Session 6: Administrative and Alternative Data Sources

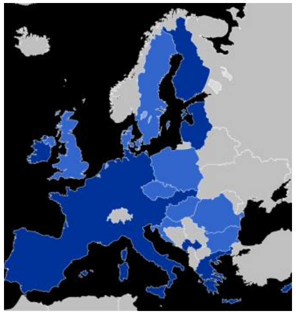
Big Data and Macroeconomic Nowcasting: From Data Access to Modelling

Dario Buono*, European Commission, Eurostat, dario.buono@ec.europa.eu
Stephan Krische*, GOPA Consultants, stephan.krische@gopa.de
Massimiliano Marcellino, Bocconi University, massimiliano.marcellino@unibocconi.it
George Kapetanios, King’s College, george.kapetanios@kcl.ac.uk
Gian Luigi Mazzi, European Commission, Eurostat, gianluigi.mazzi@ec.europa.eu
Fotis Papailias, Queen's University Management School, f.papailias@qub.ac.uk

*The views expressed are the author’s alone and do not necessarily correspond to those of the corresponding organisations of affiliation

15th Conference of International Association for Official Statistics
Abu Dhabi, 6–8 December 2016

Eurostat, the Statistical Office of EU



- About 700 people with 28 different nationalities
- Statistical Office of European Union, part of EC
- Core business:
  - **Euro-zone (19) & EU (28) aggregates**
  - **harmonization, best practices, guidelines, trainings & international cooperation**
- Methodology team: Time Series, Econometrics, SDC, Research & EA

Why interested in Big Data for nowcasting?

- **Big Data** are complementary information to standard data, being based on **different information sets**
- More **granular** perspective on the indicator of interest, both in the temporal and cross-sectional dimensions
- It is **timely** available, generally **not subject to revisions**

European research project: Apr 15 to Jul 16

eurostat

QUEEN'S UNIVERSITY
MANAGEMENT
SCHOOL

Queen's University
Belfast

GOPA
WORLDWIDE CONSULTANTS
MEMBER OF GOPA CONSULTING GROUP

KING'S
College
LONDON

UNIVERSITA COMMERCIALE
·LUIGI BOCCONI MILANO·

Research questions and findings

Can Big Data help for Macroeconomic Nowcasting?

What are the potential Big Data sources?

1. Literature review
2. **Models/methods** to be used for Big data
3. **Recommendations** on how to handle Big Data
4. **Case study**: IPI, Inflation, unemployment of some EU countries

Big Data types & dimensionality

- When the dimensionality increases, the volume of the space increases so fast that the available data become **sparse**.
- For statistically significant result, the amount of data needed often grows exponentially with the dimensionality.
- Use of a typology based on Doornik and Hendry (2015):
  - **Tall** data: many observation, few variables
  - **Fat** data: many variables, few observations
  - **Huge** data: many variables, many observations

Models race

- Dynamic Factor Analysis
- Partial Least Squares
- Bayesian Regression
- LASSO regression
- U-Midas models
- Model averaging

255 models tested, **macro-financial** & **google trend** data

Statistical Methods: findings

- Sparse regression (LASSO) works for fat, huge data
- Data reduction techniques (PLS) helpful for large variables
- (U)-MIDAS or bridge modelling for mixed frequency
- Dimensionality reduction improves nowcasting
- Forecast combination: Data-driven automated strategy with model rotation based on forecasting performance in the past works well

From Data Access to Modelling

Step-by-step approach, accompanied by specific recommendations for the use of big data for macroeconomic nowcasting, guiding to

- **the identification and the choice of Big Data**
- **pre-treatment and econometric modelling**
- **the comparative evaluation of results to obtain a very useful tool for decision about the use or not of Big Data**

Step 1: Big Data usefulness within a nowcasting exercise

Recommendations

1. *Evaluate the **quality** of the existing nowcasts and identify issue (bias or inefficiency or large errors in specific periods), that can be fixed by adding information in Big Data based indicators*
2. *Use of Big Data only when expecting to improve the timeliness and/or the quality of nowcastings*
3. *Do not consider Big Data sources with **spurious correlations** with the target variable*

Step 2: Big Data search

Recommendations

1. *Starting point for an assessment of the potential benefits/costs of the use of Big Data for macroeconomic nowcasting: identification of their source*
   - **Social Networks (human-sourced information)**
   - **Traditional Business Systems (process-mediated data)**
   - **Internet of Things (machine-generated data)**
2. *Choice is heavily dependent on the target indicator of the nowcasting exercise*

