ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY

ETLA Raportit
ETLA Reports
25 May 2016

No **54**

ETLAnow: A Model for Forecasting with Big Data

Forecasting Unemployment with Google Searches in Europe

Joonas Tuhkuri*

* ETLA – The Research Institute of the Finnish Economy, joonas.tuhkuri@etla.fi

Suggested citation: Tuhkuri, Joonas (25.5.2016). "ETLAnow: A Model for Forecasting with Big Data – Forecasting Unemployment with Google Searches in Europe".
ETLA Reports No 54. http://pub.etla.fi/ETLA-Raportit-Reports-54.pdf

Acknowledgements: ETLAnow is collaborative work of 28 European economic research institutions. We would like to thank our ETLAnow team and partners for their generous contributions. Our partners in the ETLAnow project are listed below.

Thomas Horwath and Team	The Austrian Institute of Economic Research WIFO
Bart De Ketelbutter	Bureau federal du Plan FPB
Iskra Beleva	Economic Research Institute at the Bulgarian Academy Of Sciences ERI-BAS
Iva Tomic and Ivan Zilic	Ekonomski Institut Zagreb EIZG
Alexandros Polycarpou and Nicoletta Pashourtidou	The Economics Research Centre of the University of Cyprus CypERC
Barbara Pertold-Gebicka	Charles University
Jesper Linaa	Danish Economic Council DEC
Andres Võrk	Praxis, Center for Policy Studies
Petri Rouvinen	The Research Institute of the Finnish Economy ETLA
Herve Peleraux	Observatoire Français des Conjonctures Economiques OFCE
Dominik Groll	Institut für Weltwirtschaft an der Universität Kiel IFW
Nikolaos Kanellopoulos and Ioannis Cholezas	Center of Planning and Economic Research KEPE
Tamás Székács	Kopint-Tárki Institute for Economic Research Co.
Stefania Tomasini	Associazione Prometeia AP
Uldis Rozevskis	University of Latvia
Tomas Šiaudvytis and Jurgita Pesliakaite	The Bank of Lithuania and Vilnius University at CEFER
Brian Micallef	Central Bank of Malta
Joris de Wind	Netherlands Bureau for Economic Policy Analysis
Pieter Vlag	Statistics Netherlands
Janusz Chojna	Institute for Market, Consumption and Business Cycles Research IBRKK
Pedro Martins	Queen Mary, University of London and CEG-IST, Lisbon
Marta C Lopez	Nova School of Business and Economics
Oana Popovici and Dana Tapu	Institute for Economic Forecasting, Romanian Academy
Marek Radvansky	Institute of Economic Research, Slovak Academy of Sciences
Bojan Ivanc and Darja Mocnik	Chamber of Commerce and Industry of Slovenia SKEP
Julián Pérez García and Juan Jose Mendez	Centro de Prediccion Economica CEPREDE
Åsa Olli Segendorf and Karine Raoufinia	Konjunkturinstitutet
Simon Kirby	The National Institute of Economic and Social Research NIESR

In addition, we would like to thank Kimmo Aaltonen, Harri Laine, Matti Lammi, Petteri Larjos, Juri Mattila, and Petri Rouvinen at ETLA for contributing to the project.

This document describing the use of Google searches in unemployment forecasts in the ETLAnow project draws from Tuhkuri (2014, 2015, and 2016) and some text is repeated from those sources. ETLAnow project is not described those papers or in earlier documents.

ISSN-L 2323-2447
ISSN 2323-2447 (print)
ISSN 2323-2455 (pdf)

Table of Contents

ETLAnow: A Model for Forecasting with Big Data – Forecasting Unemployment with Google Searches in Europe

Abstract

In this report we document the ETLAnow project. ETLAnow is a model for forecasting with big data. At the moment, it predicts the unemployment rate in the EU-28 countries using Google search data. This document is subject to updates as the ETLAnow project advances.

Key words: Big Data, Google, Internet, Nowcasting, Forecasting, Unemployment, Europe

JEL: C22, C53, C55, C82, E27

ETLAnow: A Model for Forecasting with Big Data – Forecasting Unemployment with Google Searches in Europe

Tiivistelmä

Tämä raportti esittelee ETLAnow-projektia. ETLAnow on suuria tietomassoja hyödyntävä talousennuste. Tällä hetkellä se ennustaa työttömyysastetta kaikissa EU-28 maissa hyödyntäen Googlen hakuaineistoja.

Asiasanat: Big Data, Google, Internet, Ennustaminen, Talousennusteet, Työttömyys, Eurooppa

JEL: C22, C53, C55, C82, E27

1 Introduction

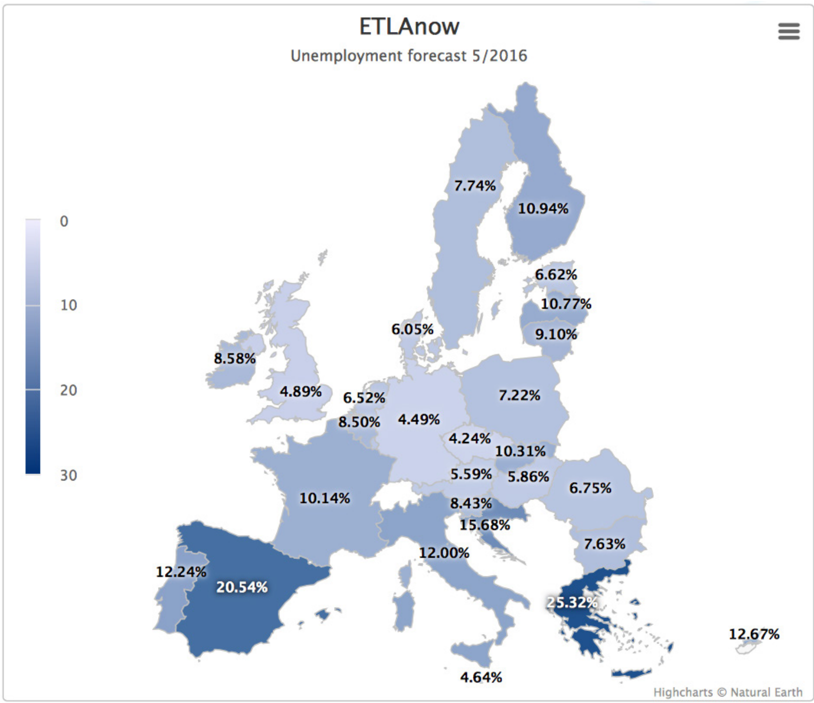
ETLAnow is an experiment run by ETLA, The Research Institute of the Finnish Economy, to use big data in economic forecasting. At the moment, ETLAnow utilizes Google search data to predict the official unemployment rate the EU-28 countries. The model is publicly available at the ETLAnow's website at http://www.etlanow.eu. To our knowledge, ETLAnow is the first publicly available economic forecast that uses Google search data.

