Simionescu, Mihaela; Zimmermann, Klaus F.

Working Paper

Big Data and Unemployment Analysis

GLO Discussion Paper, No. 81

Provided in Cooperation with:
Global Labor Organization (GLO)

Suggested Citation: Simionescu, Mihaela; Zimmermann, Klaus F. (2017) : Big Data and Unemployment Analysis, GLO Discussion Paper, No. 81, Global Labor Organization (GLO), Maastricht

This Version is available at:
http://hdl.handle.net/10419/162198

Big Data and Unemployment Analysis

Mihaela Simionescu
Institute for Economic Forecasting of the Romanian Academy, Bucharest
Centre for Migration Studies, Prague Business School, Prague
Global Labor Organization (GLO)
mihaela_mb1@yahoo.com

Klaus F. Zimmermann
Princeton University, Princeton
UNU-MERIT & Maastricht University
Global Labor Organization (GLO)
klaus.f.zimmermann@gmail.com

June 2017

Abstract

Internet or "big" data are increasingly measuring the relevant activities of individuals, households, firms and public agents in a timely way. The information set involves large numbers of observations and embraces flexible conceptual forms and experimental settings. Therefore, internet data are extremely useful to study a wide variety of human resource issues including forecasting, nowcasting, detecting health issues and well-being, capturing the matching process in various parts of individual life, and measuring complex processes where traditional data have known deficits. We focus here on the analysis of unemployment by means of internet activity data, a literature starting with the seminal article of Askitas and Zimmermann (2009a). The article provides insights and a brief overview of the current state of research.

Key-words: big data, unemployment, internet, Google, internet penetration rate

JEL Classification: *C22, C82, E17, E24, E37*

1. Introduction

Internet data, in particular the Google search activity data have been used for nowcasting, forecasting or analysis of different variables by many researchers from different fields. For instance, for economists, policy makers and academicians, a timely information regarding the actual situation of macroeconomic indicators is essential. However, in most cases, this important information is only released with a lag by national statistical offices, often substantially late and revised. An example has been late 2008 during the Great Recession when the profession was clueless about the strenght of the economic challenge. But internet data do not only provide potentially valid data for nowcasting, but also for the analysis of human, firm and institutional behavior.

This paper aims to survey the evidence on the usefulness of internet search data in various fields, mostly in modelling unemployment in different countries. Empirical studies made for different developed states confirmed the utility of big data for modelling and predicting the unemployment rate.

The digital revolution marks the evolution from analogue and mechanical electronic technology to digital electronics. It marked the start of the information age. The mass production and the widely utilization of digital logic circuits and the associated technologies (Internet, computer, digital cellular phone) are the main pillars of the digital revolution. Network computing became a part of an increasing number of objects that were integrated in our daily lives in order to have a data driven economy as well as a data driven society (Edelman, 2012).

All aspects of life could be registered. All activities of individuals and firms are present in the internet and could provide a complete picture of the market economy and of social aspects that are reflected in the big data cloud. These informational resources might be accessed by social scientists that are conscious of the huge research potential of these data. The analysis of the historical data might be repeated as to constantly update the perspective on a certain phenomenon or process. Using the internet, answers to questions are provided before they are asked which suggests a new research strategy and innovative survey designs.

Online markets of various products and services developed fast, a particular attention being given to the analysis of job markets. This progress was favoured by social media who provide a large amount of data about individual behaviour and preferences (Askitas, 2014). Given the integration of the technology in daily life, the social component develops fast in a new direction. The high dimension of the "second economy" in the macro and micro approach is explained by the recent advances in the digital technology and in the economics of information and communication technology (Arthur, 2011). The second economy is the centre of the digital age and supposes a neural system that underlie the material world. For instance, the size of the second economy in the US will soon surpass the dimension of the physical economy.

The most utilized place of the second economy is the internet where social media are operating. Nowadays, there are very popular products available like Google+, Facebook, LinkedIn, Twitter or YouTube. The data provided by official statistics will be completed by the data from the second economy, technology miniaturization, social media and the internet (Askitas and Zimmermann, 2011). Google allows the construction of real-time aggregated data with high frequencies on the search for those keywords of interest to the researcher which receive a significant amount of traffic (Askitas and Zimmermann, 2009a). However, Google does not reveal these thresholds.

