Nowcasting GDP with big data from internet

Luca Onorante (ECB). Discussant: Chris Redl (BoE)

London, Big Data & Machine Learning Conference, November 2018

Summary

- Combines raw Google search/category data with Bayesian model averaging
- Predict prelim GDP
- Raw search data $\longrightarrow$ 1000+ series
- Compete with std Macro dataset

Summary

- Robustify with related search terms
- E.g. "GDP" in UK:
  - uk gdp, gdp per capita, gdp growth, world gdp, gdp of uk
- E.g. "GDP" in Germany:
  - gdp per capita, bundespolizei, gdp growth, gdp nrw, polizei

Summary

- Simple way to address noise: Bayesian Model Averaging
  - **exponential reweighting**: shrinkage for noise

$$w_{i,t} \propto \frac{L_i\left(y_t|y^{1:t-1}, M_i\right)^{\alpha}}{\sum_{j=1}^{K} L_j\left(y_t|y^{1:t-1}, M_j\right)^{\alpha}}.$$

  -
  - **trimmed weights**: collinearities in big data

$$w'_{i,t} \propto w_i * I(w_i < Percentile(w, x))$$

  -
- Results: small improvements on AR models

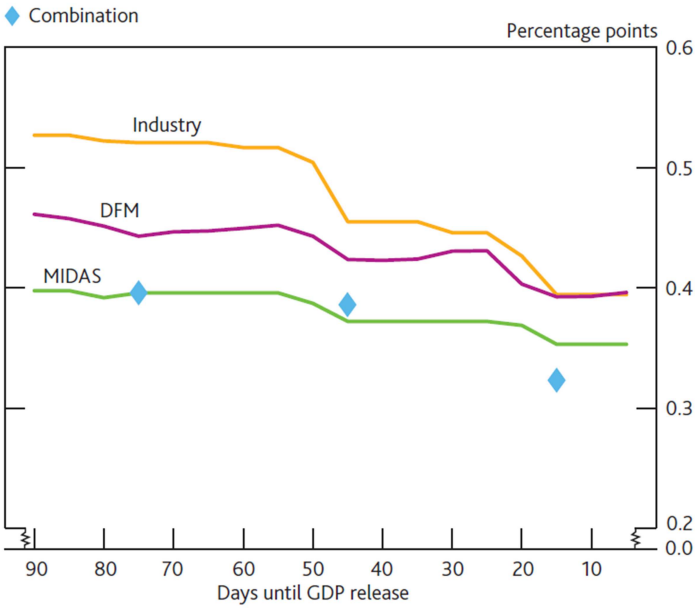
Comments

- Raw Google data $\rightarrow$ automatic feature selection
- Robustify with related search is nice idea!
- What search terms exactly? "growth", "GDP", "jobs"?
- Flavour of post-robustification terms look like? Example would be v. helpful here

Comments

- Hard to know value of the search data without comparison to performance of Nowcasting models i.e. DFM, MIDAS as used at ECB?
  - Plot over relase calendar: surely timliness of the data is its key advantage?

Chart 2 Model RMSEs through the data cycle (2007–16)



- Case study: what kind of terms do a good job - help policy & reseach where hand selection is pursued