This paper provides an overlook on how the model works. In short, we use trends in Google search volumes to predict the unemployment rate. The ETLAnow model is based on the idea that volumes of Google searches on unemployment related matters, such as unemployment benefits or jobs, could be associated with the current and future unemployment rate.

The motivation for our forecasting approach is that newly available real-time and large-scale data sources—such as Google search data—could help produce more accurate economic forecasts. These data are available earlier than official statistics. Moreover, the new data could give an early signal on the behaviour of people and firms. The forecasts, in turn, for example on the unemployment rate, would inform better labor market and monetary policy, and help real people—especially during an economic crisis.

Our earlier and first trial for real-time forecasting is documented in Tuhkuri (2014, 2015). A more detailed analysis on unemployment forecasting using Google data is provided in Tuhkuri (2016) using U.S. state-level data.

Figure 1 ETLAnow forecasts visualized on a map of Europe



In practice, the ETLAnow model automatically predicts the unemployment rate for three months ahead using data from Google Trends database and Eurostat, and publishes the updated forecasts every morning. At present, the model relies on real-time data on the volumes of unemployment-related Google searches and the latest official figures on the unemployment rate. It also features an automated *Twitter* feed that interested users can subscribe to in order to follow ETLAnow's forecasts in real time. Figure 1, a screenshot from the ETLAnow website, visualizes the forecasts on a map of Europe.

Previous literature has shown that Internet search query data could help predict, for example, influenza epidemics (Ginsberg et al. 2009), video game sales (Goel et al. 2010), and housing market transactions (Wu and Brynjolfsson 2015). In summary, studies on Internet searches suggest that the variation in the volumes of Internet searches could reveal intentions or sentiment of the population that uses the Internet. From an economic perspective, each Internet search is someone expressing an interest in or demand for something (Brynjolfsson 2012). We use that information to forecast the economy.

More closely related to our project, our own work and the previous literature show that Google search volumes could help predict the unemployment rate[1] (See, for example, Askitas and Zimmermann 2009; Choi and Varian 2012; and Tuhkuri 2014; 2016).

Our previous findings from Finland (Tuhkuri 2014) tell that, compared to a simple benchmark, Google search queries improved the prediction of the present by 10 % measured by mean absolute error. Moreover, predictions using search terms performed 39 % better over the benchmark for near future unemployment 3 months ahead. In particular, we found that Google search queries tended to improve the prediction accuracy around turning points. Those are often hard to predict. We find that real-time information from Google searches tends to be useful for forecasting purposes during the economic crisis. More generally, in Tuhkuri (2014), we concluded that Google searches would contain useful information on the present and the near future unemployment rate.

On the other hand, using more granular U.S. state-level data, we find in Tuhkuri (2016) that predictive power of Google searches tend to be limited to short-term predictions, and the improvements in forecasting accuracy are sometimes only modest. This is more in line with previous literature on the topic, such as, Choi and Varian (2012). In general, our two studies illustrate both the potentials and limitations of using big data to predict economic indicators.

One of the motivations to use timely data, such as Google data, is that the traditional statistics are released with a lag. In that sense, the ETLAnow project is closely related to the more general and rapidly expanding literature on macroeconomic monitoring and real-time data analysis (see, Croushore 2006; Aruoba and Diebold 2010; Bańbura et al. 2013, and the references therein). Real-time assessment of current macroeconomic activity is also called *nowcasting* (Giannone et al. 2008). The underlying idea is that real-time data could help to nowcast the current level of an economic indicator.

[1] The studies on unemployment forecasting with Google searches have been performed in Germany (Askitas and Zimmermann 2009), the U.S. (Choi and Varian 2012; D'Amuri and Marcucci 2012, Tuhkuri 2016), the UK (McLaren and Shanbhogue 2011), Israel (Suhoy 2009), Finland (Tuhkuri 2014), Italy (D'Amuri 2009), Norway (Anvik and Gjelstad 2010), Turkey (Chadwick and Sengul 2012), France (Fondeur and Karamé 2013), Spain (Vicente et al. 2015), Czech Republic, Hungary, Poland, and Slovakia (Pavlicek and Kristoufek 2014).

But real-time data sources could also have practical relevance for several economic agents. For example, central banks are interested in acquiring real-time information on the economy, and recently, several central banks have shown interest in using Internet search data for economic forecasting (see, for example, Suhoy 2009 and McLaren and Shanbhogue 2011). Several other government institutions and NGOs worldwide, such as national unemployment offices, would also be better equipped if they had more timely information on the unemployment rate.

Recent studies document that the Internet plays an important role in the labor market (see, for example, Kuhn and Skuterud 2004; Stevenson 2008; Kroft and Pope 2014; and Kuhn and Mansour 2014). The Internet is used to search for jobs in a variety of ways, including contacting public employment agencies and submitting job applications (Kuhn and Mansour 2014). In particular, Google searches could offer information on the unemployment rate and labor market activity (Baker and Fradkin 2014). But there are also other promising applications of using Internet data for economic forecasting.