Even if methodological progress in economics have manifested in big steps, measurement still has many deficiencies, while many indicators being just estimated or often

revised. In this context, internet search data represent a relevant alternative and substantial potential, even if there are limits. For analysing and predicting unemployment, Google search data could be very useful.

In this paper, in section 2 we first discuss internet activity data in general. Then, in section 3 we focus on the internet data used in modelling the unemployment rate starting with the valuable experience of Askitas and Zimmermann (2009a). Finally, In section 4 we draw some conclusions.

2. Internet activity data

In the 1980s, when the internet activity was at the beginning, researchers in the social sciences considered it as a good environment for collecting data using online surveys or other methods. The advantages were related to price and speed. In the 1990s, the internet spread very quickly and became a part of the people's lives and home because of many advantages: fast communication via email and other facilities, surfing or searching for a specific answer (Askitas and Zimmermann, 2015). In the 2000s, web technologies developed more coming up with better techniques. The individuals used more intensively the internet and the internet provided huge amounts of data. At the very beginning, the individuals even did not know that data about them were collected and stored. Unlike with traditional surveys, where data collection has the consent of the subjects investigated, individuals are now studied by observing behavior and preferences in the privacy of their homes or offices. Personal information of various types was spread with the entrance of Google into the market. Popular internet data sources next to Google are Google+, Facebook, Wikipedia, LinkedIn, Twitter, among others.

Constant and Zimmermann, (2008) and Askitas and Zimmermann (2009a) were among the very first to study the usefulness of google activity data for the analysis of social science issues, including the US presidential elections, unemployment and the Great Recession. Goel et al. (2010) provided a large survey in the field of internet activity data, describing the advantages, but also the limitations of big data. Internet data have a large number of advantages: easy to store, organize and work with, because they are generated digitally. They allow for accurate measurement in cross-sectional and transversal dimension since they are geo-tagged and time-stamped (Askitas and Zimmermann, 2009a). Internet data could be utilized in more informed, timely and effective policy making for society benefits, mostly in crisis times. In this context, the relationship between theory and the empirical data is changed. Big data supposes large numbers of observations and allows for flexible experimental settings and conceptual forms. Search activity data permit analyses in various combinations of time, space and contexts, favoring multidisciplinary research and providing indirect panel survey data. In times of crisis, the breaking trends are timely identified, because the data are provided with high frequency and in almost real time.

The Disadvantages might be related to the fact that the data are only made available in aggregate form (Askitas and Zimmermann, 2015). The data methodology is not well documented. The internet activity is captured by the chosen search keywords. However, the relevance of those keywords may change across regions and in time. The Google page rank can affect demand and supply. The geo-location is considered using the IP address that is only available at country level. Improvements are necessary for smaller areas. The samples may not be representative for the entire population even if the samples are based on a large number of activities since internet use might be biased. For instance, a study by McLaren and Shanbhogue (2011) showed that internet use depends on age and income.

Selection bias remains an important issue because of the differences between individuals and countries in responding to new technological waves (Zagheni and Weber, 2015). The internet penetration rate represents the percentage of the entire population of a given country that uses the internet. There are countries where internet penetration is more than 90 percent of the population, but in other states the coverage is lower. In the European Union, the internet penetration rate is 80.1%, according to the European Union Internet Statistics (2016) lastly updated on the 30th of June 2016. In 2016, the internet penetration rate was 89% in Germany, 91.6% in the UK, 95.9% in Denmark 96.3% in Norway, and only 88.1% in the US. Even in countries with high internet penetration rates, not everyone uses social media or smart phone which bring selection bias.

In the future, new data will be increasingly measured by (objective) embedded sensors that will provide information about individual vital signs, location, human and economic activity. As a consequence, the economy will be more data dependent and the research opportunities will grow. New technologies and their combination will provide additional new data and new challenges as Askitas and Zimmermann (2013) have shown.

