BANCA D'ITALIA
EUROSISTEMA

Nontraditional Data & Statistical Learning with Applications to Macroeconomics

Welcome address by Daniele Franco
Senior Deputy Governor of the Bank of Italy

Rome, 11 November 2020

Ladies and Gentlemen,

I am glad to open this virtual conference jointly organized by Banca d'Italia and the Federal Reserve Board. I would like to welcome all the participants from more than 60 countries.

1. This conference is about the impact of digital activities on economic analysis.

An astonishing amount of digital activity is occurring at any given moment. This explosion stems from the aggregate action of 4.5 billion internet users, a number that is projected to rise even further in the coming years.

Since the outbreak of COVID-19, technology has played an even greater role in our daily lives.

This development has increased the production of data. In fact, data are constantly being generated by our clicks, reactions on social media, shares, streaming of videos, digital transactions, digital recordings of our personal and work activities, digital circulation of scientific texts, and so on and so forth.

These data can give us a better understanding of the state of the economy at both the micro and macro level, provided – obviously – that we have access to them.

Data ownership is therefore extremely important.

As recently highlighted in *The Economist* ('Who owns the web's data?', 22 October 2020), we are now experiencing a sort of 'digital feudalism', where a handful of big tech companies have full control over data.

Tech companies such as Facebook, Amazon, and Google have reached an impressive position, becoming some of the biggest companies in the world. In this process, they have caught up with longer-standing titans like Apple and Microsoft.

The use of data is actually one of the world's biggest businesses. We see this in the market capitalization of some companies, like Alphabet (the owner of Google) and Facebook. Even once the value of their cash, physical and intangible assets, and accumulated R&D has been stripped out, their capitalization remains huge.

These developments are causing increasing concerns among policymakers, from both a political and economic point of view. As to the former, the big techs' business model depends on network effects to influence users and sell more advertising, which may result in collective behaviour that can harm public discourse. Economic concerns stems from the fact that the size of big tech firms may shrink competition.

2. Banca d'Italia has always based its decisions on the data. Micro-level statistical information on firms has been collected 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.

At the beginning of the 1960s, the Bank began surveying household finance. Nowadays, this survey allows us to evaluate income and wealth distribution at the micro level in order to analyse individual preferences and the determinants of their economic decisions. In recent years, we have interviewed a sample of more than 8,000 households in each survey; about 50 per cent of them were part of a panel.

In 2016, Banca d'Italia started its project on Big Data and Machine Learning applications. We wanted to fine-tune the accuracy of our economic analysis by combining all available skills ranging from economics to statistics, from computer science to engineering, and from mathematics to applied physics.

This year we have stepped up our action by strengthening the team with new resources from universities and by widening our data collection from private companies.[1]

3. The COVID-19 pandemic has radically transformed our lives in a number of different ways. Direct personal contacts have been partly replaced by online events. Most meetings around the world are now organized via multiple IT applications. This has allowed many of us to keep working in safety. It has also reduced both travelling cost and carbon emissions.

These changes have been made possible by the progress of IT in many areas, from core technology advances – like further doublings of basic computer power every 18 months – to successful investment in complementary innovations, like cloud infrastructure.

One of the most important developments is the wave of recent improvements in artificial intelligence, especially machine learning.

---

1 Such as, for example, Immobiliare.it, Mutuionline.it, Twitter, Indeed, Dow Jones, etc.

Machine learning represents a paradigm shift from the first wave of computerization. Historically, most computer programs were created by accurately codifying human knowledge, mapping inputs to outputs as prescribed by the programmers.

In contrast, machine-learning systems use categories of general algorithms (e.g. neural networks, random forest and gradient boosting) to figure out mappings and are typically driven by voluminous amounts of data. By employing huge data sets and big data processing resources, machines have made impressive progress. For instance, error rates in labelling the content of photos on ImageNet, a dataset of over 10 million images, have fallen from over 30 per cent in 2010 to as low as 2.2 per cent in 2019. This has, inter alia, big implications for the identification of fake banknotes, which is part of our central bank work.

The exponential growth of digital data, such as images, videos and speeches, from numerous sources (e.g. social media, Internet-of-Things, etc.) is driving the search for tools that allow us to extract relevant information from disparate kinds of data.

Private companies have played an important role in collecting and organizing behavioural microdata on firms and households. In this respect, we can mention a recent paper on the economic impact of COVID-19 by Professor Raj Chetty and his team at Opportunity Insights.[2] Using anonymized data from a large number of private companies, Professor Chetty and his co-authors develop a publicly accessible platform, which measures some important macroeconomic variables such as spending, employment, and other outcomes both at a very high frequency, and at a very granular level.