Our forecasting approach builds upon improvements in economic measurement. In this case, we get information on private actions on labor market through Internet search logs. These new data sources are sometimes called *big data*. It is a broad term that refers to new massive data sets—the amount of information created until 2003 is now created every two days (Einav and Levin 2013, and the references therein). The broad theme of the ETLAnow project is to understand whether big data could improve macroeconomic forecasts.

2 The Model

2.1 Data

The primary data sources for ETLAnow forecast model are the *Google Trends* database developed and maintained by *Google Inc.* and the Labor Force Statistics from *Eurostat*.

Unemployment

ETLAnow uses harmonized and non-seasonally adjusted unemployment rates published by Eurostat, as we are interested in short-term predictions. Unemployment statistics are available with at least a one-month lag.

Recent evolution of the unemployment rate in most EU-28 countries was characterized by a sudden increase in the level of unemployment rate between 2008 and 2010. It was associated with the economic crisis. The abrupt increase in unemployment was hard to predict—or at least, many predictions failed. New big data sources, such as Internet search data, could help produce more accurate forecasts.

Google

The Google search data for the ETLAnow model comes from the *Google Trends* database through a special API that was built for that purpose. Google data are available in real time.

Google Trends tells us how many searches on certain search terms have been made, compared to the total number of Google search queries in the same period. The data are publicly available from 2004 onwards. The data are location specific; we use the data at the EU member country level. Google data is documented in more detail in Tuhkuri (2016) and Choi and Varian (2012).

In most EU member countries, more than 90 percent of internet users use Google.[2] And according to Eurostat, Internet use varies between 50 to almost 100 percent in the EU. From another perspective, economic literature provides support for using Internet data for labor research; the Internet is commonly used as a tool in the labor market (Kuhn and Mansour 2014). For example, according to Kuhn and Mansour (2014), the proportion of young unemployed in the US who looked for work online was 74% in 2009.

In order to use Google search data, we needed to select the keywords we would use in each EU-28 country. In short, we use Google search terms that specifically an unemployed person, or a person expecting unemployment, would search for in each country. The underlying idea is that more searches would give a signal of a higher unemployment.

But each country is different. People make searches in their local languages and the content of searches depends on the institutional context that country, and many other factors. That is, the most useful set of search terms for prediction is likely to be different from one country to another. In order to solve this issue, we utilized expert knowledge from local labor economists in each EU country in order to define specific Google search terms that we would use to make predictions. As a result, we use search terms in 22 languages.

Most of the search terms we that we use are related to being laid-off or seeking for new employment. According to our research described in Tuhkuri (2014, 2015, and 2016) and to earlier literature (Choi and Varian 2012, Askitas and Zimmermann 2009), being laid-off tends to result in searches for unemployment benefits, new jobs or simply searches for being laid off. People search for unemployment benefits in various ways, for example, using different names for the benefits, name of the organization that distributes the benefits, or name of the benefit system. Seeking for new employment includes searches using terms related to jobs, new jobs, employment websites or recruitment agencies.

We only use search terms that have a solid theoretical or institutional background in labor market. This is to avoid using search terms that might have been good predictors in the past only by chance. Those terms would not necessarily produce reliable predictions in the future. More to the point, to facilitate our expert's work, we encouraged them to think through the Internet search; what would be most reasonable search queries? How would an unemployed person proceed over the Internet after being laid off in their country?

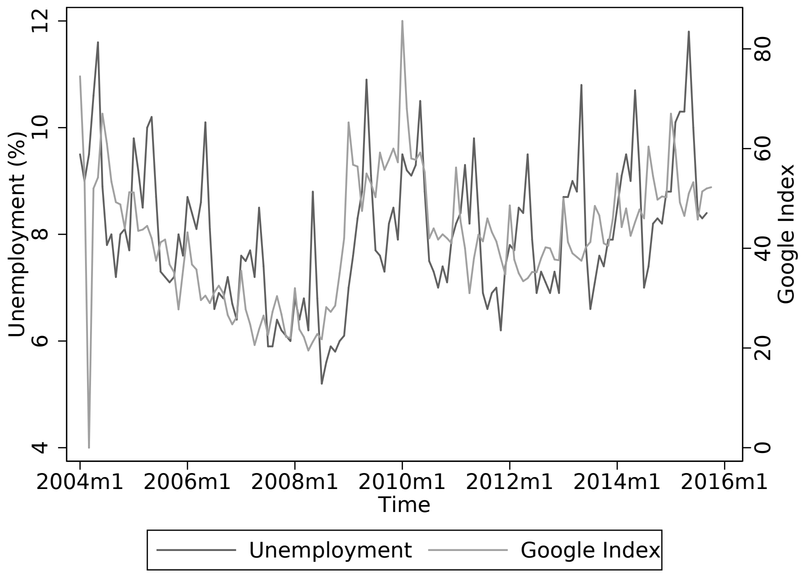
Our experience and previous studies (see, for example, Tuhkuri 2016; McLaren and Shanbhogue 2011) suggest that different wordings and spellings of the same concept are useful. For example, "unemployment benefits" and "labor market subsidy" might both be useful terms, as might "UI benefits" and "labour market subsidy" be as well. We have also noticed that short terms are usually better than long. In each country, we have included many terms in order to extract a more robust signal.

2 Source: PTG Media, 2011.

After selecting the set of search terms, we follow the method proposed by Tuhkuri (2016) in order to construct a variable—we call it *Google Index*—for each country from the Google data. Google Index represents aggregate search activity for the selected unemployment-related search queries. It is normalized between 0 and 100.

