Discussion of "Macroeconomic nowcasting with big data through the lens of a sparse factor model" by Laurent Ferrara and Anna Simoni

Francesca Monti
Bank of England

ECB, Frankfurt am Main 19 June 2018

Outline of the discussion

- Summary of the paper
  - Methodology
  - Main findings
- My comments
  - Nowcasting: Alchemy or science?
  - The evaluation of the model
    - Real time?
    - More timely official data?

Summary of the paper

- The paper asks the question: Can Google Trends data be useful in macroeconomic nowcasting?
- To assess this $\Rightarrow$ Combine Google data with "official" soft and hard data to assess their relative importance.
  - Summarise the Google data with some factors.
  - Estimate a different *factor-augmented bridge regression* for each week of the quarter with these factors and the data available at that point in time.
  - Pseudo real time evaluation against (some) competing nowcasting models

Methodology

1. Extract factors from Google search data with Sparse Principal Component Analysis (Zou et al. 2006)
   1. Pre-select a subsample of Google data before estimating a factor model (3 pre-selection methods considered).
   2. Start from ordinary principal components analysis, which can be formulated as regression-type optimisation problem.
   3. $\Rightarrow$ Enforce sparsity by adding two terms to the PCA minimisation problem: ℓ_1 type penalty and a Ridge-type quadratic penalty to deal with possible multicollinearity
2. For every week of the quarter w estimate:
   $Y_t = \alpha_{0,w} + \alpha'_{1,w} F_t^w + \alpha_{2,w} S_t + \alpha_{3,w} IP_t + \varepsilon_t$
   Because of frequency mismatch, the twelve models include different numbers of predictors.
3. Compare with 1) bridge regression with ordinary PCA factors and 2) bridge regression with LASSO ℓ_1 type penalty (no factors) and across pre-selection method.

Main findings

- Accuracy improvements with Google trends data in the first three weeks of the quarter, *when the model incorporates no other information*
- The improvement in forecasting accuracy due to Google Trends data is negligible once the first soft data is available.
- Pre-selection of the Google data that go in the factor model helps.

Nowcasting: Alchemy or science?

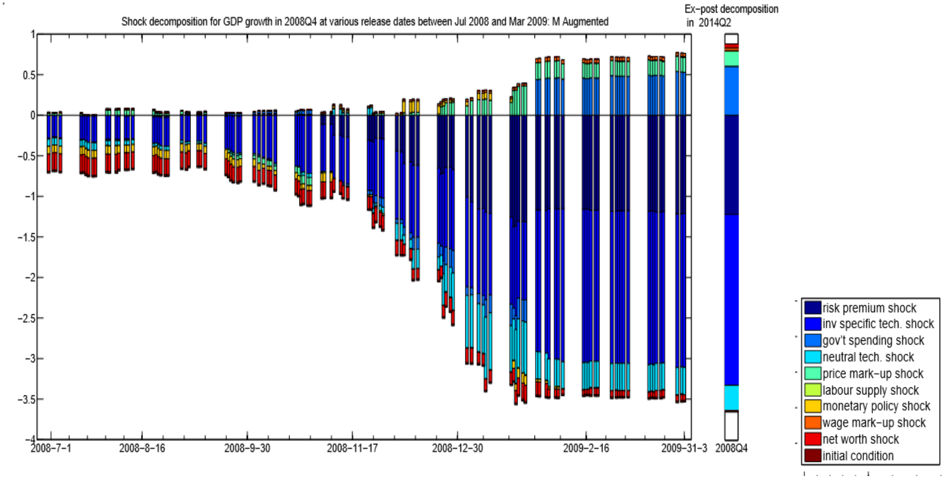
- There are a lot of moving parts in this paper: google trends, pre-selection methods, a factor model, shrinkage, and processing of the news that changes according to the day of the month.
- *"While this approach uses data, it is not scientific in the sense of being replicable, using well-understood methods, quantifying uncertainty, being amenable to later evaluation [...] and being internally inconsistent".* Stock and Watson (2017, JEP)
- But should we give up internal consistency and coherency?
- Stock and Watson (2017, JEP) point to the improvements of the methods for real-time macroeconomic monitoring as one of the 10 key developments in time-series econometrics in the last 20 years.

