



THE ROLE OF BIG DATA FOR NOWCASTING AND FORECASTING MACRO EVOLUTIONS

L. FERRARA

LAURENT FERRARA
ECONOMIC AND FINANCIAL REGULATION IN THE ERA
OF BIG DATA CONFERENCE

NOVEMBER 24, 2017

» WHAT DO WE MEAN BY « BIG DATA »?

- A difference has to be done between large « official data » and « alternative data »
- « Official data »
 - Released by National Statistical Institutes or Central Banks
 - Disaggregated / Granular data (micro and macro)
- « Alternative data » = **Big data**
 - Stemming from Google Trends, web scraping, scanner data, crowdsourcing with mobile phones ...
 - Data issues: outliers, structural breaks, seasonalities, filters
 - Reliability of the data and revisions (cf. debate H. Varian vs S. van Norden)
- From « Official » to « Big » data ⇔ **From the 3 V's to the 5 V's**
 - (i) Volume / (ii) Variety (source and formats) / (iii) Velocity
 - (iv) *Variability* / (v) *Veracity*



MODEL CHARACTERISTICS TO ACCOUNT FOR « BIG DATA »

Base equation:

$$y_t = \beta_1 x_{1t} + \dots + \beta_k x_{kt} + \varepsilon_t \text{ with } \varepsilon_t \sim N(0, \sigma^2)$$

- Big data = large k
- Models have to account for:
 - Computational efficiency
 - Frequency mismatch between y and x_i (mixed data sampling approaches)
 - Ragged-edge data (different reporting lags)
- Standard statistical inference is not a good idea
 - Too many parameters to estimate
 - High degree of uncertainty in estimates
 - Overfitting and poor out-of-sample accuracy



MODEL CHARACTERISTICS TO ACCOUNT FOR « BIG DATA »

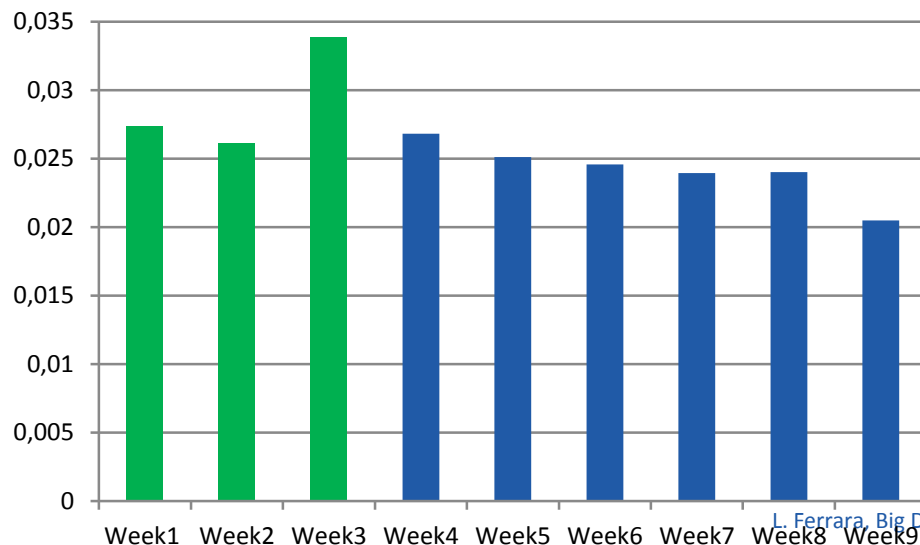
- Methods to address curse of dimensionality (Giannone, Lenza, Primiceri 2017)
 - **Sparse models** : Lasso = some β are constrained to 0
 - **Dense models** : Dynamic Factor Models (DFM) = all the variables have a role to play)
- Are more data always needed ?
 - No: Theoretical evidence by Boivin and Ng (2005) when forecasting with Dynamic Factor models (DFM)
 - No: Empirical evidence by Barhoumi, Darné, Ferrara (2010) on nowcasting French GDP
 - A good idea: Pre-selection of data = **Focus on core sets**:
 - Ex: Targeted DFM by Bai and Ng (2008), Application by Schumacher (2011) on German GDP

- Is there a gain from using big data?
 - **No significant gain when controlling by « official » data (hard, soft and financial data)**, see e.g. Choi and Varian (2012) vs Li (2016)
- When is there a gain?
 - **Significant gain when there is a lack of information or only fragmented**
- Examples:
 - Emerging and Low Income Countries (Carrière-Swallow, 2013, for GDP in Chile; Assessing real-time inflation in Venezuela and Argentina by Cavallo and Rigobon at MIT « The Billion Prices Project »...)
 - Low frequency information (Annual global GDP, Annual National Accounts in LICs)
 - Lagged information (QNA nowcasts, flash estimates, National Bank of Poland: eCPI)
 - Measuring unobserved variables : Uncertainty (Baker, Bloom, Davis, 2016)
 - ...