Their work shows how new nontraditional data can be used to obtain quasi-real-time insights into the effects of the COVID-19 pandemic. Along the same lines, a recent study by colleagues of Banca d'Italia, Banque de France and the Bundesbank points to the role of nontraditional data and indicators for a rapid assessment of the implications of the pandemic.[3]

4. Banking and financial industries have been profoundly impacted by these developments. Central banks are now using big data to improve their capacity to carry out their missions: monetary policy, banking supervision and payment system oversight.

Over the next two days, we will hear about several studies applying new data sources to different issues relevant to central bank activities.

Some papers focus on the new data and tools available to improve the trade-offs between the objectives for inflation, employment/growth and financial stability. Others show the advantages provided by machine learning methods for estimating and

---

2 'The Economic Impacts of COVID-19: Evidence from a New Public Database Built from Private Sector Data', NBER Working Paper No. 27431 DOI 10.3386, June 2020. Professor Raj Chetty is 2013 John Bates Clark Medal and is a prominent member of the Harvard faculty.

3 See G. Veronese, C. Biancotti, A. Rosolia, R. Kirchner and F. Mouriaux (2020*),* 'COVID-19 and Official Statistics: A Wakeup Call?', Banca d'Italia, Deutsche Bundesbank and Banque de France. Forthcoming in Proceedings of the *8th IMF Statistical Forum: Measuring the Economics of a Pandemic*.

forecasting macroeconomic variables and the relevance of nontraditional sources of information, especially when it is impossible to run a classical survey, as was the case during the COVID-19 pandemic.

Besides its long history of cooperation with commercial banks, with Istat (our National Statistical Institute) and other public institutions, Banca d'Italia has already started to collect nontraditional data from social media, newspapers, payment platforms and on real estate and home mortgages.

The harnessing of larger data sets, faster computational power, and more advanced analytical techniques is spurring progress on a range of micro- and macroeconomic problems.

At Banca d'Italia, we approached the data deluge by building a multidisciplinary team of economists, statisticians, econometricians, mathematicians, engineers and IT specialists. By connecting these data experts with their frontline colleagues, we began to identify projects that can increase both the accuracy and performance of our policy action.

5. The Big Data revolution poses some fundamental challenges. Let me briefly mention two of them.

First, big data can lose their statistical relevance if they are not used in a methodologically sound way. Such data typically target only a particular layer of the population; increasing the sample size will not improve the accuracy of estimates if the estimation methods do not correct for this kind of distortion. Social media data for example, like many other internet-based sources of big data, are usually based on a biased sample of the population.

Second, the availability of a huge amount of data raises the importance of its integrity, confidentiality and privacy. Personal data protection is central to our societies. The sheer amount of personal data now available, and the growing ease with which individual information can be connected across databases, have far-reaching implications for privacy and freedom. Indeed, this has prompted the development of regulations concerning the treatment of digital data (think of the GDPR or the CCPA in California).[4]

Rules on data management are sometimes different across jurisdictions and data domains. Ideally, one would like to have global standards. Therefore, international cooperation is to be encouraged as far as possible.

Furthermore, the adoption of machine learning and artificial intelligence requires as input huge amounts of granular and unstructured data, which should be collected through close collaboration between public institutions and private companies. We also need specific platforms or data lakes to store and analyse such data so as to preserve

---

[4] The General Data Protection Regulation (GDPR) is the European Union law on data protection and privacy; the California Consumer Privacy Act (CCPA) is the statute aimed at enhancing privacy rights and consumer protection for residents of California.

privacy and security. This can be done through a collectively trusted cloud where all actors, both public and private, can operate to gain insights for their core business.

The availability of privacy-preserving algorithms or the use of federated learning, recently suggested by researchers at Google, could provide the enabling technology. Federated learning allows machine learning algorithms or artificial intelligence on data to be distributed on different servers, avoiding any data exchange. This will require further investment from both the public and the private sector. It will also require tighter cooperation between the private sector, which typically owns most of the new data, and the public sector, which uses such data for policy reasons and for the common good.

6. The challenges generated by big data and machine learning technologies require continuous organizational and technological innovation in the institutions that want to use them effectively and fairly. Banca d'Italia has taken several steps in recent years to enhance its ability to collect, store and exploit huge amounts of data, using state-of-the-art hardware and software. Other institutions are doing the same.

Let me conclude my talk by thanking, once again, all the speakers and participants for joining us today. We hope to welcome you here in person in the future. Special thanks go to those who have contributed to organizing the workshop, which brings together leading economists, statisticians, artificial intelligence and machine-learning specialists, data scientists from about 60 central banks, 57 universities, 24 government agencies, 7 multilateral organizations, and 20 private companies. This will ensure a broad variety of perspectives and a lively discussion.

I am sure that you can count on having two interesting and productive days.

Designed by the Printing and Publishing Division of the Bank of Italy