The geographic gap of surveyors in terms of sample size, scale and frequency was covered by the internet data with no marginal cost in case of online surveys or e-mails as Askitas and Zimmermann (2015) indicated. Seen as a survey platform, the internet came up with new methodological challenges and with a high potential. Being present anywhere, the internet allows the construction of representative and random samples. In case of full access to data, the bias selection is eliminated, because the online population is very close to the general population. Therefore, the samples become representative and they are random. In this context, sampling is not anymore necessary since we work with unlimited quantities of data. A famous example of large-scale survey based on internet data is the Wage Indicator Survey of the Wage Indicator Foundation[1]. Wages based on individual reports are provided in more than 20 languages and for more than 60 countries. Harmonized data on wages are provided for a large sample of countries. Issues of selection bias were observed, but research is ongoing to eliminate this disadvantage.

The surveys constructed on the internet became an essential tool for collecting data. The Information and Communication Technology and the internet have the advantage of reducing frictions in matching tasks in almost any kind of market. Matching is not essential only in real life, but also in economics where matching problems and the optimal solution is the objective of research in economics (matching long distance passengers to airplane seats or travelers to taxis). Matching individuals in the job market (Kuhn, 2014) or in the marriage market (Hitsch et al., 2010) are other examples that underline the internet advantage of reducing the search frictions which offer new business opportunities like job board services or online dating services. This new data potential about economic behavior in different contexts helps in reconsidering fruitfully old but still unsolved questions. Actually, the internet is the one that allowed the replacement of different labour markets. For example, if someone needs the help of a medical doctor, lawyer, fitter etc., he/she has only to type these words and find in a very short time hundreds of options. On the other hand, many employers search for human resources using the internet (for example, through LinkedIn). The Great Recession (2008-2011) confirmed the huge potential of the internet, since it revealed the activity of people that intensively searched there to find a job.

The market of internet search engines matches the demand and the supply for documents. The demand for information is correlated to the supply of documents that include this type of information. In this case, an overall image of the demand in time for this

[1] http://www.wageindicator.org/main/Wageindicatorfoundation/researchlab/wageindicator-survey-and-data

information is provided and consequently, we know the state of the individuals searching for this type of information. Google Trends and the Google's business model provide us with a global image of this demand. This idea was promoted in the studies of Askitas and Zimmermann (2009a), Askitas and Zimmermann (2011) and Askitas and Zimmermann (2015) for Google Trends data and in the work of Askitas and Zimmermann (2013) for technological data.

The data-provisioning tool called Google Trends was introduced in the summer of 2008 to provide a public view into the relative internet search volumes in case of some queries, whereas the user can freely define the keywords for the queries.. Google Trends then provides a time series index corresponding to the volume of the queries of the users that were introduced into Google in a specific geographic zone. The query index is computed as the total query volume for a specific keyword that was searched within a certain geographic area divided by the total number of queries in the same area and during a certain period. The maximum query share in a period is normalized at 100, while the query share at the initial date is normalized to zero (Choi and Varian, 2012).

The advantages and the limits of Google Trend are well described in the paper of Askitas (2015). The Google Trends team used the term of "sessionization" to show that search data are standardized as to reduce the noise from typing errors, frivolous repetitions, rewrites and other types of acts. The search session can be geo-located based on the IP address where the sessionis initiated. The scientific potential relies on the ability of the user to define the set of relevant variables and construct their content by defining and merging keywords. It is then possible to easily examine the consequences of different concepts.

However, the tool permits only an aggregate image of the behavioral microdata. The methodology is not well described and it lacks versioning. Google Trends is efficient for large search volumes and in places with high internet penetration. The IP address is available only at the country level. Moreover, the access to data is conditioned by Google that can change the commitment regarding data provision. It is also important to note that the data provided is based on a representative subsample that is freshly drawn when a new data set is created. Hence, the researcher needs to store the data to undertake exact replications of the studies.

An important issue for economists is the way to register and measure all transactions that are made using internet. Issues related to ownership and data custody as well as data privacy should be solved by keeping in mind that privacy protection is an individual right (Askitas and Zimmermann, 2011). The institutional structures for data provision should be improved as to avoid the data monopoly of some companies. In most cases, the data are not broadly available. On the other hand, there are a lot of questions related to government use of data about citizens. Internet data might be used in economic policy making. However, the banks could monitor the clients' transactions behavior in real time and data protection of their clients is not ensured. McLaren and Shanbhoge (2011) explained how web search data may be utilized for economic nowcasting by the national banks.

