Differences between risk estimates of different banks are due to different actual risks and risk management practices (warranted variability) but also due to different default- and loss definitions, modelling methodology, assumptions and other factors such as regulatory add-ons.
- Due to the variety of modelling practices allowed under Basel and the CRR, there is no clear view on what exactly would constitute an acceptable level of RWA variability.
- Explicitly estimating bank-specific effects could lead to a better understanding of the differences between minimum capital requirements of banks.

Analysis

- Get historical loan tape data (100Gbs) in the cloud; at least initially only mortgage data
- Extend existing code base of OSBE/IMK to build simple credit risk / pillar 1 shadow models
- Estimate bank-specific effects
- Compare typical (scorecard) methods with more sophisticated ('machine learning') methods
- Nice to have:
  - Bayesian methods, with Stan or on a tensorflow backend
  - More advanced modelling techniques (state-space/graphical models)

Deliverables

- Replication / extension of current RAN secure RStudio+MonetDB analysis environment in the cloud
- IRB challenger modelling methodology and results
- A better understanding of RWA variability for Dutch banks
- A more generally applicable methodology for risk-corrected RWA benchmarking

PoC 2 EURO CCP — Eric Hogewoning

Importance

- Central Counterparties (CCPs) are becoming more and more important. Post-crisis regulation and technical innovations draw increasing market volume to these central nodes.
- With their increased systemic importance, it is imperative to understand the risks in CCPS. In particular, how CCP transaction data can support the CCP-oversight function.
- In order to have a firm base for data driven oversight/supervision we need to develop risk indicators from transaction and initial margin data.

Analysis

- Take the transaction and initial margin data of EuroCCP for one year.
- Explore the data and develop risk indicators with R.
- Develop a method to set a threshold for medium or substantial change in an indicator with R.
- Write reusable code for the different CCP risk indicators and threshold method.

Deliverables

- A better understanding of the potential of the data for deriving risk indicators
- A selection of risk indicators linked to the Principles for Financial Market Infrastructures
- Know whether it is possible to derive a adequate threshold for medium or substantial risk of the indicators

Proofs of concepts II

PoC 3 Contagion — Dieter Wang

Importance

- Credit Default Swaps (CDS) prices reflect the perceived credit risk of the underlying entity, e.g. a bank. A higher CDS price indicates higher credit risk.
- To understand what drives these prices and risks, researchers usually analyze the role of bank-specific (stock prices, leverage ratio) or economy-wide (stock index, bond yields) variables. However, these variables explain CDS prices very poorly.
- Not only that, the part of credit risk that we cannot explain is not random. Instead, we know that there is another *hidden* variable which failed to include.
- In other words: **Something is driving credit risk, that we cannot explain!**

Analysis

- It is not too surprising, that the bank-specific and economy-wide variables do not explain credit risk of banks very well – we didn't account for *contagion*!
- Thus, we look at the networks between banks based on asset holdings similarity. Why? In case a bank under stress starts a fire-sale of its assets, this network will tell us who´s likely to be affected by the sale (namely those with similar holdings).
- We use the resulting portfolio overlap network to capture how much of this stress or credit risk spills over from one bank to another.
- Lastly, we estimate the importance of the network over time, because contagion is likely to be more important during stress times than calm times.

Deliverables

- We hope to find proof that the portfolio overlap network can capture the hidden part of credit risk that eluded the other variables.
- We would like to see how the importance of the network varies over time.
- Ideally, we can use our resulting model to conduct stress tests in the system. I.e. a dashboard that tells us, how a shock will pass through the banking system.

PoC 4 IRS — Iman van Lelyveld

Importance

- Interest rate derivative markets are a key component in how interest rates are managed by financial an non-financial firms. Daily volume is USD 2.7 trillion.
- Recent regulatory chances force more and more firms to post margin. This reduces counterparty credit risk. But transforms it into liquidity risk.
- Especially for entities that are not used to posting margin (and without access to the discount window == pension funds) this might be an issue
- Moreover, from an FS perspective, the sinkhole effect of unexpected IM calls might be an issue

Analysis

- Collect the data: take IRS data from 3 TRs (DTCC, REGIS, ICE), apply cleaning steps, add auxiliary data, join. We do this for a limited number of days.
- Analyse the data in a mix of Stata, R, Python, Gephi
- Write modular code for the deliverables: community detection, pricing, and stress testing modules.