Figure 2 gives an example of the resulting data set for an individual country. The figure describes the evolution of the Google Index and the unemployment rate in Finland from January 2004 until October 2015. The series seem to behave in a similar manner. However, association is not as clear in every country covered by our ETLAnow forecast model. This depends on the selected search terms, and on how the Internet is used in those countries. We expect to improve the Google indices as we learn more about Internet behavior in each country. Figures describing the evolution of the Google Index and the unemployment rate in every EU country are given in the Appendix.

Figure 2 Unemployment rate and the Google Index that describes search activity for unemployment benefits in Finland 2004–2016



Sources: Eurostat and Google Trends.

2.2 Methods

ETLAnow model is an autoregressive seasonally adjusted time-series model extended with Google data. It uses the past unemployment rate and a real-time variable constructed from the Google search volumes in order to predict the unemployment rate. ETLAnow model's schematic structure is given in Figure 3.

Figure 3 ETLAnow model's schematic structure

The mathematical exposition is given in the equation below. ETLAnow uses a seasonal AR(1) model with an exogenous Google variable.

$$log(y_t) = \beta_{00} + \beta_{10}log(y_{t-1}) + \beta_{20}log(y_{t-12}) + \beta_{30}x_t + e_t$$

The unemployment rate in the present month t is denoted by y_t, in the previous month by y_{t-1}, and a year ago by y_{t-12}. The contemporaneous value of the Google Index is denoted by x_t. Moreover, e_t stands for the error term. Coefficients and the constant term are denoted by β:s using different subscripts. The described model is also used in Tuhkuri (2015, 2016) and is closely related to the work of Choi and Varian (2012) and Goel et al. (2010).

Google data are available a month earlier than the official unemployment statistics. That gives the Google data a meaningful forecasting lead (Choi and Varian 2012).

Each forecast horizon has its own model. That is, we construct separate models for each forecast horizon into the future, so that every model uses the most recent information when producing dynamic forecasts for the future. But the idea in each forecast model is the same, only the time subscripts change depending on the most recent data that is available for that horizon. Optimal forecasts are produced recursively.

The selected model in the ETLAnow project is a starting point. Empirical research has shown that simple models often yield better out-of-sample predictions than complex models (Mahmoud 1984). That is why a simple univariate autoregressive model is a relevant benchmark in our forecasting environment. More to the point, Montgomery et al. (1998) document that an autoregressive model is appropriate for short-term unemployment forecasting. But in future work, we might be able to improve the predictions by using more sophisticated forecasting techniques. Tuhkuri (2016) provides a discussion on model selection in the context of forecasting unemployment with Google searches.

In the ETLAnow forecast model, both variables, the unemployment rate and the Google Index, are measured in levels rather than in differenced values, because both are bounded between 0 and 100. For this reason, they cannot exhibit global unit root behavior (Koop and Potter 1999). Furthermore, during the last one hundred years, the unemployment rates have had no visible trend and economic theory does not suggest they should have had one (see, for example, Cochrane 1991).

A seasonal autoregressive term, y_{t-12}, is included in the ETLAnow's AR model to accommodate some of the seasonality in the unemployment series. In the literature on assessing the relevance of Internet data sources, Choi and Varian (2012) and Wu and Brynjolfsson (2015) apply the same approach. Additionally, we perform a logarithmic transformation on the unemployment series since changes in unemployment rate are most naturally discussed in percentage terms and also because logarithmic transformation helps stabilize the variance of the series (Lütkepohl and Xu 2012).

Tuhkuri (2016), together with related literature (see, for example, Choi and Varian 2012; Askitas and Zimmermann 2009), provides an assessment on how far into the future, when, and how much Google searches could improve unemployment forecasts. Tuhkuri (2016) also provides a forecast comparison, comparing simple models that include Google variables to those models that do not. This comparison includes the models that we use in ETLAnow. In future, we plan to provide an overall analysis of the ETLAnow model's forecast performance.

3 The user interface

ETLAnow provides forecasts for all 28 EU countries and computes the aggregate EU-28 average. The ETLAnow forecasts are given in tables, such as, the one depicted below. In the first two rows of Table 1, the model reports the most recent official unemployment statistics in the EU from the *European Union Labor Force Survey* and the ETLAnow forecasts for the next three months, this month, and the past three months. For example, in May, the model predicts the unemployment rate until August, while the official statistics are from March. But Google data are available in real-time.

Table 1 An example of ETLAnow forecasts

EU-28

Unemployment rate (%)	2/2016	3/2016	4/2016	5/2016	6/2016	7/2016	8/2016
Official	9.30	9.10	.	.	.	.	.
ETLAnow	9.21	9.28	9.12	**9.08**	9.08	8.99	8.88
Change (pp)			-0.58	-0.42	-0.22	-0.01	-0.02
							Date: 24.5.2016

. = official data not available
Last update 24.5.2016.
Next update 25.5.2016.
Next official release on the unemployment rate 31.05.2016.
Export ETLAnow forecasts for this table or all ETLAnow forecasts.