Internet data can be applied in many fields to solve human resource issues: nowcasting (relevant information is gotten earlier than through traditional ways of collecting data) as in McLaren and Shanbhoge (2011), Askitas and Zimmermann (2013), Carrière-Swallow and Labbé (2013), Chen et al. (2015), forecasting (for example, forecasts for unemployment rate, consumption of various goods, tourism arrivals, festival winners) as in Askitas and Zimmermann (2009a), Vosen and Schmidt (2011), Choi and Varian (2012), Artola et al. (2015), identification of health issues and well-being (malaise, flu, and ill-being in times of economic crisis) as in Ginsberg et al. (2009), Yang et al. (2010), Tefft (2011), Lazer et al. (2014), Askitas and Zimmermann (2015), documenting the matching process in different situations of life (e.g. partnership, jobs, shopping) as in Askitas and Zimmermann (2009a), Kuhn and Mansour (2014), Kuhn (2014), Kureková et al. (2014) and measurement of complex

processes when traditional data have deficits (e.g. collective bargaining agreements in developing countries, international migration) as in Hitsch et al. (2010), Reips and Buffardi (2012), Billari et al. (2013), Besamusca and Tijdens (2015) and Bellou (2015).

3. Modelling unemployment using internet data

In most cases, macroeconomic time-series are published with a significant delay and may be subjects to various revisions. Unemployment data are also published with delays. In this context, there is an increasing demand for the real-time estimation of changes in unemployment (Fondeur and Karamé, 2013).

For the EU countries, the European Community ask for large sets of data used in making economic analyses. The data are drawn from many scale surveys based on censuses and samples. These requirements are specified in the EC regulation on short-run business statistics since August 2005, corresponding to the launch of the Action Plan on EMU Statistical Requirements by the Central European Bank and Eurostat and with the support of the EU national institutes of statistics. The main objective was to decrease the long period that is necessary for the creation and circulation of the essential indicators used in short-run economic analyses of the EU economies (Naccarato et al., 2015).

Given the recent economic and financial crisis with the strong decline in the economic activity, the unemployment is a macroeconomic indicator of particular interest not only to the general public, but also for research.

Short-run information on unemployment was required during the Great Recession, but it was unavailable. The seminal paper of Askitas and Zimmermann (2009a) showed for the first time the strong correlations between monthly unemployment rates in Germany and specific Google keyword searches. The observed structure was used by the authors to forecast unemployment behavior under the changing and complex circumstances of the upcoming Great Recession. Askitas and Zimmermann (2009a) used the time-series Granger causality approach to explain the monthly German unemployment rate through changes in other variables of interest. Error correction models were built using the seasonally unadjusted unemployment rate in the period January 2004 - April 2009. The authors used various search search terms like ‘unemployment rate’, ‘unemployment office or agency’, ‘most popular search engines in Germany’ and ‘Personnel Consultant’.

In Askitas and Zimmermann (2009b), the authors have re-estimated updated models using improved keywords to study the quality of unemployment analysis and the prediction performance comparing it also with prominent rival labour market indicators. We shortly summarize this effort here to explain the steps of the strategy and explain the major contributions. The core regression equation is the prominent error correction model (with Y: the unemployment rate and X: the indicator vector):

$$(1) \qquad \Delta Y_t = \alpha + \beta y_{t-12} + \sum_{i=1}^{k} (\gamma_i \Delta X_{i,t} + \delta_i X_{i,t-12})$$

with $\Delta Y_t = Y_t - Y_{t-12}$; $\Delta X_t = X_t - X_{t-12}$; Δ: lag operator of lenght 12

and t = 1, 2, ..., n

The internet activity indicators or search keywords have been related to "labour office", "short-term work" and "jobsearch", see Askitas and Zimmermann (2009b) for technical details using German data at the beginning of the Great Recession. Table 1 summarizes the estimates

based on the standard Ordinary Least Squares (OLS) estimation technique and contains evaluation measures like the corrected Coefficient of Determination (R^2), the Bayesian Information Criterion (BIC) and the Mean Absolute Error (MAE). From the first three lines it is obvious that the internet indicators affiliate very well with the actual unemployment rate, while the model with all three indicators performs best (see row 2), eg. the R^2 is 0.943. The estimation coefficients are all well statistically significant, and their signs indicate that searching for jobs reduces unemployment while searching for information about the labour office and financial support through short-term work is affiliated with a rise in unemployment. These findings make much sense.