Deliverables

1. Overview of the Dutch IRS markets How many observations/reporters/contracts do we have? What kind of contracts do these parties trade? Who is buying/selling risks?

2. How does TR reporting match up with prudential reporting Match TR with other prudential data: FINREP/COREP, Basel International Data Hub data

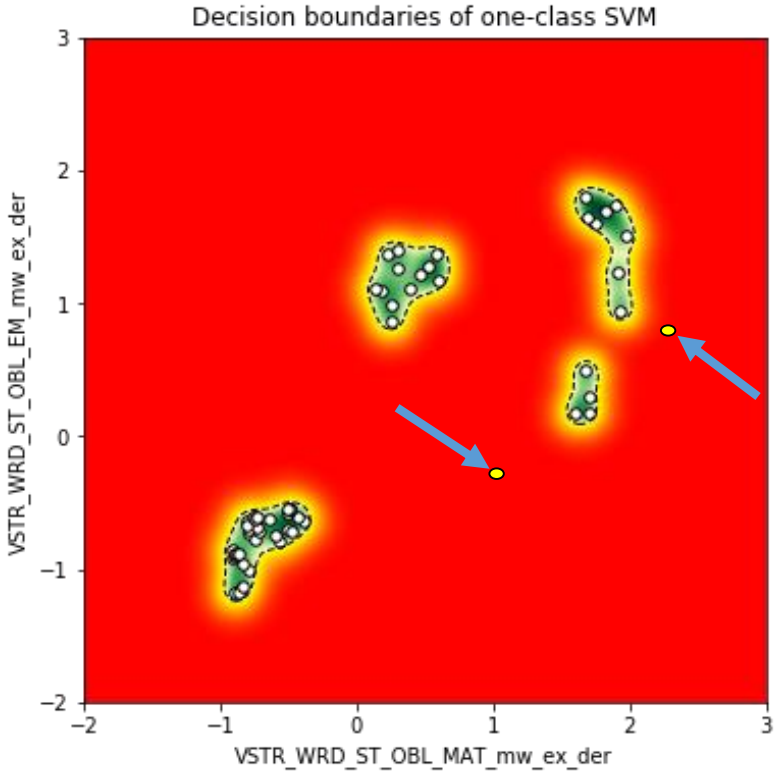
3. How to price IRS with TR data? Given that we don't trust the MtM in the data, we want to price ourselves. What are the options? How can we estimate all of this?

4. How to come to sensible estimate of margin demands? Given the price model, what model can we use to have a simple robust model that can reliably estimate margin demands across the entire market

5. A stress test of the Dutch IRS market Given data, prices and margin demands: how would changing some of the parameters affect a) solvency and b) margin required?