Step 3: Assessment of big-data accessibility and quality

Recommendations

1. *Privilege data providers with guarantee of **continuity** and of the availability of a good **metadata** associated to the Big Data*
2. *Privilege Big Data sources ensuring sufficient time and cross-sectional coverage*
3. *If a bias is observed a **bias correction** can be included in the nowcasting strategy.*
4. *To deal with possible instabilities of the relationships between the Big Data and the target variables, nowcasting models should be **re-specified on a regular basis** (e.g. yearly) and occasionally in the presence of unexpected events.*

Step 4: Big data preparation

Recommendations

1. *Big data often unstructured: proper mapping*
2. *Pre-treatment to remove deterministic patterns*
   - **Outliers, calendar effects, missing observations**
   - **Seasonal and non-seasonal short-term movements should be dealt accordingly to the characteristic of the target variable**
3. *Create a **specific IT environment** where the original data are collected and stored with associated **routines***
4. *Ensure the availability of an **exhaustive documentation** of the Big Data conversion process*

Step 5: Big Data modelling strategy Recommendations

1. *Identification of appropriate econometric techniques*
2. *First dimension: choice between the use of methods suited for large but not huge datasets, therefore applied to summaries of the Big Data (Google Trends)*
   - **nowcasting with large datasets can be based on factor models, large BVARs, or shrinkage regressions**
3. *Huge datasets can be handled by* ***sparse principal components****, linear models combined with heuristic optimization, or a variety of* ***machine learning*** *methods such as* ***LASSO & LARS regression***
4. *In case of mixed frequency data, methods such as UMIDAS and, as a second best, Bridge, should be privileged.*

Step 6: Results evaluation of Big Data based nowcasting

Recommendations

1. *Run a critical and comprehensive **assessment of the contribution** of Big Data for nowcasting the indicator of interest based, e.g., on standard criteria such as **MSE or MAE**.*
2. *In order to reduce the extent of data and model snooping, a cross-validation approach should be followed:*
   - **various models and indicators, with and without Big Data, estimated over a first sample and selected and/or pooled according to their performance**
   - **then the performance of the preferred approaches re-evaluated over a second sample**

Case study

*- Implementation of all these steps for nowcasting **IP growth, inflation and unemployment in several EU countries** in a **pseudo out of sample context**, using Google trends for specific and carefully selected keywords for each country and variable*

- Big Data specific features: transform unstructured into structured data, time series decompositions, handling mixed frequency data

*- Overall, the <u>**results are mixed**</u> but there are several cases where Google trends, when combined with rather sophisticated econometric techniques, yield forecasting gains, though generally small.*

- Gains in term of timeliness or revisions have not been considered

Literature contribution

Big Data and Macroeconomic Nowcasting: from data access to modelling

[illegible]

2016 edition

eurostat

Eurostat Statistical Working Paper

"Big Data and Macroeconomic Nowcasting: From data access to modelling"

- Methodological finding will be included in 2 chapter of the **Eurostat/UNECE Handbook on Rapid Estimates** currently under 2nd peer review, (forthcoming in 2017)

What's next? Big Data Econometrics

2017, a new project focusing on:

- Econometrics, Filtering issues, advanced Bayesian estimation and forecasting methods
- **Real time** empirical evaluations (including a direct comparison with Eurostat flash estimates),
- **New ways and new metrics** to present nowcasts
- Possible data **timeliness/accuracy gains**
- Big data handling tool developed as **R package**
- Scientific summary for Big Data Econometric **strategy**

Thank you for your attention!!

Some References:

- Eurostat, Big data and macroeconomic nowcasting, preliminary results presented at the ESS methodological working group (7 April 2016, Luxembourg) http://ec.europa.eu/eurostat/cros/content/item21bigdataandmacroeconomicnowcastingslides_en

- Big data CROS portal, http://ec.europa.eu/eurostat/cros/content/big-data_en

- Marcellino, M. (2016), "Nowcasting with Big Data", Keynote Speech at the 33rd CIRET conference.

- Harford, T. (2014, April). Big data: Are we making a big mistake? Financial Times. Available at http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html #ixzz2xcdlP1zZ

- Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014). "The Parable of Google Flu: Traps in Big Data Analysis", Science, 143, 1203-1205.

- Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso", Journal of the Royal Statistical Society B, 58, 267-288.