A prominent traditional labour market indicator collected by the Ifo-Institute which is based on individual company data is Ifo-BB, a measure that is often used as a reference variable predicting the labour market. As has been demonstrated by Askitas and Zimmermann (2009b), the DAX, a German stock-market index lagged for one year (meaning one year earlier!) is predicting equally well and his highly correlated with Ifo-BB. Both indicators alone and separate also affiliate well with the German unemployment rate, see Table 1, row 7 for the DAX and row 11 for Ifo-BB. However, their performance is much worse that the pure internet activity model of row 2 involving the full set of internet data used. However, it is also true that the prediction quality of the internet variables can be increased by using traditional variables. Judging this by following the BIC measure (see Table 1), the BIC value of row 2 (all three internet variables only) of value 28.8 can be reduced to value 11.4 when Ifo-BB is added (see row 9) and to value 3.2 when instead DAX is added (see row 6).

The conclusion from this historical example is that there is valuable, useful and useable information in the internet activity data. However, we need more experience with using the new technique and see to what extend this new data can replace traditional sources of information. It is not yet a priori clear that one can fully replace traditional data by internet data.

The concept of Askitas and Zimmermann (2009a) for modelling unemployment using internet activity data was also followed by researchers in other countries. The empirical findings suggest that Google or other internet activity data add relevant additional information for explaining unemployment compared to business cycle indicators or traditional time-series models. Similar studies were made for the unemployment rate in the UK (McLaren and Shanbhogue, 2011), France (Fondeur and Karamé, 2013), Israel (Suhoy, 2009), Italy (D'Amuri, 2009, Naccarato et al., 2015), Norway (Anvik and Gjelstad, 2010), Turkey (Chadwick and Sengül, 2015), Brazil (Lasso and Snijders, 2016), for unemployment levels in Spain (Vicente et al., 2015) and Ukraine (Oleksandr, 2010), for claims regarding unemployment benefits in the US (Choi and Varian, 2009; Choi and Varian, 2012) and for unemployment internet search indicators from Google and Baidu in China (Su, 2014). According to the Granger causality test, unemployment-related search indices have the ability to improve forecasts of different macroeconomic indicators also in China (Su, 2014).

Before Google activity data were available, Ettredge et al. (2005) used internet search engine keyword usage data from WordTracker's Top 500 Keyword Report that was weekly published by Rivergold Associates Ltd. This report covered the Web's largest meta-search engines. The authors used six expressions that could be mostly used by people seeking for a job and used them to predict unemployment rate in the US: jobs, job listings, namely job search, resume, employment and monster.com.

Most of the mentioned studies employ a large set of Google queries. Some principal components might be extracted in order to reduce the dimensionality. These components are introduced as explanatory variables in models like ARMAX. Choi and Varian (2009) selected two indicators: "welfare & unemployment" and "jobs". In case of the US, Choi and Varian (2009) found that unemployment and welfare-related searches may improve forecasts of initial

jobless benefit claims. For the US, D'Amuri and Marcucci (2009) only used the keyword 'jobs'. They showed that Google index (internet job-search indicator) is the best leading indicator to forecast the unemployment rate in the US. For Germany, Askitas and Zimmermann (2009a) use four groups with one to eight terms and the operator 'or'. For Spain, Vicente et al. (2015) used Google Trends indicators for queries like 'oferta de trabajo' and 'oferta de empleo' (job offers).

For Italy, Naccarato et al. (2015) analyzed the cointegration relationship between the official unemployment rate from Labour Force survey and the Google Trend query 'offerte di lavoro' (job offers). In previous studies for Italy, D'Amuri and Marcucci (2009) and D'Amuri (2009) showed that 'Offerte di lavoro' is the most popular keyword used for job searches in this country. Naccarato et al. (2015) showed that Google search is a useful tool in nowcasting the Italian unemployment rate. The same keyword 'Offerte di lavoro' was previously used by Francesco (2009) to show that the models based on Google search data improve the out-of-sample forecasts of the unemployment rate in Italy.

