



BANCA D'ITALIA
EUROSISTEMA

Big Data & Machine Learning Applications for Central Banks

Closing Remarks by Daniele Franco
Deputy Governor of the Bank of Italy

Rome, 22 October 2019

The workshop on 'Big Data & Machine Learning Applications for Central Banks' is coming to an end. Let me begin by thanking the speakers, discussants and participants.

This is the third workshop organized by Banca d'Italia on these issues. And it will not be the last. Why are we so interested in these topics?

A few days ago, the Chairman of the Federal Reserve, Jerome Powell, noted that monetary policy is data dependent. It is adjusted over time in response to new information about the state of the economy. Monetary policy must be based on timely and accurate data.

This is why central banks have always made great efforts when it comes to collecting and analysing data.

Throughout its history, Banca d'Italia has drawn extensively on data published by the National Statistical Institute and other national and international agencies. But it has also been an active producer of statistics, not only on banking, financial and fiscal variables, but also on firms and households.

Banca d'Italia has collected micro-level statistical information on companies since the early 1950s. Back then only a few aggregate statistics were available. We began by surveying the balance sheets of a small group of firms, an activity that was progressively expanded until it became what is now the Survey on Industrial and Service Firms. These days, we collect detailed data on about 5,000 firms.

In the early 1960s, the Bank started the survey on household finance. This survey allows us to study the distribution of income and wealth at the micro level in order to analyse the preferences of individuals and the determinants of their economic and financial decisions. In recent years, we have interviewed about 8,000 households in each survey.

The digitalization of the economy is changing our lives, the organization of production and our jobs. It also poses new challenges, requiring changes in the way central banks collect and analyse data.

The typical indicators used by central banks to measure inflation, investment or economic activity may not fully capture the changes determined by digitalization.

For example, online shopping has caused a disruption in the retail sector around the globe along with changes in pricing behaviour. When measuring inflation we should take account of the prices of online stores such as Amazon or eBay.

Gross domestic product is one of the pillars of macroeconomic statistics. It is still the most important measure of economic activity and wellbeing. However, the growth of the digital economy, whose output is usually immaterial, limits the role of GDP in measuring economic activity and wellbeing.

Digitalization has dramatically changed how businesses are conducted and goods and services purchased. This brings us to the productivity paradox: productivity in industrialized countries is growing less than in past decades while economies are experiencing rapid digitalization. Jerome Powell noted that current statistics may understate productivity growth because they miss the value that we increasingly derive from new technologies, such as internet connections and smartphones.

The measurement of the digital economy is now an important challenge for statisticians and economists. We need new methodological insights.

To answer some of these questions, in 2016 Banca d'Italia put together a team to assess the potential benefits and hidden risks of Big Data, Artificial Intelligence and Machine Learning.

Despite controversies about privacy issues, despite password hacking and other worrisome aspects of online life, the world continues to embrace the internet and social media. It is estimated that more than 5.1 billion people (two thirds of the world's population) use a mobile phone and 4.4 billion people take advantage of internet services. In 2018 the number of internet users recorded 9 per cent growth (see the [We Are Social](#) blog).

So we have plenty of new data which can be mined to improve our knowledge of economic developments.

To understand and exploit these data we need to bring together different skillsets: those of economists, statisticians, engineers and computer scientists.

What are the advantages of these data? In his lecture on 'Big Data and Measurement: From Inflation to Discrimination', the MIT Professor Roberto Rigobon said that the biggest advantages of organic data (i.e. Big data) as against designed data (such as administrative data) are that: (i) they reduce the cost of collecting and providing information and (ii) they do not introduce distortions arising from the behaviour of respondents. People don't lie to their GPS or to Google.

The banking and financial industries have also been profoundly affected by these developments. Central banks must react in a timely manner and develop new tools to

carry out their missions: monetary policy, banking supervision and payment system oversight.

In these two days you have seen many applications which can improve central bank activities.