Is there a trade-off between accuracy and coherency?

The last 20 years have seen the development of platforms for real-time forecasting that combine formal models for big data and Filtering into nowcasting.

- Aruoba, Diebold & Scotti (2009): dynamic factor model blending weekly, monthly and quarterly stock and flow data.
- Banbura et al. (2013) handle in a systematic daily, weekly, monthly and quarterly data, essentially all the information that moves markets.
- Brave, Butters & Justiniano (2016) nowcast with a mixed-frequency BVAR, exploiting Schorfheide and Song (2015) to do exact Bayesian posterior mixed-frequency analysis in a high-dimensional model.
- Giannone, Monti & Reichlin (JME, 2016): nowcasting with a DSGE model!

Interpreting 2008Q4 in real time



Real time model evaluation

Real time model evaluation

1. Recursive estimation
2. Real-time data flow
3. Real-time data
4. Ex-ante model specification

Real time model evaluation

Real time model evaluation

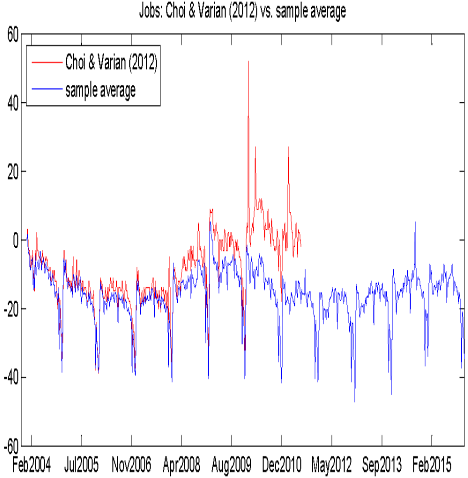
1. **Recursive estimation**
2. **Real-time data flow**
3. Real-time data
4. Ex-ante model specification

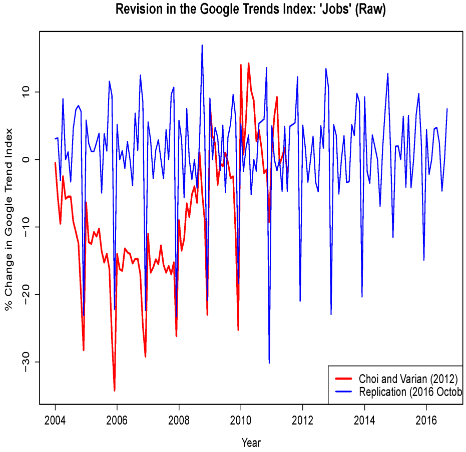
Google trends data

- Product of numerous algorithms and decisions invisible to the user and that are adapted in time (Lazar et al. 2014)
- Data revisions are large, what you observe today is different from what was available at the time the forecasts were made.
  - Noise might be large, need to clean from seasonality and outliers.
  - Unlike official data, which go through systematic quality checks, these series are not audited.

Google trends data

Li (2016) Van Norden (2017) http://econbrowser.com/archives/2017/05/guest-contribution-big-data-and-fake-forecasts





Real time model evaluation

Paper finds that Google trends data are useful when other data are not available.

Real time model evaluation

Paper finds that Google trends data are useful when other data are not available.

But this marginal predictive power of Google data could be an illusion, because the data being exploited are revised.

Real time model evaluation

Paper finds that Google trends data are useful when other data are not available.

But this marginal predictive power of Google data could be an illusion, because the data being exploited are revised.

Real time model evaluation

1. **Recursive estimation**
2. **Real-time data flow**
3. **Real-time data** $\Rightarrow$ Would be great to have vintages!
4. Ex-ante model specification

Other official data?