Moreover, Barreira et al. (2013) analyzed the usefulness of Google search for more South-western countries and concluded Google Trends data improved the unemployment in Italy, France and Portugal, but not in Spain. The keywords were related to unemployment and its benefits: 'disoccupazione', 'disoccupazione ordinaria', 'INPS disoccupazione' (INPS is the Italian National Institute for Social Security) in Italy, 'chomage', 'indemnites de chomage', 'allocations chomage', 'allocations de chomage' in France, 'desemprego' and 'subsidio desemprego' in Portugal and 'desempleo', 'subsidio de desempleo' and 'prestacion desempleo' in Spain.

McLaren and Shanbhogue (2011) analyzed the relationship between the official unemployment rate in the UK and some search term data ('unemployment ', 'jobs', 'unemployed', 'JSA', 'Jobseeker's Allowance', 'unemployment benefit') using autoregressive models. The authors proved that search data include useful information compared to existing surveys. The JSA model explained the unemployment better than baseline model that uses only official data for unemployment.

Fondeur and Karamé (2013) built unobserved components models treated with Kalman filter and the maximum likelihood estimation method. This approach allows the restoration of unobservable components and the estimation of unknown parameters. The authors used as variables the Google index and the claimant count for people between 15 and 24 years old with data for France.

The limited access to the internet and lower literacy rate in transition countries make it more difficult to extrapolate western models. Hence, for Ukraine, Oleksandr (2010) did not confirm the usefulness of internet data for explaining the unemployment rate. This may, however, change as soon as the internet plays a more substantial role in Ukrainian economic life. Or it could result from the fact that the successful strategy was not yet revealed. As should be expected, the stability of the structures identified with internet data may be limited over time even for developed countries. This suggests larger challenges for transition and developing countries. However, those challenges are also present in traditional data and models.

Pavlicek and Kristoufek (2015) analyzed the relationship between the monthly unemployment rate and job-related searches in the Visegrad countries (V4 countries) in the period January 2004 - December 2013. Only for Czech Republic and Hungary the Google searches contribute valuable information in explaining the unemployment rate. This might be related to the fact that these countries sent many migrants abroad that were interested in jobs outside the country of origin. The situation was not yet examined for Poland and the Slovak Republic.

On the other hand, for the emerging economy Brazil, Lasso and Snijders (2016) showed that Google searches strongly correlated with unemployment, but seasonal patterns had a

higher impact. The authors used more keywords: ‘empregos’, ‘seguro desemprego’, ‘décimo terceiro salário’, ‘FGTS’ (Severance Indemnity Fund), ‘INSS’ (National Institute for Social Security), ‘job vacancies index’ and ‘unemployment and social benefits index’. For Turkey, Chadwick and Sengül (2015) used terms like: ‘unemployment’, ‘unemployment insurance’, ‘job announcements’, ‘looking for a job’, ‘cv’ and ‘career’. Using the framework of Bayesian Model Averaging, the authors obtained that Google search data is useful in nowcasting the monthly unemployment rate in Turkey only in nonagricultural sectors. The official data for unemployment rate were provided by Household Labor Survey, while the internet data were collected using Google Insights for Search.

4. Conclusions

In recent years, the availability of internet data encouraged researchers to use them in order to analyze or forecast macroeconomic indicators. This might be explained not only by the fact that the data are accessible, plentiful, economical and digitally organized, but also by the reason that the internet became a part of everyday life of individuals and measures increasingly reality in terms of behavioral trends.

Approximations of the changes in unemployment are mostly based on official governmental sources or on surveys that might not be always reliable. Furthermore, in developing countries, the responsible institutions are often unable for various reasons to provide valuable estimates of macroeconomic indicators like unemployment. Most of the previous research on unemployment nowcasting deals with developed countries like US, UK, Italy, Germany, Finland or Belgium. Few studies are dedicated to non-western states (V4 countries, Ukraine, Turkey, Brazil) with weaker public institutions.

In this paper, we examined the use of internet activity data in different fields, focusing on their use in modelling unemployment. The empirical studies reviewed here suggest that there is a strong potential that needs to be further explored. In most of the countries, internet data improved the models and the forecasts of unemployment. However, the forecast accuracy depends on the internet penetration in each country, the age structure of the internet users and the stability of the constructed internet variables.