Some papers focused on the extraction of quantitative information from texts and images: this can be useful for forecasting financial market variables, predicting headline inflation, and analysing the real estate market.¹

We have seen the utility of machine learning algorithms for combating illicit financial activities and increasing the statistical accuracy of micro- and macro-economic indicators (for example, forecasts of the Index of Industrial Production).²

Some indicators have moved rapidly from the experimental to the production stage. For example, at Banca d'Italia we have developed and now regularly use an indicator of the situation of the main banks obtained using selected tweets.³

We have improved our estimates of Foreign Direct Investment for the Balance of Payments by introducing Machine Learning classifiers.⁴

We are also working on other sources of information, such as credit and debit card payment data, which should help us to assess consumption, retail sales and other macroeconomic variables.⁵

These developments present significant challenges. I will mention three.

First, there is a risk of biased statistical inference. Big data and, in particular, unstructured data collected from social media may reflect only a part of the population.

¹ See M. Loberto, 'The rental market in Italian cities', Banca d'Italia, Temi di Discussione (Working Papers), 1228, 2019. See M. Loberto, A. Luciani, and M. Pangallo, 'The potential of big housing data: an application to the Italian real-estate market', Banca d'Italia, Temi di Discussione (Working Papers) ',1171, 2018. See also M. Bernardini, P. Cariello, M. De Leonardis, J. Marcucci, F. Quarta, and A. Tagliabracci, 'Deriving Indicators from a Large Corpus of Italian Documents', Bank of Italy Workshop on Big data & Machine Learning Applications for Central banks, October 2019; V. Astuti, G. Bruno, S. Marchetti, and J. Marcucci, 'News and Banks' Equities: Do Words Have Predictive Power?', Bank of Italy Workshop on Big data & Machine Learning Applications for Central banks, October 2019.

² See V. Aprigliano, G. Bruno, S. Emiliozzi, S. Marchetti, J. Marcucci, and D. Nicoletti, 'Neural Networks for Macroeconomic Forecasting', Bank of Italy Workshop on Big data & Machine Learning Applications for Central banks, October 2019.

³ See M. Accornero and M. Moscatelli, 'Listening to the buzz: social media sentiment and retail depositors' trust', Banca d'Italia, Temi di Discussione (Working Papers), 1165, 2018.

⁴ A. Carboni and A. Moro, 'Imputation techniques for the nationality of foreign shareholders in Italian firms', External sector statistics: current issues and new challenges. Irving Fisher Committee on Central Bank Statistics, 2018.

⁵ See G. Ardizzi, S. Emiliozzi, J. Marcucci and L. Monteforte, 'News and consumer card payments', Banca d'Italia, Temi di Discussione (Working Papers), 1233, 2019; V. Aprigliano, G. Ardizzi and L. Monteforte, 'Using the payment system data to forecast the Italian GDP', Banca d'Italia, Temi di Discussione (Working Papers), 1098, 2017.

Increasing the sample does not improve the accuracy of the estimates if the source of the bias is not understood and dealt with. The use of massive amounts of rough data may result in an exceptionally good in-sample fit but may perform poorly out-of-sample (the curse of overfitting).

Moreover, when using Big data, special attention should be paid to the robustness and stability of the estimates over time.

The second challenge is privacy. Protecting the integrity and confidentiality of personal data is of paramount importance to our societies. Storing personal data in digital repositories belonging to public and private institutions can be risky. We should avoid any use for illegitimate purposes such as the second-hand trading of personal information.

At Banca d'Italia, we are exploring new avenues for utilizing microdata without jeopardizing their confidentiality. In this way data from different institutions can be combined to produce aggregate statistics that are important for understanding economic, financial and welfare issues. We have started experimenting some privacy-preserving algorithms with Istat. We are also discussing these issues with Eurostat and the National Authority for Personal Data Protection.

The third challenge concerns the preservation of data over time. While we keep producing a huge amount of data, we have not yet agreed on a shared set of criteria for making these data accessible in the future. It is crucial that technology-neutral standards be established to ensure data availability for future generations. Coordination between statistical agencies, governments, academia and private companies is equally crucial.

Timely and accurate information derived from non-traditional data sources, as well as new analytical techniques can help central banks to improve their knowledge of the economy and society, allowing them to make better-informed data-driven decisions.

Banca d'Italia is firmly committed to carrying out further research on these issues and to cooperating with other institutions to strengthen methodologies and applications.

Future events will certainly be organized to share the results of our research activity.