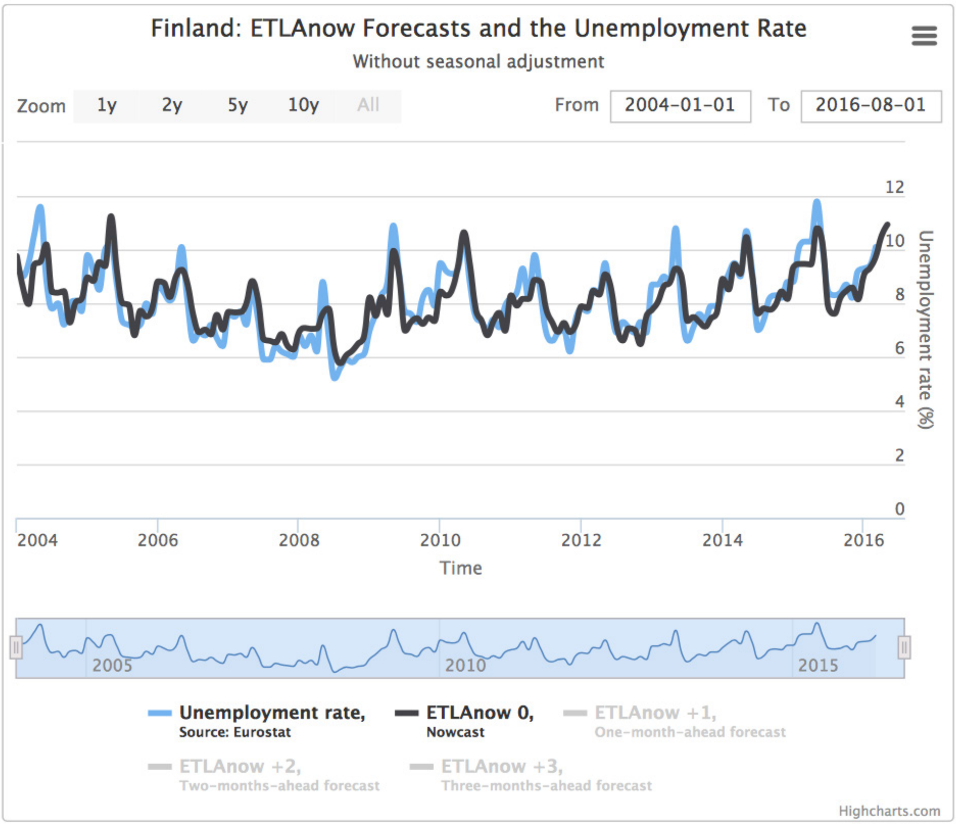
The forecasts for the past months are reported—although it may sound strange—because the official records on the state of the economy are published with a delay. In other words, we predict the past, present and future. Each forecast becomes more accurate toward end of the month as we gather more information on the Internet search activity. Reported historical forecasts are those released on the last day before official statistics were released.

Last row of each table compares the ETLAnow forecasts to the official unemployment rate one year ago. This comparison tells whether the ETLAnow model predicts rising or falling unemployment rate. For example, +0.1 in the last row indicates that the unemployment rate is expected to be 0.1 percentage points higher than in the corresponding month a year ago. Similarly, -0.1 indicates that the unemployment rate is expected to fall 0.1 percentage points as compared to the rate a year ago.

A user can export the current and past forecasts from links provided below each table. Furthermore, simulated historical data for the forecasts will estimate what predictions ETLAnow would have done since 2004, if it were in use. But, ultimately, we can evaluate the accuracy of ETLAnow forecasts every month when new official data become available. Then we can compare the forecast to the actual unemployment figures.

ETLAnow also visualizes the forecasts on an interactive time series graph, depicted in Figure 4.

Figure 4 ETLAnow forecasts visualized in an interactive time-series graph



ETLAnow also provides a portal through which the user can explore and modify the set of Google search terms that were used in forecasting. As a reference for our approach, Brynjolfsson et al. (2014) present a crowd-sourcing based variable selection method. They find that it improves unemployment predictions when using Google search data. Human interaction with the model might help identify when the language and search behavior are changing.

4 The model performance

In this section, we provide an analysis of the data underlying the ETLAnow model. We explore whether Google searches include useful information on the unemployment rate in the EU and could help improve unemployment forecasts. This analysis reflects the current data and is subject to updates as we develop ETLAnow.

4.1 Cross correlation

Table 2 displays the values of the estimated cross-correlation function between unemployment-related Google searches and the unemployment rate. We find strong contemporaneous correlations between Google searches and the unemployment rate, presented in the column labeled by zero. Furthermore, in many countries, the values of the cross-correlation function between past Google search volumes and the present unemployment rate appear to be larger than that of the opposite case. In those cases, Google searches *now* are better predictors of the *future* unemployment rate than they are of the present. That is, in many countries of our sample, Google searches *anticipate* the unemployment rate.

Table 2 Cross-correlation function between the unemployment rate and the Google Index

	CCF								
h	-4	-3	-2	-1	0	1	2	3	4
AT	0.17	0.16	0.19	0.28	0.32	0.26	0.27	0.23	0.25
BE	-0.30	-0.38	-0.37	-0.28	-0.17	-0.11	-0.27	-0.50	-0.52
BG	0.12	0.10	0.09	0.06	0.04	0.04	0.03	0.02	0.03
CY	0.47	0.45	0.44	0.41	0.40	0.37	0.36	0.34	0.32
CZ	0.34	0.31	0.30	0.34	0.40	0.39	0.38	0.34	0.27
DE	-0.49	-0.52	-0.53	-0.49	-0.42	-0.43	-0.46	-0.50	-0.51
DK	0.69	0.68	0.68	0.68	0.66	0.63	0.58	0.55	0.53
EE	0.37	0.38	0.38	0.39	0.39	0.40	0.40	0.41	0.40
FI	0.64	0.55	0.45	0.42	0.45	0.36	0.32	0.38	0.31
FR	0.72	0.73	0.74	0.72	0.70	0.63	0.57	0.54	0.52
GR	0.51	0.53	0.54	0.55	0.58	0.59	0.58	0.59	0.59
HR	-0.41	-0.43	-0.45	-0.48	-0.52	-0.58	-0.63	-0.66	-0.68
HU	0.72	0.74	0.75	0.76	0.77	0.74	0.71	0.68	0.66
IE	0.83	0.82	0.81	0.81	0.79	0.76	0.73	0.70	0.67
IT	-0.08	-0.10	-0.03	0.05	0.14	0.01	-0.01	0.02	0.10
LT	0.23	0.19	0.15	0.12	0.08	0.05	0.02	-0.01	-0.05
LU	0.59	0.59	0.58	0.57	0.58	0.60	0.60	0.61	0.62
LV	0.64	0.64	0.63	0.62	0.60	0.58	0.56	0.54	0.51
NL	0.75	0.73	0.72	0.71	0.72	0.64	0.60	0.57	0.56
PL	0.89	0.90	0.91	0.92	0.92	0.91	0.87	0.87	0.85
SE	0.56	0.57	0.56	0.52	0.52	0.35	0.45	0.40	0.36
SI	0.66	0.62	0.61	0.60	0.58	055	0.51	0.48	0.45
SK	0.30	0.30	0.31	0.34	0.37	0.34	0.30	0.27	0.26
UK	0.10	0.06	0.05	0.08	0.08	0.02	-0.05	-0.10	-0.11

n = 141, h = lag of Google Index, CCF = value of cross-correlation function. The values of CCF on the left-hand side tell the correlation coefficients between past Google search volumes and the present unemployment.