References

Anvik, C., & Gjelstad, K. (2010). Just Google it. Forecasting Norwegian unemployment figures with web queries. *Working Paper*, 11, Center for Research in Economics and Management, Oslo.

Arthur, W.B. (2011). The second economy. McKinsey Quarterly, October.

Artola, C., Pinto, F., & de Pedraza, P. (2015). Can internet searches forecast tourism inflows? *International Journal of Manpower*, 36(1), 103-116.

Askitas, N. (2014). Social media: eine technologische und ökonomische Perspektive, in Rogge, C. and Karabasz, R. (Eds), Social Media im Unternehmen – Ruhm oder Ruin, Springer Vieweg, Wiesbaden, pp. 155-166.

Askitas, N. (2015). Google search activity data and breaking trends. *IZA World of Labor*.

Askitas, N., & Zimmermann, K. F. (2009a). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, *55*(2), 107-120.

Askitas, N., & Zimmermann, K. F. (2009b**).** Googlemetrie und Arbeitsmarkt. Wirtschaftsdienst, 89 (7), 489-496.

Askitas, N., & and Zimmernann, K.F. (2011). Detecting mortgage delinquencies. *IZA DP* 5895, IZA, Bonn.

Askitas, N., & Zimmermann, K.F. (2013). Nowcasting business cycles using toll data. *Journal of Forecasting*, 32(4), 299-306.

Askitas, N., & Zimmermann, K.F. (2015). Health and well-being in the Great Recession. *International Journal of Manpower*, 36(1), 26-47.

Barreira, N., Godinho, P., & Melo, P. (2013). Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends. *NETNOMICS: Economic Research and Electronic Networking*, *14*(3), 129-165.

Bellou, A. (2015). The impact of Internet diffusion on marriage rates: evidence from the broadband market. *Journal of Population Economics*, *28*(2), 265-297.

Besamusca, J., & Tijdens, K. (2015). Comparing collective bargaining agreements for developing countries. *International Journal of Manpower*, *36*(1), 86-102.

Billari, F., D'Amuri, F., & Marcucci, J. (2013). Forecasting births using Google. Annual Meeting of the Population Association of America, PAA, New Orleans, LA.

Carrière□Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289-298.

Chadwick, M. G., & Sengül, G. (2015). Nowcasting the Unemployment Rate in Turkey: Let's Ask Google. *Central Bank Review*, *15*(3), 15.

Chen, T., So, E. P. K., Wu, L., & Yan, I. K. M. (2015). The 2007–2008 US Recession: What Did the Real-Time Google Trends Data Tell the United States?. *Contemporary Economic Policy*, *33*(2), 395-403.

Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc*, 1-5.

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

Constant, A., & Zimmermann, K. F. (2008). Im Angesicht der Krise: US-Präsidentschaftswahlen in transnationaler Sicht. *DIW Wochenbericht*, *44*, 688-701.

D'Amuri, F. (2009). *Predicting unemployment in short samples with internet job search query data*. University Library of Munich, Germany.

Edelman, B. (2012). Using internet data for economic research. *The Journal of Economic Perspectives*, *26*(2), 189-206.

Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, *48*(11), 87-92.

European Union Internet Statistics (2016). Internet Usage in the European Union. Available online at: http://www.internetworldstats.com/stats9.htm

Fantazzini, D. (2014). Nowcasting and Forecasting the Monthly Food Stamps Data in the US Using Online Search Data. *PloS one*, *9*(11), e111894.

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

Francesco, D. A. (2009). Predicting unemployment in short samples with internet job search query data. *MPRA Paper*, 18403, 1-18.

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), 1012-1014.

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*, *107*(41), 17486-17490.

Hitsch, G. J., Hortaçsu, A., & Ariely, D. (2010). Matching and sorting in online dating. *The American Economic Review*, *100*(1), 130-163.

Kuhn, P. J. (2014). The internet as a labor market matchmaker. *IZA World of Labor*.

Kuhn, P., & Mansour, H. (2014). Is Internet job search still ineffective? *The Economic Journal*, *124* (581), 1213-1233.