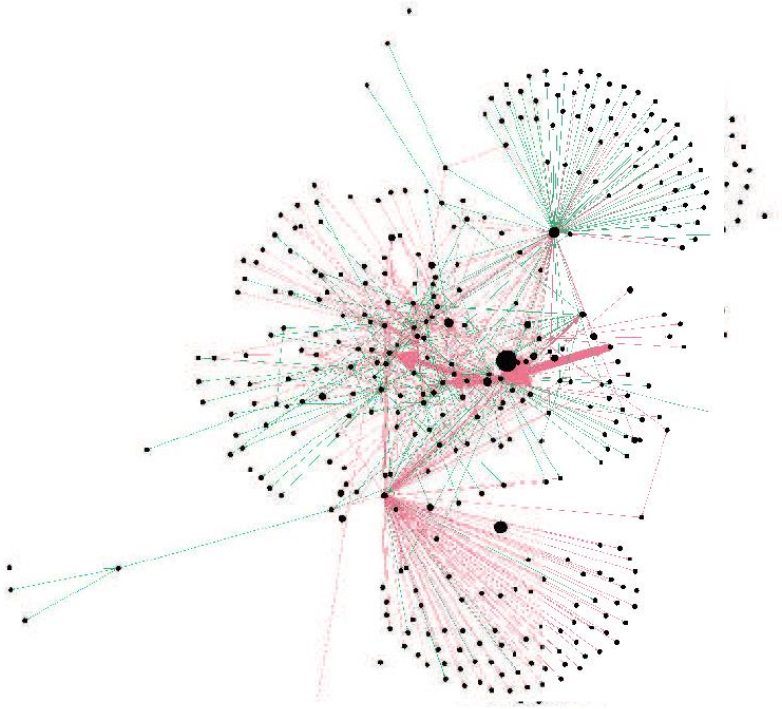
Dazzling Data Science …

Machine Learning



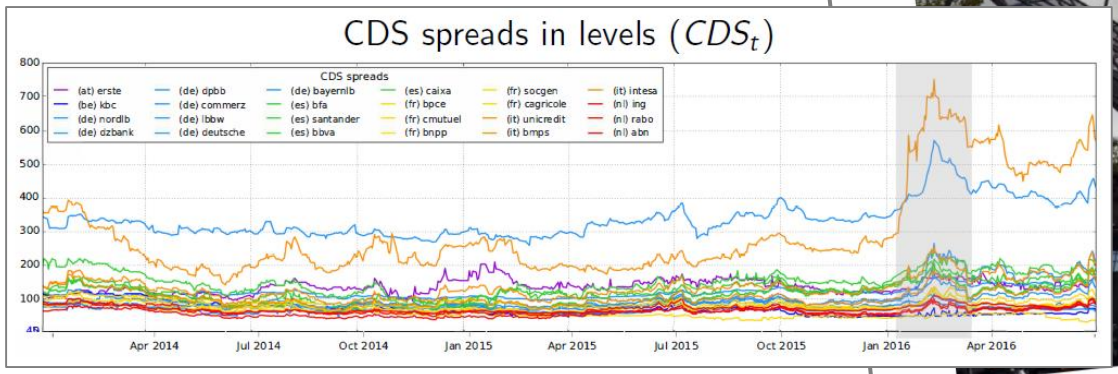
Detecting outliers in regular reporting

Stress testing IRS



How big will margin stress be if interest rates rise?

PoC 3 - Contagion in bank CDS



The Credit Spread is a Puzzle

- Credit Default Swaps (CDS) prices reflect the (perceived) credit risk of the underlying entity (ie. bank)
- To understand what drives these prices and risks, researchers usually analyze the role of *bank-specific* (stock prices, leverage ratio) or *economy-wide* (stock index, bond yields) variables. However, these **structural variables explain poorly**.
- Not only that, the part of credit risk that we cannot explain is not random. Instead, we know that there is another *hidden* variable which we failed to include.
- **Something is driving credit risk, that we cannot explain!**

Main hypothesis

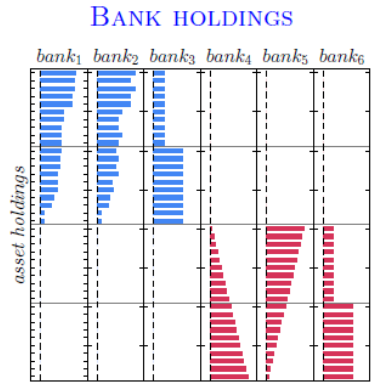
- It is not too surprising, that the bank-specific and economy-wide variables do not explain credit risk of banks very well – we didn't account for overlapping business models!
- Our hypothesis

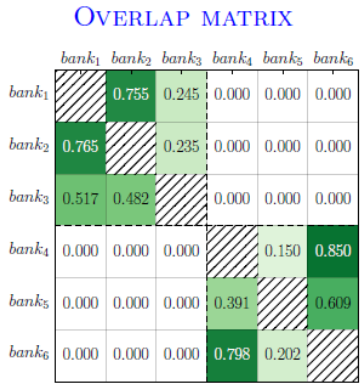
  The credit spread puzzle is a result of the commonality in banking business

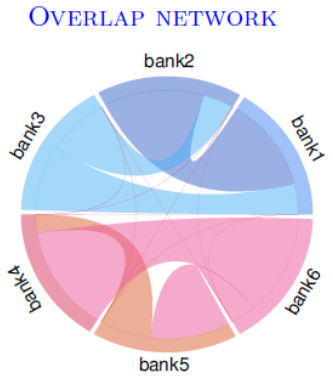
Capturing the underlying network

- Banks with similar business models (holdings) likely affect each other in stress times

➔ What is the portfolio overlap network?