4.2 Granger causality

Table 3 shows statistics for testing Granger non-causality (Granger 1969). The first reported specification in the table is a standard Granger non-causality test based on first-order VAR model. In the second specification, we use a lead of the Google variable, because it is observed at least a month before the unemployment rate (see, Tuhkuri 2016 for more details). In most countries, the null hypothesis that Google searches do not Granger cause unemployment can be rejected at the 1% or 5% level. We also observe that the unemployment rate alone in many cases do not offer useful information in predicting the Google search volumes. This result suggests that Google searches could offer new and useful information on the unemployment rate.

Table 3 Statistics for testing Granger non-causality

	Null hypothesis							
	VAR(1)				VAR(1) using lead of x			
	$y \nrightarrow x$		$x \nrightarrow y$		$y \nrightarrow x$		$x \nrightarrow y$	
Country	χ^2	p-value	χ^2	p-value	χ^2	p-value	χ^2	p-value
AT	<0.001	0.99	0.65	0.42	1.78	0.18	3.56	0.059
BE	0.20	0.66	11.3	0.001***	2.20	0.14	10.6	0.001***
BG	0.089	0.77	4.02	0.045*	0.049	0.82	3.93	0.047*
CY	1.81	0.18	5.06	0.024*	1.97	0.16	4.61	0.032*
CZ	9.18	0.002**	0.32	0.57	9.85	0.002**	1.98	0.16
DE	9.25	0.002**	25.0	<0.001***	11.4	0.001**	0.11	0.74
DK	0.94	0.33	6.86	0.009**	0.20	0.66	15.2	<0.001**
EE	1.50	0.22	0.12	0.73	0.94	0.33	0.014	0.91
FI	0.78	0.38	4.12	0.043*	1.31	0.25	11.2	0.001**
FR	11.1	0.001**	11.6	0.001**	7.17	0.007**	26.8	0.000**
GR	15.5	<0.001**	0.80	0.37	15.9	<0.001**	0.41	0.52
HR	14.0	<0.001**	5.75	0.016**	16.5	<0.001**	15.8	<0.001**
HU	11.9	0.001**	1.60	0.21	8.77	0.003**	5.49	0.019*
IE	0.37	0.54	33.0	<0.001**	0.20	0.65	42.7	<0.001**
IT	0.30	0.58	11.2	0.001**	0.01	0.93	35.4	<0.001**
LT	0.011	0.92	17.8	<0.001**	0.08	0.76	10.49	0.001**
LU	15.2	<0.001**	1.10	0.29	13.7	<0.001**	.082	0.78
LV	7.30	0.007**	9.77	0.002**	5.53	0.019*	8.01	0.005**
NL	0.19	0.67	2.44	0.12	1.23	0.27	58.9	<0.001**
PL	9.35	0.002**	15.7	<0.001**	4.87	0.027*	26.7	<0.001**
SE	0.03	0.86	9.37	0.002**	14.0	<0.001**	28.8	<0.001
SI	3.62	0.057	4.33	0.037*	1.99	0.158	10.13	0.001**
SK	3.83	0.05*	0.41	0.52	2.73	0.098	4.58	0.032**
UK	0.01	0.92	0.015	0.900	0.52	0.468	40.50	<0.001**

y = unemployment rate, x = Google Index.

The sample period is Jan 2004–Sept 2015 (n = 141). Both models estimated are first-order VARs, which, based on the Schwarz criterion, are statistically adequate simplifications of second-order VARs. Asterisks * and ** denote significance at the 5% and % levels, i.e., Granger non-causality $' \nrightarrow '$ is rejected.

4.3 Panel data

We estimate a country-level fixed-effect model using the panel aspect of the data. We use the following fixed effects model with lagged dependent variables.

$$log(y_{i,t}) = \beta_1 log(y_{i,t-1}) + \beta_2 log(y_{i,t-12}) + \beta_3 x_{i,t} + \alpha_i + e_{t,i},$$

As in the previous equation describing the ETLAnow model, the unemployment rate is denoted by $y_{i,t}$ and the Google Index by $x_{i,t}$. Each EU member state is denoted by i. The fixed effects model has 28 different intercepts denoted by α:s, one for each state. We predict the unemployment rate using past unemployment rates and present Google searches. And using the model, we control for constant country-level differences.

The results of the model are given in Table 4. In summary, the Google Index's coefficient is significant at the 1% level. More generally, the estimation results suggest that the Google searches are associated with the unemployment rate in the EU—even after controlling for the country-level, lagged, and seasonal effects.

Table 4 Estimation results of the fixed effects model

Model		FE (AB)	FE (OLS)
Variables			
$log(y_{t-1})$		0.959**	0.958**
		(0.00580)	(0.00545)
$log(y_{t-12})$		0.00889	0.00891*
		(0.00550)	(0.00515)
x_t		0.000745**	0.000763**
		(0.000072)	(0.000068)
Summary statistics for FE (OLS)			
R^2	within	0.963	
	between	0.997	
	overall	0.980	
F test that state fixed effects = 0		4.67	
		(<0.0001)	

y = unemployment rate, x = Google Index.

