Big Data: Google Searches Predict Unemployment in Finland

Joonas Tuhkuri (joonas.tuhkuri@etla.fi)[1]

Keywords: big data, Google, Internet, forecasting, unemployment

1. Introduction

There are over 100 billion searches on Google every month. [1] Could data from Google searches help to predict the unemployment rate in Finland?

Predicting the present and the near future is of interest, as the official records of the state of the economy are published with a delay. Furthermore, the data are subject to revisions, sampling variation, and measurement error. It would be helpful to have more timely information on unemployment, especially during an economic crisis. From a policy perspective, more accurate knowledge could inform better decisions that might help to reduce the unemployment.

Data on Google searches are publicly available in real-time. Real-time information might help to *nowcast* the present unemployment rate, which is uncertain. Furthermore, Google search queries might be associated with the future expectations and thus help to *forecast* the future unemployment.

To answer these questions, this paper compares a simple univariate autoregressive model of unemployment to a model that contains a variable, Google Index, formed from Google data. In addition, cross-correlation analysis and Granger non-causality tests are performed. Furthermore, to study the robustness of the results, I explore the sensitivity of the findings to the selected search terms. The Google Index is constructed from the Google data using approximately 2 million [2] search queries that are related to search for unemployment benefits. The underlying idea is that Google searches in these topics might be related to actual filings for unemployment benefits. Moreover, the Internet plays an important role in the labor market [3–5]. That is why Google searches might be able to offer information especially on the unemployment rate. To be clear, I do not claim any clear causality in this paper. However, Google searches might offer a signal on the future unemployment rate. A new data set could also offer new insights on the unemployment.

Previous literature suggests that Google searches could be useful in predicting the unemployment rate in Germany [6], the United Kingdom [7], the United States [8], and predicting the initial claims for unemployment in the United States [9]. This paper offers an extensive review of the literature on forecasting with Internet search data. In summary, the previous studies on Internet searches hint that the variation in volumes of Internet search terms could reveal intentions or sentiment of the population that uses the Internet. However, the topic is still relatively little studied. Previous results serve as a motivation to study further the possibility to use Google searches for predicting the unemployment rate, in this case in Finland. This is the first paper to use Google data to study the Finnish economy.

[1] The Research Institute of the Finnish Economy and the University of Helsinki

2. Methods

The primary data sources for this paper are *Google Trends* database by *Google Inc.* and the Labor Force Survey published by the *Statistics Finland*. *Google Trends* measures the incidence of specific search terms on Google compared to other searches terms. To my knowledge, this is the first paper that additionally provides statistics on the actual search volumes on Google. Although there have been efforts to use the data on Google searches in economic research, the Google data has not been well documented. This paper fills this gap by offering a careful documentation and a discussion on the Google data.

This paper uses an extended version of the methods initially suggested by Choi and Varian [9] and Goel et al. [10] to find out whether Google searches predict the unemployment rate in Finland. In summary, I construct a one single Google Index that will simultaneously describe the evolution of several search terms that are related to unemployment, such as "unemployment benefits". Figure 1 describes the evolution of Google Index and the unemployment rate 2004–2014. The search terms are selected based on prior knowledge of the labor market. Google Index is available in real time, while the unemployment rate is only available after the end of each month. This gives the Google data a meaningful forecasting lead.

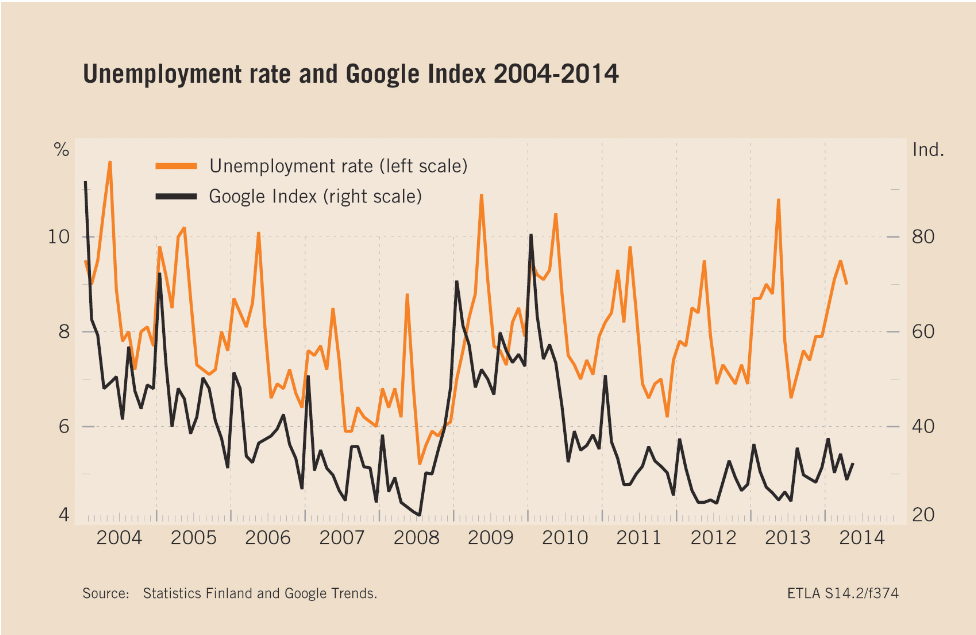

Figure 1. Evolution of Google Index and the unemployment rate in Finland 2004–2014. Source: Statistics Finland and Google Trends.