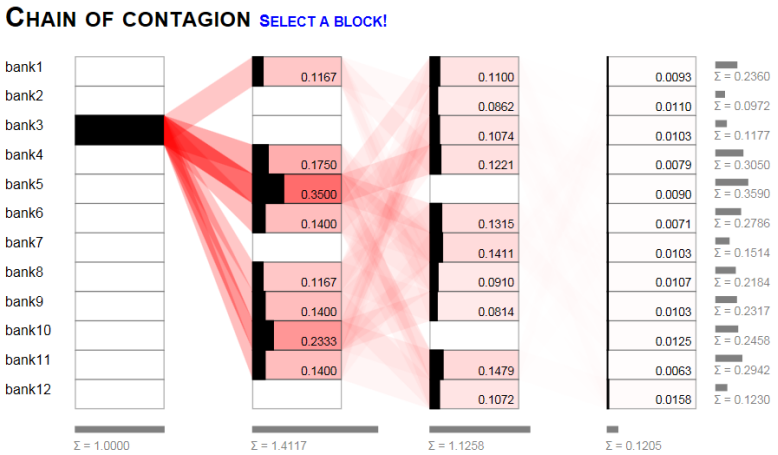






Network stress-testing

- Once the network is available, we can use our model to conduct stress-tests



http://dieter.wang/contagionchain

How do we work?

- Organisation
  - Sprint: 3 weeks, 2 week break
  - Physical location
- Agile: goals are clear for each PoC
  - Success factors for PoCs as defined in user stories
- Working towards responsible data use
  - Coding Hygiene document
  - GIT: code repository

Community

- Python & R lunches
  - Purpose: get to know each other, exchange ideas
  - Frequency: every 6 weeks
  - Big succes: 30 people on average
- Training
  - Overview of training possibilities
  - Open source (Coursera) complemented with bespoke training

Questions?

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC – Bank Indonesia International Workshop and Seminar on *"Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data"*

Bali, Indonesia, 23-26 July 2018

How do central banks use big data to craft policy?[1]

Bruno Tissot,

Bank for International Settlements

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

BANK FOR INTERNATIONAL SETTLEMENTS

How do Central Banks use Big Data to craft policy?

Bruno TISSOT
Head of Statistics and Research Support, BIS
Head of Secretariat, Irving Fisher Committee on Central Bank Statistics (IFC)

International Seminar on Building Pathways for Policy Making with Big Data – Bali, 26 July 2018
Panel discussion

Overview

- Central banks' growing interest for Big Data
- Main results of an IFC survey
- Working with Financial Big Data
- Lessons identified during the Workshop
- Looking forward

Central banks' growing interest for Big Data

- **Private sector** use big data to produce new & timely indicators
- Opportunities for **Central Banks: new type of information**
- Not just the internet: **3 key developments in Financial Big Data**
  - The **internet** of things
  - **Digitalisation**
  - Expansion of **micro financial data-sets** in the aftermath of the Great Financial Crisis of 2007-09

Central banks' growing interest for Big Data

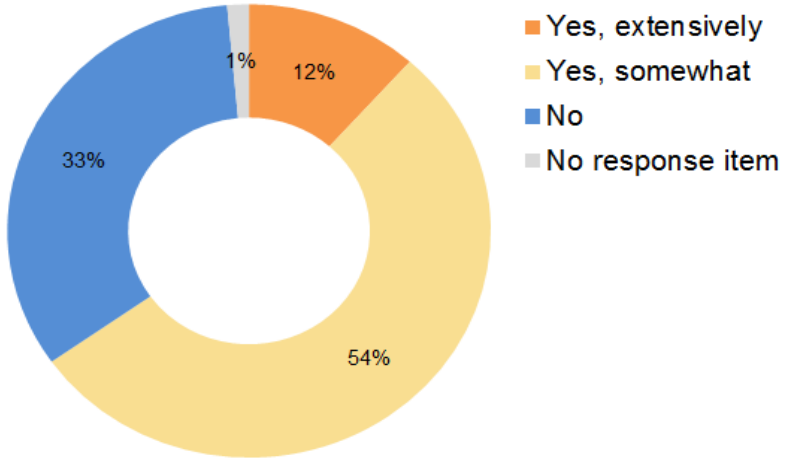
- Recent IFC survey to **assess experiences and interest**
- **Report** on www.bis.org/ifc/publ/ifc-report-bigdata.pdf
- Big Data concept is **not clearly defined**: different understanding and interest among institutions
- Not so much interest in exploring Big Data in general terms…