Asterisks * and ** denote statistical significance at 5% and 1% levels using a two-sided test with standard errors of Arellano (1987). In the first column, the model is estimated by method of Arellano and Bond (1991). In the second column, the model is estimated by the ordinary least squares (OLS) method. The sampling period is Jan 2004–Sept 2015.

References

Anvik, C., & Gjelstad, K. (2010). "Just Google it". Forecasting Norwegian unemployment figures with web queries. *CREAM Publication 11*.

Arellano, M. (1987). Computing Robust Standard Errors for Within-groups Estimators. *Oxford Bulletin of Economics and Statistics*, *49*(4), 431–434.

Arellano, M., & Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, *58*(2), 277–297.

Aruoba, S. B., & Diebold, F. X. (2010). Real-time macroeconomic monitoring: Real activity, inflation, and interactions. *American Economic Review*, *100*(2), 20–24.

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

Baker, S. R., & Fradkin, A. (2014). The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data. *Working Paper, Stanford University*.

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 2, pp. 195–237). Elsevier.

Brynjolfsson, E. (2012). Big Data: A revolution in decision-making improves productivity. *MIT Sloan Experts*.

Brynjolfsson, E., Geva, T., & Reichman, S. (2015). Crowd-Squared: Amplifying the Predictive Power of Large-Scale Crowd-Based Data. *MIS Quarterly (Forthcoming)*.

Chadwick, M. G., & Sengul, G. (2012). Nowcasting unemployment rate in Turkey: Let's ask Google. *Central Bank of the Republic of Turkey Working Paper 12/18*, (June).

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

Cochrane, J. H. (1991). A critique of the application of unit root tests. *Journal of Economic Dynamics and Control*, *15*(2), 275–284.

Croushore, D. (2006). Forecasting with Real-Time Macroeconomic Data. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 1, pp. 961–982). Elsevier.

D'Amuri, F. (2009). Predicting unemployment in short samples with internet job search query data. *MPRA Working Paper 18403*.

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

Einav, L., & Levin, J. D. (2013). The Data Revolution and Economic Analysis. *NBER Working Paper 19035*.

Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, *30*, 117–125.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–14.

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

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Koop, G., & Potter, S. M. (1999). Dynamic Asymmetries in U.S. Unemployment. *Journal of Business & Economics Statistics*, *17*(3), 298–312.

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

Kuhn, P., & Mansour, H. (2014). Is Internet Job Search Still Ineffective? *Economic Journal*, *124*(581), 1213–1233.

Kuhn, P., & Skuterud, M. (2004). Internet Job Search and Unemployment Durations. *American Economic Review*, *94*(1), 218–232.

Lütkepohl, H., & Xu, F. (2012). The role of the log transformation in forecasting economic variables. *Empirical Economics*, *42*(3), 619–638.

Mahmoud, E. (1984). Accuracy in Forecasting: A Survey. *Journal of Forecasting*, *3*(2), 139–159.

McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, *Q2*, 134–140.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., & Tiao, G. C. (1998). Forecasting the U.S. Unemployment Rate. *Journal of Americal Statistical Association*, *93*(442), 478–493.

Pavlicek, J., & Kristoufek, L. (2015). Nowcasting Unemployment Rates with Google Searches: Evidence from the Visegrad Group Countries. *PLoS ONE*, *10*(5), e0127084.

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

Suhoy, T. (2009). Query Indices and a 2008 Downturn: Israeli Data. *Bank of Israel Discussion Paper 2009.06*.

Tuhkuri, J. (2016). Forecasting Unemployment with Google Searches. *ETLA Working Papers 35*.

Tuhkuri, J. (2015). Big Data: Google Searches Predict Unemployment in Finland. *NTTS 2015 Conference Proceedings*.

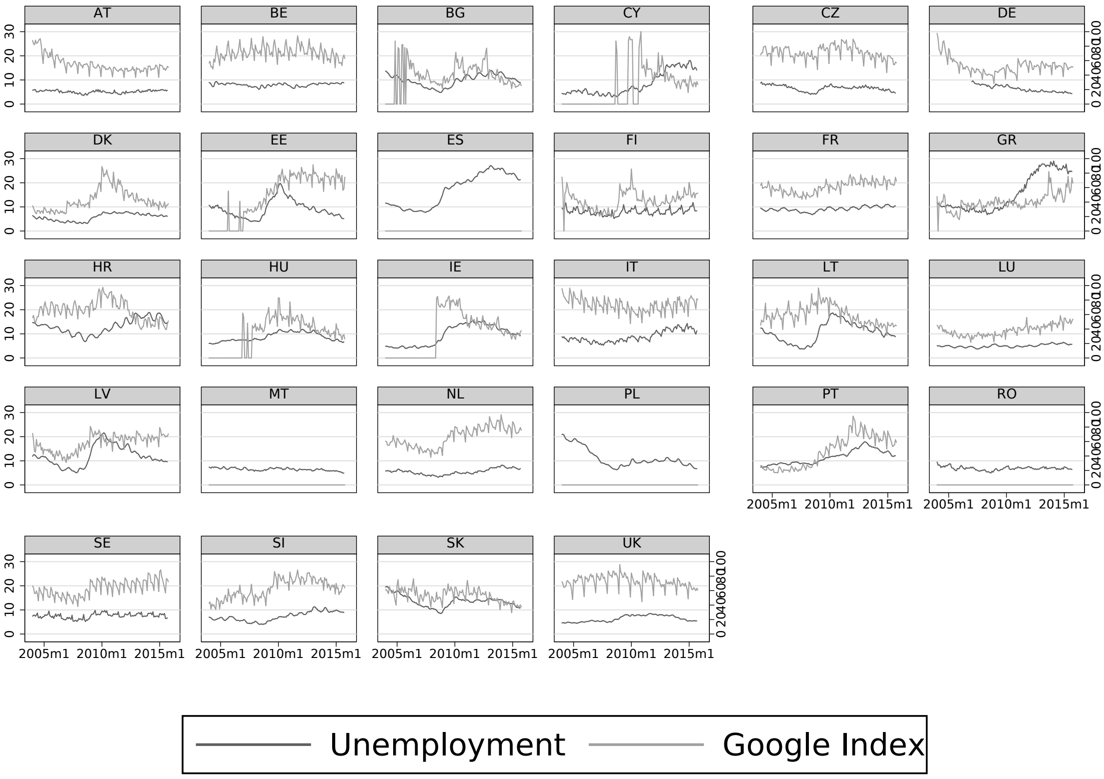
Tuhkuri, J. (2014). Big Data: Google Searches Predict Unemployment in Finland. *ETLA Reports 31*.