I apply standard ARIMA model selection procedures and select seasonal AR(1) model as a benchmark for predicting the unemployment rate. Then I add the most recent value of Google Index to the model as an additional predictor. Finally I compare the properties of the two models. In specific, I study how the out-of-sample forecasts improve, measured by mean absolute error using a rolling window forecast. For each prediction I compare dynamic forecasts that contain the most recent information available at the date of prediction. I study particularly the turning points since they are hardest to forecast. Last I run Granger non-causality tests and study the cross-correlation function.

One concern would be that the results were sensitive to the choice of the set of search terms. I explore this sensitivity by estimating the models with different search terms.

3. RESULTS

The results tell that a simple time series model extended with Google data predicts the unemployment rate better than the same model without data on Google search volumes.

Table 1 summarizes the out-of-sample nowcasting and forecasting accuracy of the benchmark (0.0) and the extended models. Compared to a simple benchmark, Google search queries improve the prediction of the present by 10% measured by mean absolute error. Moreover, predictions using search terms perform 39% better over the benchmark for near future unemployment 3 months ahead. The paper also suggests that Google search queries tend to improve the prediction accuracy around turning points.

Table 1. Nowcasting and forecasting accuracy of the benchmark and the extended models.

	Model	MAE	△
t	(0.0) (1.0)	7.8 % 7.0 %	10.0 %
t + 1	(0.0) (1.1)	9.3 % 7.7 %	16.9 %
t + 2	(0.0) (1.2)	10.5 % 7.0 %	32.9 %
t + 3	(0.0) (1.3)	11.1 % 6.7 %	39.2 %
t + 4	(0.0) (1.4)	11.3 % 7.7 %	30.5 %
t + 5	(0.0) (1.5)	11.3 % 8.4 %	25.3 %
t + 6	(0.0) (1.6)	11.4 % 9.0 %	20.5 %

MAE = mean absolute error
△ = improvement in forecasting accuracy

The estimation results of the models support the findings. The coefficient of the present Google Index is statistically significant at 1% level and it has a positive sign, which means that the searches related to unemployment benefits are positively connected to the present unemployment rate. More to the point, the coefficient is 0.0056, which means that 1% increase in current search intensity is associated with 0.5% increase in current unemployment rate.

Extending the benchmark model with *Google Trends* data increases the coefficient of determination by 14.8% and decreases the values of both Akaike and Bayesian information criteria. These findings suggest that the Google searches offer useful information in explaining variation of the unemployment rate.

The correlation coefficient between monthly unemployment and Google Index is 0.87. I observe that the cross-correlation coefficients between present unemployment volumes and past Google searches appear to be larger than the coefficient of the opposite case. In other words, the Google Index presents a classic pattern of a leading indicator. According to the Granger non-causality test, Google searches offer useful information in predicting the unemployment rate. In contrast, the lagged values of unemployment rate do not seem to offer useful information in predicting the Google searches.

The results indicate that Google searches might offer genuinely new information on the unemployment that is not already included in the unemployment series itself. Robustness checks suggest that the results are not sensitive to the selected search terms.

4. CONCLUSIONS

I have found that a simple seasonal first-order autoregressive model, which includes relevant Google variables, tends to outperform models that exclude these predictors in predicting the unemployment rate in the short run. Moreover, joint analysis of the series suggests that the changes in Google searches, which are related to unemployment benefits, more often than not precede the changes the unemployment rate. The results are in line with the previous findings on Google searches and the unemployment [6–9].

As a result, ETLA, the Research Institute of the Finnish Economy has launched a trial for a real-time forecast model ETLAnow that automatically predicts the unemployment rate for three months ahead using data from Google Trends and Statistics Finland, publishing the results every morning. Currently we are building a model that would predict the unemployment rate for each EU-28 country in real-time using big data.

The results suggest that Google searches could offer useful information for predicting the Finnish unemployment rate. The results also demonstrate that big data can be utilized to forecast official statistics.

REFERENCES

[1] Google Internal Data, (2014).

[2] Google Adwords, (2014).

[3] B. Stevenson, The Internet and Job Search, NBER Working Paper 13886 (2008).

[4] P. Kuhn and H. Mansour, Is Internet Job Search Still Ineffective? The Economic Journal 124(581) (2014), 1213–1233.

[5] K. Kroft and D. G. Pope, Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist, Journal of Labor Economics 32(2) (2014), 259–303.

[6] N. Askitas and K. F. Zimmermann, Google Econometrics and Unemployment Forecasting, Applied Economics Quarterly 55(2) (2009), 107–120.

[7] N. McLaren and R. Shanbhogue, Using internet search data as economic indicators, Bank of England Quarterly Bulletin 1 (2011) 134–140.

[8] F. D'Amuri and J. Marcucci, The Predictive Power of Google Searches in Forecasting Unemployment, Bank of Italy Working Paper 891, (2012).

[9] H. Choi and H. R. Varian, Predicting the Present with Google Trends, Economic Record 88(s1) (2012), 2–9.

[10] S. Goel, J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts, Predicting consumer behavior with Web search, Proceedings of the National Academy of Sciences of the United States of America 107(41) (2010), 17486–90.