Lasso, F., & Snijders, S. (2016). The power of Google search data; an alternative approach to the measurement of unemployment in Brazil. *Student Undergraduate Research E-journal!*, *2*.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203-1205.

McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. Bank of England Quarterly Bulletin, June 2011.

Naccarato, A., Pierini, A., & Falorsi, S. (2015). *Using Google Trend Data To Predict The Italian Unemployment Rate* (No. 0203). Department of Economics-University Roma Tre.

Oleksandr, B. (2010). *Can Google's search engine be used to forecast unemployment in Ukraine* (Doctoral dissertation, Kyiv School of Economics).

Pavlicek, J., & Kristoufek, L. (2015). Nowcasting unemployment rates with google searches: Evidence from the visegrad group countries. *PloS one*, *10*(5), e0127084.

Reips, U. D., & Buffardi, L. E. (2012). Studying migrants with the help of the Internet: methods from psychology. *Journal of Ethnic and Migration Studies*, *38*(9), 1405-1424.

Su, Z. (2014). Chinese online unemployment-related searches and macroeconomic indicators. *Frontiers of Economics in China*, *9*(4), 573-605.

Suhoy, T. (2009). *Query indices and a 2008 downturn: Israeli data*. Bank of Israel.

Tefft, N. (2011). Insights on unemployment, unemployment insurance, and mental health. *Journal of Health Economics*, *30*(2), 258-264.

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 and Social Change*, *92*, 132-139.

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey□based indicators vs. Google trends. *Journal of Forecasting*, *30*(6), 565-578.

Yang, A. C., Huang, N. E., Peng, C. K., & Tsai, S. J. (2010). Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. *PloS one*, *5*(10), e13728.

Zagheni, E., & Weber, I. (2015). Demographic research with non-representative internet data. *International Journal of Manpower*, *36*(1), 13-25.

Table 1: Results of Regression Models and One-step-ahead Forecasts

Model	Labour Office	Short-term Work	Jobsearch	Ifo-BB	DAX	R^2-a	BIC	MAE
Prediction 1	L*** + K +		L*** - K -			0.862	69.010	0.434
Prediction 2	L*** + K* +	L** + K*** +	L*** - K*** -			0.943	28.802	0.354
Prediction 3		L*** + K*** +	L*** - K -			0.923	38.986	0.420
Prediction 1 + DAX	L*** + K +		L*** - K*** -		L*** - K*** -	0.950	21.589	0.263
Prediction 2 + DAX	L*** + K +	L + K*** +	L*** - K*** -		L*** - K* -	0.969	3.178	0.297
Prediction 3 + DAX		L*** + K** +	L*** - K*** -		L*** - K* -	0.955	16.177	0.429
DAX - Prediction					L*** - K*** -	0.887	53.216	0.314
Prediction 1 + ifo-BB	L*** + K** +		L*** - K*** -	L*** - K*** -		0.950	21.645	0.333
Prediction 2 + ifo-BB	L*** + K** +	L + K*** +	L*** - K*** -	L*** - K** -		0.963	11.368	0.414
Prediction 3 + ifo-BB		L*** + K +	L*** - K* -	L*** - K* -		0.938	32.593	0.550
ifo-BB - Prediction				L*** - K*** -		0.863	63.213	0.541

Notes: Adapted from Askitas and Zimmermann (2009b), p. 495. Data are from Arbeitsamt.de, Ifo-Institute and Google Insights. Ifo-BB: Employment indicator of the Ifo-Institute, Munich. DAX: Stock-market index. The used offical monthly unemployment rate is seasonally unadjusted, but seasonality got covered through the 12th difference in modelling. For more details on the keywords see Askitas and Zimmermann (2009a, 2009b). All standard regression models following equation (1) were estimated using monthly data for January 2005 to May 2009. K represents the change and L the 12th lag of the level of the corresponding variable. +, - are the signs of the estimated coefficients and "*" reflects the statistical significance (* P < 0,05, ** P < 0,01, **** P < 0,001). The One-step-ahead Forecasts were executed for March 2008 to June 2009. R^2-a is the corrected Coefficient of Determination, BIC the Bayesian Information Criterion and MAE the Mean Absolute Error.