- Paper finds that Google trends data are useful when other data are not available $\rightarrow$ exploit their **timeliness**.
- Only one survey series is used, and it is not the most timely.
- We know that for the US there is a plethora of high frequency private sector weekly and daily data to be exploited.
  - e.g. Stock and Watson (2014) – Weekly Economic Index

Data from Stock and Watson WEI

Private Sector Weekly & Daily Data

Description of Variables

Series Name	Frequency	Release Date	Transformation
A. Consumer Spending Variables			
Gallup Consumer Spending	Daily	Early afternoon next day	-
ICHS Same-Store Retail Sales	Weekly	Tuesday Following Week	Year-over-Year Change (%)
Johnson Redbook Same-Store Retial Sales	Weekly	Tuesday Following Week	Year-over-Year Change (%)
Composite Sales Index (Redbook + ICHS)	*Weekly*	-	Year-over-Year Change (%)
MBA Purchase Applications	Weekly	Wednesday Following Week	-
B. Consumer Confidence Variables			
Gallup Economic Confidence	Daily	1PM Following Day	-
Rasmussen Consumer Index	Daily	11AM Same Day	-
Bloomberg Consumer Comfort Index	Weekly	Thursday Following Week	
C. Employment Variables			
Gallup Job Creation Index	Daily	1PM Following Day	-
Unemployment Insurance (Initial Claims)	Weekly	Thursday Following Week	-
D. Industrial Production Variables			
Raw Steel Production	Weekly	Monday Following Week	Year-over-Year Change (%)
Lumber Production: Western Woods	Weekly	Thursday Following Week	Year-over-Year Change (%)
Car Production	Weekly	Tuesdsay Following Week	Year-over-Year Change (%)
E. Financial Variables			
Chicago Fed National Financial Conditions Index	Weekly	Thursday Following Week	-
CBOE Market Volatility Index (VIX)	Daily	COB Same Day	-
Moody's BAA-AAA Corp. Bond Spread	Daily	COB Same Day	-
10YR-3Mo. Secondary Treasury Spread	Daily	COB Same Day	-

Data for EU

TRADING ECONOMICS | CALENDAR | INDICATORS | MARKETS | FORECASTS | Search | APPS

Time	Country	Event	Actual	Previous	Consensus	Forecast
07:45 AM	FR	Current Account MAY		€-1.1B		€-1.8B
09:00 AM	IT	Retail Sales MoM MAY		-0.7%		0.4%
09:00 AM	IT	Retail Sales YoY MAY		-4.6%		1.2%
Monday July 09 2018			Actual	Previous	Consensus	Forecast
07:00 AM	DE	Balance of Trade MAY		€20.4B		€21.2B
07:00 AM	DE	Balance of Trade s.a MAY		€19.4B		
07:00 AM	DE	Exports MoM s.a MAY		-0.3%		
07:00 AM	DE	Imports MoM s.a MAY		2.2%		
07:00 AM	DE	Current Account MAY		€22.7B		€14.6B
Tuesday July 10 2018			Actual	Previous	Consensus	Forecast
05:30 AM	NL	Manufacturing Production YoY MAY		5%		5.5%
07:45 AM	FR	Industrial Production MoM MAY		-0.5%		0.8%
09:00 AM	IT	Industrial Production YoY MAY		1.9%		2.9%
09:00 AM	IT	Industrial Production MoM MAY		-1.2%		-0.1%
09:40 AM	ES	12-Month Letras Auction		-0.348%		
09:40 AM	ES	6-Month Letras Auction		-0.468%		
10:00 AM	EA	ZEW Economic Sentiment Index JUL		-12.6		4.5
10:00 AM	DE	ZEW Current Conditions JUL		80.6		82.5
10:00 AM	DE	ZEW Economic Sentiment Index JUL		-16.1		-10.2
Wednesday July 11 2018			Actual	Previous	Consensus	Forecast
10:40 AM	DE	10-Year Bund Auction		0.48%		
10:45 AM	IT	12-Month BOT Auction		0.550%		
	EA	ECB Non-Monetary Policy Meeting				
Thursday July 12 2018			Actual	Previous	Consensus	Forecast

My Conclusion

Google search data are very useful when official data are not available:

- Emerging economies, e.g. Carriere-Sallow & Labbé 2013.
- Lagged data, e.g. Coble & Pincheira (2017) predict US building permit data.
- Measuring unobserved variables: e.g. Uncertainty (Baker, Bloom, Davies, 2016)

This paper makes progress in the evaluation their relevance for nowcasting, but...

- Given the relevance of data revisions, it would be great to have this assessment with actual real-time