BANCA D'ITALIA
EUROSISTEMA

Harnessing Big Data & Machine Learning Technologies for Central Banks

Opening Remarks by the Deputy Governor of the Bank of Italy

Fabio Panetta

Rome, 26 March 2018

It is my great pleasure to open this workshop on 'Harnessing Big Data & Machine Learning Technologies for Central Banks', and to welcome the speakers and participants. I would like to take this opportunity to thank the organizers for putting together such a broad and topical program.

Big Data and machine learning are the products of digital technology, whose widespread adoption has important implications for how communication occurs, education is delivered, and knowledge is spread. As we become accustomed to the new digital environment, we are seeing significant shifts in society. Around 80 per cent of Europe's citizens own a smartphone, personal computer or tablet. The internet is widely used to gather information, communicate and carry out activities that directly or indirectly affect people's behavior.

In Italy 80 per cent of consumers have access to the Internet via a smartphone; half of them have direct access to their checking account and buy goods and services online. Around one third of Italian firms has automated both administrative and back-office activities.

The effects of these changes are particularly evident in the financial sector. Globally, the number of users of payment solutions offered by tech giants such as Apple, Google, Samsung or Android has risen from 18 to 144 million in the last two years alone. Amazon Lending[1] issued more than one billion dollars in loans last year (from June 2016 to June 2017), three times more than the average over the previous five years.

In broad terms Big Data is the application of new techniques to digital information on a size and scale that extends well beyond traditional approaches.[2] The word 'big' encompasses the volume, diversity and speed with which the data are generated. This idea is synthesized in the three V's: volume, velocity and variety, as illustrated below.

---

[1] Amazon Lending, the lending arm of Amazon, is a financing service for firms selling through the Amazon platform.

[2] Lohr, S., 'How Big Data Became So Big', *The New York Times*, 11 August (2012).

Big Data: expanding on three fronts at an increasing rate
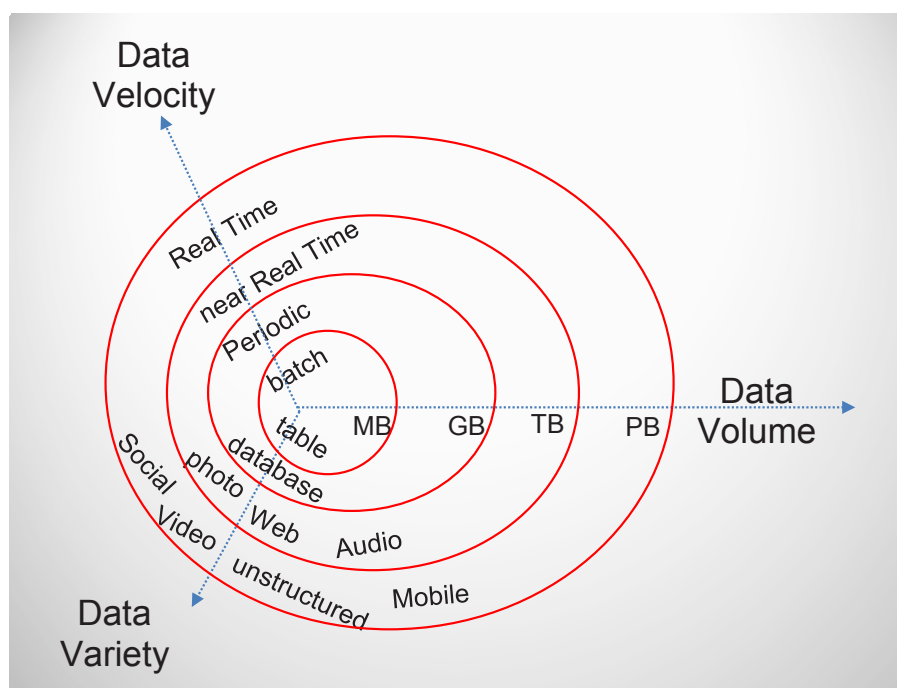
Central banks could, and definitely should, play an active role in exploiting digital technologies and the enormous amount of data they generate.

The flow of high-volume, high-velocity and high-variety information on agents' preferences and choices can be put to good purpose by policy makers to make better decisions. Along with the traditional data stored in well-defined records, today mobile networks and social media generate a 'data rainforest' with diverse sources of semi-structured (such as XML or JSON) and unstructured data (audio/video and free text). The traditional and the new data sources can be used to construct better and more timely measures of economic activity.

There are prominent examples of Big Data being used for policy analysis. Big Data is used to estimate unemployment rates or the inflation rates, to improve the forecasts of policy-relevant variables, to compute measures of sentiment of consumers and firms.[3]

---

[3] For example, D'Amuri and Marcucci (2017) show that models using Google Trends data give better predictions of US unemployment than those using standard leading indicators. Using online data from the Billion Price Project (BPP), Cavallo (2013) shows that while Argentina's government announced an average annual inflation rate of 8 per cent from 2007 to 2011, the online data suggested it was actually over 20 per cent, in line with the estimates of some provincial governments, and consistent with the results drawn from surveys of household inflation expectations. Goolsbee and Klenow (2018) use online transactions for millions of products from 2014 to 2017 to demonstrate that online inflation in the US is lower than in the CPI (by around 1.3 percentage points per year) for the same categories. In a similar vein, Daas and Puts (2014) compute a measure of sentiment of the Dutch public using social media messages, which are highly correlated and cointegrated with the monthly consumer confidence index obtained from the classic surveys.