  … but **focus on issues <u>related</u> to central banks' mandates**

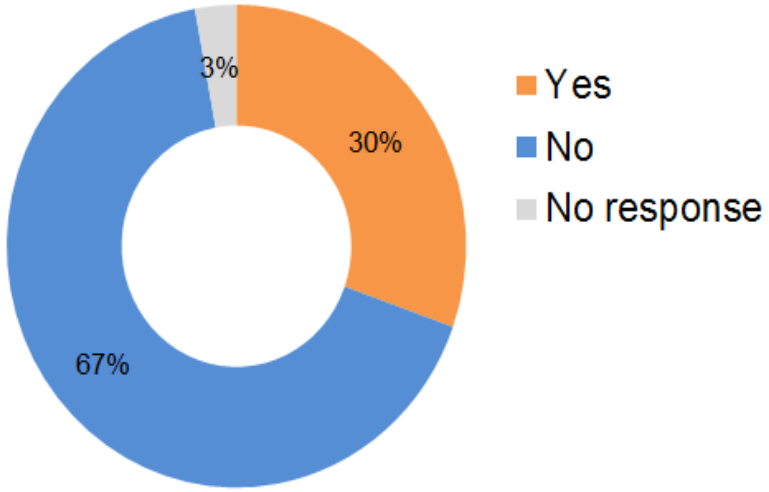
Main results of an IFC survey

- **Significant interest** in big data within the central banking community, esp. at **senior policy level**

- Yet central banks have **limited experience** in use of big data



Is the topic of big data being formally discussed within your central bank?

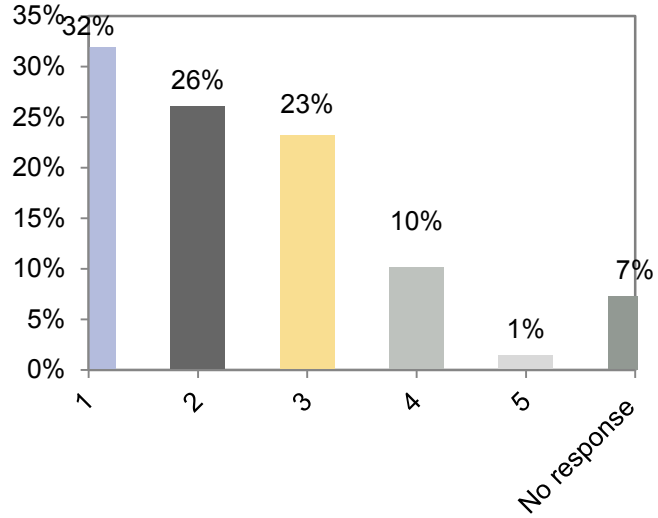


Are you already using big data sources?

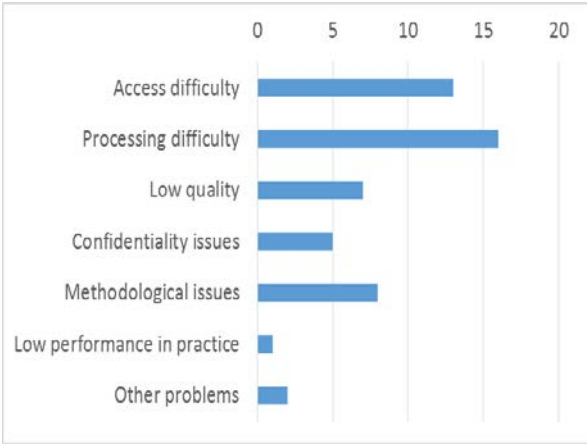
Main results of an IFC survey (2)

- About 60% of the central banks **are not ready to start a regular production** and analysis of big data

- Main **challenge relates to accessing and processing** the data



How would you rate the readiness of your central bank to start regular production and/or analysis of big data (1: low readiness/not ready to 5: high readiness)