EX1: NOWCASTING GDP WITH GOOGLE TRENDS

- Source : Weekly data on Google Trends
- Objective: Nowcasting EA GDP growth on a weekly basis (Ferrara & Simoni, 2017)
- Model: Bridge equation with Survey + IPI, augmented with a Google Factor
- Results:
 1. During the first 3 weeks of the quarter, the model is only based on Google Trends information and provides reasonable forecasting errors (ie: similar to those obtained with official data)
 2. When official data are available, no marginal gain from using Google data

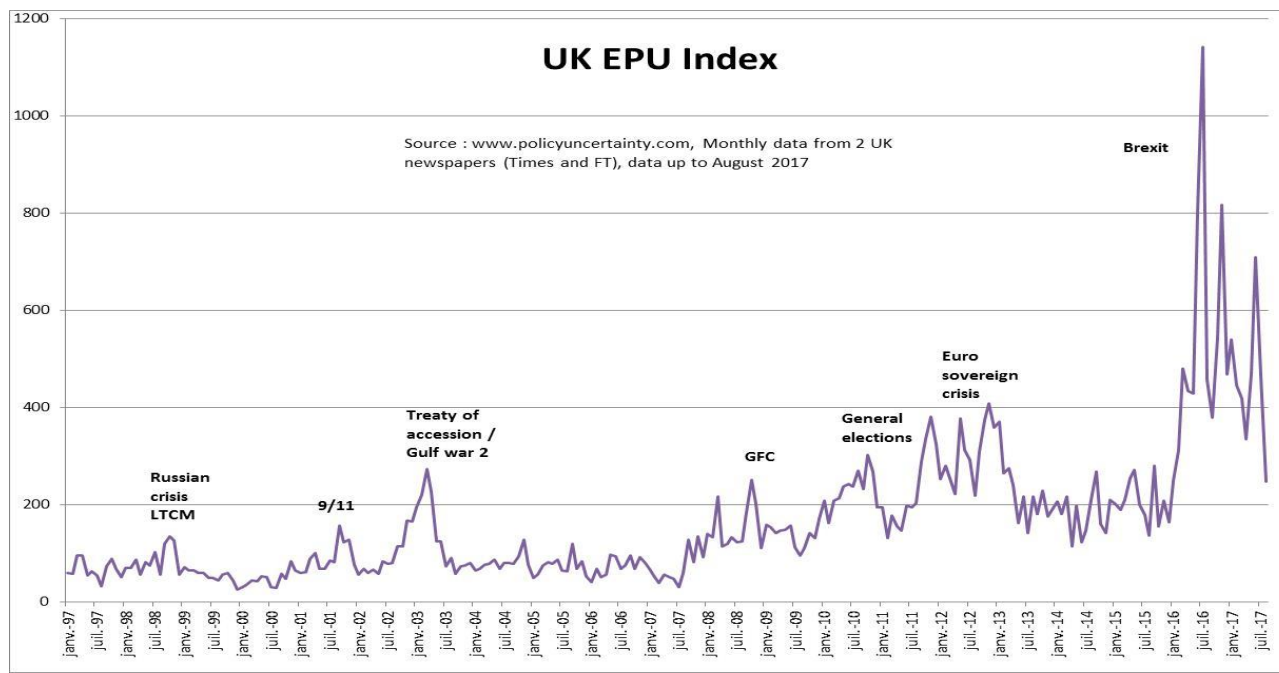


Source:
Ferrara and Simoni
(2017, mimeo)



EX2: MEASURING UNCERTAINTY (BAKER, BLOOM, DAVIS, 2016)

- Source : Media scraping
- Objective: Assess **Economic Policy Uncertainty** in real-time for many countries
- Approach: Counting articles referring to Economy/Uncertainty/Policy
- Results:
 1. EPU indexes for countries (see graph for the UK) and the world economy
 2. Literature is showing that the coincidence of uncertainty and financial shocks can lead to adverse macro outcomes (Caldara et al., 2016, Candelon et al., 2017)





Main take-aways:

- Big data appear as good complement to official statistics for nowcasting and forecasting macro variables
- Big data are useful when there is no, or only fragmented, information on macroeconomic activity
- Research is crucial for developing new sources of data, new tools, new models ... to have a more accurate and rapid assessment of current and future macro evolutions