The number of potential applications for central banks is enormous, but there are challenges ahead. The technical issues will be discussed in detail in this workshop, but I would like to mention one general, methodological problem now: before using the relationships estimated from the Internet or social media for policy, we have to first ascertain that they are sufficiently robust, representative, and reliable. To do this we have to invest in research.

An important field that is deeply affected by the innovative technologies designed to extract value from Big Data is financial stability. Fintech firms use Big Data and new technologies to create new products and grant continuous access to financial services. This is a welcome development, as it increases competition and stimulates productivity. However, the development of new lines of business and closer, direct interconnections between banks, other intermediaries and investors may well affect the stability of the system. Public authorities must carefully examine the changes induced by digital technologies and value innovative projects to preserve the stability, the efficiency, and the security of the financial sector.

The complexity is increased by the fact that central banks are at once both users and producers of information and Big Data. Proof of this is the collection of highly granular data on individual bank loans by the Single Supervisory Mechanism (the so called AnaCredit), on daily money market statistical reporting by the European System of Central Banks, and on the trading repositories envisaged by the European market infrastructure regulation.

The sheer amount of these granular data makes it clear that the necessary validation processes cannot be performed manually, but will require sophisticated algorithms and techniques such as machine learning or artificial intelligence.

Central banks should master these new technologies. The Bank of Italy has created an internal multidisciplinary team on Big Data which includes economists, statisticians and computer scientists from different departments, working in close cooperation with the Directorate General for Information Technology. The team has built a hardware and software infrastructure in order to deal with different kinds of Big Data for both macroeconomic and microeconomic issues. Some of the results will be discussed during this workshop. The analyses focus on using unstructured textual data from social media, in particular Twitter, to compute inflation expectations[4] or to gauge retail depositors' trust.[5]

---

[4] See C. Angelico, J. Marcucci, M. Miccoli, and F. Quarta, 'Can We Measure Inflation Expectations Using Twitter?'.

[5] M. Accornero and M. Moscatelli in 'Listening to the buzz: social media sentiment and retail depositors' trust', a measure of retail depositors' trust by checking Twitter comments. Banca d'Italia, Temi di Discussione (Working Papers), 1165 (2018).

Social media are also used to assess customers' sentiment towards specific companies and its effect on stock returns, volatility and trading volumes.[6] Twitter and news are used to measure economic policy uncertainty and to focus on payment card scams, to then relate them to trends in consumer payments. Another strand of research uses data from single online real estate ads extracted from the web to understand the microstructure of the Italian real estate market.[7]

Let me conclude by thanking once again all the speakers, discussants, panelists and participants for being here. I look forward to hearing more about your findings on Big Data and machine learning, and their applications for central banks and policy analysis. It is important that this workshop brings together researchers from central banks and academia, who can provide a broad variety of perspectives. I am sure we will have two productive and interesting days.

---

[6] See G. Bruno, P. Cerchiello, J. Marcucci, and G. Nicola, 'Twitter Sentiment and Banks' Equities: Is there any causal link?'.

[7] See M. Loberto, A. Luciani, and M. Pangallo, 'The Potential of Big Housing Data: an application to the Italian Real Estate Market', based on data from www.immobiliare.it.

References

Accornero M. and M. Moscatelli, 2018, "Listening to the buzz: social media sentiment and retail depositors' trust", Bank of Italy Working Paper, No. 1165.

Angelico C., J. Marcucci, M. Miccoli and F. Quarta, 2018, "Can We Measure Inflation Expectations Using Twitter?", mimeo.

Bruno G., P. Cerchiello, J. Marcucci and G. Nicola, 2018, "Twitter Sentiment and Banks' Equities: Is there any causal link?", mimeo.

Cavallo A., 2013, "Online and official price indexes: Measuring Argentina's inflation", *Journal of Monetary Economics*, 60(2): 152-165.

Daas P. J. H. and M. J. H. Puts, 2014, "Social Media Sentiment and Consumer Confidence", ECB Statistics Paper Series, n. 5.

D'Amuri F. and J. Marcucci, 2017, "The predictive power of Google searches in forecasting US unemployment", *International Journal of Forecasting*, 33(4): 801-816.

Goolsbee A. D. and P. J. Klenow, 2018, "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce", Stanford University working paper.

Loberto M., A. Luciani and M. Pangallo, 2018, "The Potential of Big Housing Data: an application to the Italian Real Estate Market", mimeo.

Lohr S., 2012, *How Big Data Became So Big*, The New York Times, 11 August 2012.