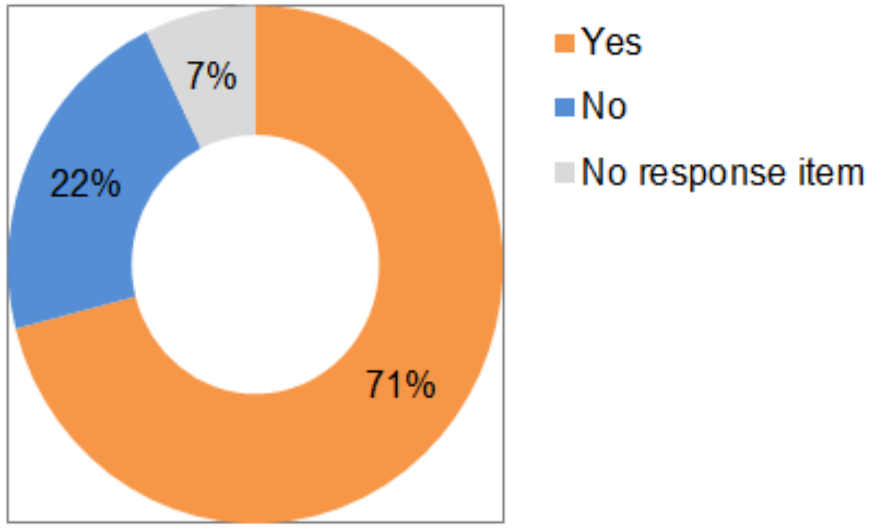


Challenges of using big data

Main results of an IFC survey (3)