Vicente, M. R., López-Menéndez, A. J., & Pérez, R. (2015). Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting & Social Change*, *92*, 132–139.

Wu, L., & Brynjolfsson, E. (2015). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales. In A. Goldfarb, S. Greenstein, & C. Tucker (Eds.), *Economic Analysis of the Digital Economy* (pp. 89–118). University of Chicago Press.

Appendix

Figure A1 Unemployment rate and the Google Index that describes search activity for unemployment benefits in 28 EU countries 2004–2016



Source: Eurostat and Google Trends.

Aikaisemmin ilmestynyt ETLA Raportit-sarjassa (ennen ETLA Keskusteluaiheita)

Previously published in the ETLA Reports series (formerly ETLA Discussion Papers)

No 39 *Paavo Suni*, EMU-eron vaikutukset – Simulointeja NiGEM-mallilla. 14.11.2014. 30 s.

No 40 *Niku Määttänen – Olli Ropponen*, Listaamattomien osakeyhtiöiden verotus, voitonjako ja investoinnit. 9.12.2014. 24 s.

No 41 *Tarmo Valkonen – Eija Kauppi – Paavo Suni*, Simulointeja yhteisöveron alennuksen dynaamisista vaikutuksista Suomessa. 11.12.2014. 30 s.

No 42 *Jari Juhanko (toim.) – Marko Jurvansuu (toim.) – Toni Ahlqvist – Heikki Ailisto – Petteri Alahuhta – Jari Collin – Marco Halen – Tapio Heikkilä – Helena Kortelainen – Martti Mäntylä – Timo Seppälä – Mikko Sallinen – Magnus Simons – Anu Tuominen*, Suomalainen teollinen internet – haasteesta mahdollisuudeksi. Taustoittava kooste. 5.1.2015. 61 s.

No 43 *Annu Kotiranta – Antti-Jussi Tahvanainen – Peter Adriaens – Maria Ritola*, From Cleantech to Cleanweb – The Finnish Cleantech Space in Transition. 25.3.2015. 61 p.

No 44 *Juri Mattila – Timo Seppälä*, Laitteet pilveen – vai pilvi laitteisiin? Keskustelunavauksia teollisuuden ja yhteiskunnan digialustojen uusista kehitystrendeistä. 18.5.2015. 16 s.

No 45 *Juri Mattila – Timo Seppälä*, Blockchains as a Path to a Network of Systems – An Emerging New Trend of the Digital Platforms in Industry and Society. 13.8.2015. 16 p.

No 46 *Annu Kotiranta – Joona Widgrén*, Esiselvitys yhteiskunnallisesta yrittämisestä – Katsaus yhteiskunnallisiin yrityksiin ja vaikuttavuusinvestoimiseen Suomessa. 23.10.2015. 37 s.

No 47 *Timo Seppälä – Marco Halén – Jari Juhanko – Heidi Korhonen – Juri Mattila – Päivi Parviainen – Jaakko Talvitie – Heikki Ailisto – Kirsi-Maria Hyytinen – Jukka Kääriäinen – Martti Mäntylä – Sampsa Ruutu*, "Platform" – Historiaa, ominaispiirteitä ja määritelmä. 23.11.2015. 14 s.

No 48 *Jesper Bagger – Mika Maliranta – Niku Määttänen – Mika Pajarinen*, Innovator Mobility in Finland and Denmark. 13.1.2016. 20 p.

No 49 *Paavo Suni – Vesa Vihriälä*, Finland and Its Northern Peers in the Great Recession. 15.1.2016. 33 p.

No 50 *Antti Kauhanen – Vesa Vihriälä*, Työn määrä: Miksi Suomessa pitäisi tehdä enemmän työtä? 18.2.2016. 29 s.

No 51 *Tero Kuusi – Mika Pajarinen – Petri Rouvinen – Tarmo Valkonen*, Arvio t&k-verokannusteen vaikutuksista yritysten toimintaan Suomessa. 11.3.2016. 55 s.

No 52 *Joonas Tuhkuri – Hans Lööf – Ali Mohammadi – Petri Rouvinen*, Offshoring R&D. 4.5.2016. 13 p.

No 53 *Jyrki Ali-Yrkkö – Timo Seppälä – Juri Mattila*, Suurten yritysten ja niiden arvoketjujen rooli taloudessa. 18.5.2016. 37 s.

Sarjan julkaisut ovat raportteja tutkimustuloksista ja väliraportteja tekeillä olevista tutkimuksista.

Julkaisut ovat ladattavissa pdf-muodossa osoitteessa: www.etla.fi » julkaisut » raportit

Papers in this series are reports on research results and on studies in progress.

Publications in pdf can be downloaded at www.etla.fi » publications » reports

ETLA
Elinkeinoelämän tutkimuslaitos
The Research Institute of the Finnish Economy
Arkadiankatu 23 B
00100 Helsinki

Puh. 09-609 900
www.etla.fi
etunimi.sukunimi@etla.fi

ISSN-L 2323-2447, ISSN 2323-2447, ISSN 2323-2455 (Pdf)