

“Google it!” Forecasting the US unemployment rate with a Google job search index

Discussion by Rocco Mosconi

Roma January 19, 2010

1 Summary

- The paper introduces an Internet job-search indicator (the Google Index, GI), and explores its usefulness for forecasting unemployment.
- Out-of-sample forecasting comparison considers 529 models for 5 different transformations of the unemployment rate (the focus is on first differences). 520 of them are ARMAX, and 9 are non linear. Initial Claims are considered as an alternative leading indicator.
- The models augmented with the GI seem to outperform the traditional ones (linear and non linear) in predicting the monthly unemployment rate, whichever transformation is considered. This evidence holds for both the federal and the state-level forecasts.

- The best predicting models, which invariably include the GI, are then compared with the quarterly forecasts provided by the Survey of Professional Forecasters. Also in this case the forecasts based on the GI beat the competitor

2 Data issues

- The authors discuss some weakness of the Google Index. In particular, individuals looking for a job through the internet may well be not randomly selected among job seekers. This lack of representativeness may become much worst if the Google Index based on searches containing "jobs" becomes standard in the analysis of unemployment. In this case, somebody might manipulate it by running a huge

number of automated searches (the same holds true for other google based indexes, like the one recently used for flu, where pharmaceutical companies might be interested in biasing the index).

- Google Index should be adjusted to account for changes of the share of google among web search engines.
- Why using seasonally adjusted data?

3 Model issues

- The unemployment in month t is published on the 5th day of month $t + 1$. Moreover, the Bureau of Labor Statistics revises its initial monthly estimates twice, in the immediately succeeding 2 months, to incorporate additional sample receipts from respondents in the survey and recalculated seasonal adjustment factors. It would be interesting to consider the usefulness of the GI in nowcasting in the current month the official data to be published in two months.
- The focus of the paper is on univariate models, where leading indicators are included as exogenous and forecasted using simple ARMA. I would consider in the forecast race also a multivariate model, like a VAR, for u_t , IC_t and GI_i . Maybe other macro variables might be introduced, at least for the 3 months horizon.

- Unemployment appears to be stationary in the final part of the sample. This would suggest to focus on the levels rather than first differences.
- The predictive power of the GI is somewhat puzzling, since it is not clear why job search anticipates unemployment (should'nt it be coincident?). It would be interesting to introduce some behavioural interpretation for this stylized fact.