- A vast majority of central banks

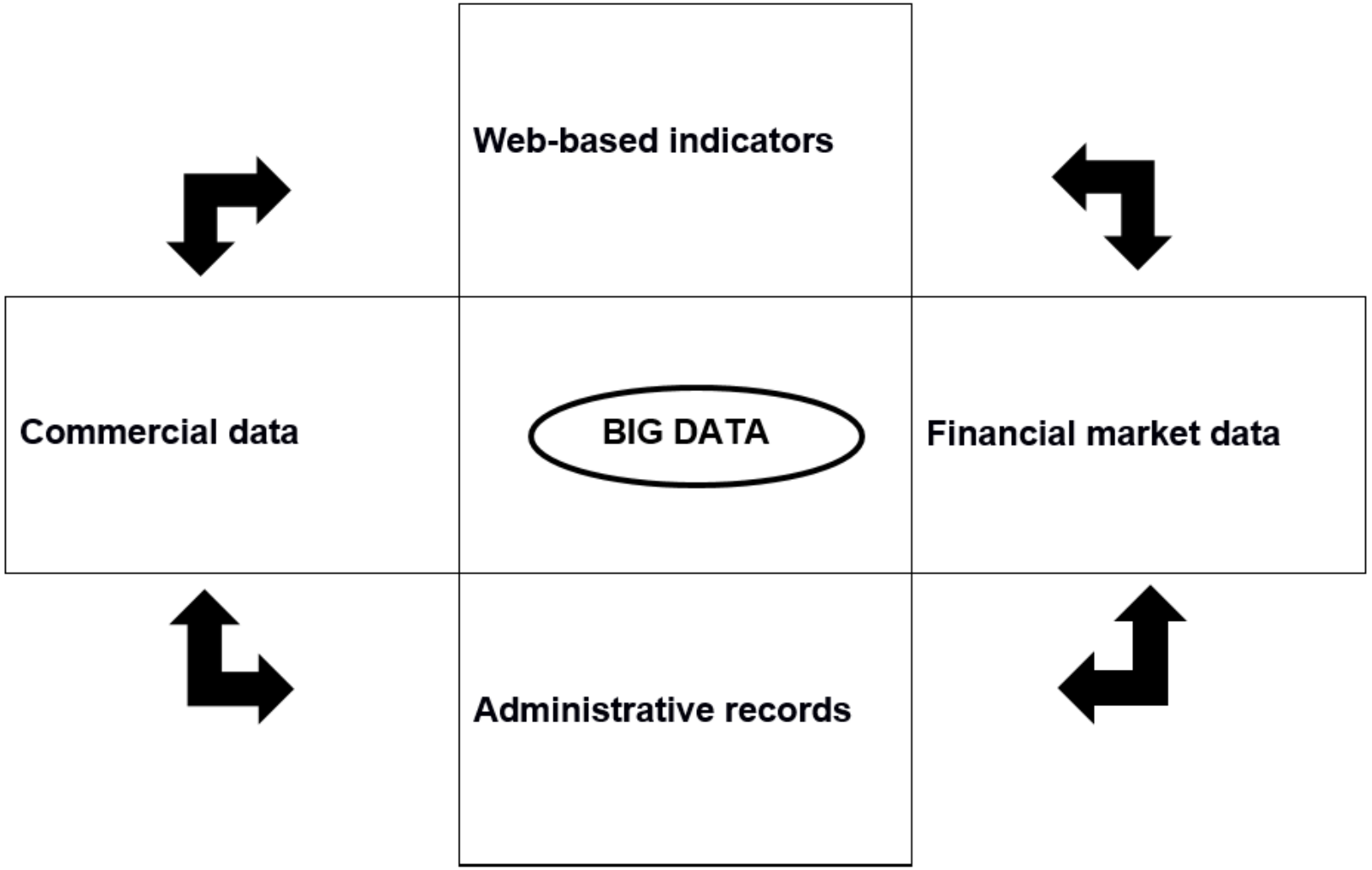
want to cooperate together



Would you be willing to cooperate with other IFC members and engage your central bank in the area of big data?

- **Selected big data pilot projects**
  - Administrative dataset (eg corporate balance sheet data)
  - Web search data set (eg Google type search info)
  - Commercial dataset (eg credit card operations)
  - Financial market data (eg high-frequency trading, bid-offer spreads)

Working with Financial Big Data: 4 main areas



Working with Financial Big Data: Exploration

- Key objective for central banks is to **better understand**
  - The new data-sets and related methodologies for their analysis
  - The value added in comparison with "traditional" statistics
- Focus on **pilots:** how big data can help to
  - Better monitor the economic and financial situation
  - Enhance the effectiveness of policy
  - Assess the impact of policy actions
- Possible tasks may well **further expand**
  - Constant creation of new information/research needs
  - Exploring behaviours in a "virtual economy"

Working with Financial Big Data: Opportunities

- **Focus on sources** that can effectively support micro- and macro-economic as well as monetary and financial stability analyses
  - Other big data – eg geospatial information – of lower interest
- **Feedback loop** inherent to policy-making authorities
  - Big data sources can affect policy-making
  - In turn policies implemented can generate new data-sets
- Big data provide **new "business opportunities"** for central banks
  - Qualitative statements to decipher central banks' communication
  - Large number of big data pools generated by financial regulations
  - In turn, big data can strengthen supervisors' capacity

Working with Financial Big Data: Challenges

- **Handling big datasets** requires significant resources and proper arrangements for managing the information
- **Using big data in policy-making** is not without risks
  - **Conveying a false sense of accuracy** and precision
  - **Undermining** effectiveness / reputation / legitimacy of **policy**
  - **Altering decision-making**

  → bias towards responding quickly and more frequently to news, encouraging shorter horizons?

  → risk of excessively fine-tuning policy communication based on perceived expectations rather than actual economic developments

Lessons identified during the workshop (1)

1. **Potential use cases have expanded for central banks,** as monetary policy-makers and micro- & macro-prudential authorities

2. **Authorities need to both have a bird's eye view of the financial system and also be able to zoom in depending on circumstances**

3. **Information needs evolve over time:** the building up of fragilities will typically require aggregated statistics to spot "abnormal patterns"; resolution work after a crisis calls for timely & granular data

Lessons identified during the workshop (2)

4. **Decisions on data have become of strategic importance**

5. **What matters** is less the way public authorities organize their information management than **the coherence of the process transforming "data" into (useful) "information"**

6. **Proper information frameworks needed to enhance the governance of big data-sets** collected / used by central banks

7. **Important challenges when accessing private information that is a by-product of commercial & administrative activities**

Looking forward (1)

- What is still unknown is whether and how far big data will trigger **a change in central banks' "business models"**
  - They are **relatively new** in exploiting big data, in contrast to the greater experience gained by statistical offices
  - They have traditionally been **data users rather than producers**, but the situation has changed since the crisis
  - Central banks are **in a key position** to ensure that big data can be transformed in useful information supporting policy

Looking forward (2)

- Big data sources of information **still under evaluation**
  - **Cooperation** (both **internationally** among authorities working on financial stability issues as well as **domestically** among statistical authorities) is the way to go to learn from each other
  - **IFC ongoing collaborative work** to explore the synergies and benefits of using big data for policy purposes
  - Other initiatives to enhance **information sharing** should be promoted