The power of Google search data; an alternative approach to the measurement of unemployment in Brazil

Fernando Lasso (406436) and Sebastian Snijders (414579)
Bachelor Honours Class 2015/2016
Erasmus School of Economics

(Dated: May 29, 2016)

Abstract: By the use of econometric techniques, this paper extends the application of predictive methods for unemployment rates by the use of Google Trends from developed western countries to the case of Brazil. Indeed, Google search volumes of keywords related to job-search turn out to contain significant predictive power and biweekly search data can predict the direction of the unemployment rate with around 80 percent accuracy, beating baseline methods using seasonal patterns by more than 10 percent.

I. INTRODUCTION

Since the advent of the internet, there has been an explosion in collected and obtainable data. However, in the domain of economic indicators, a significant amount of the available data is still not being used to its full potential. Several studies have shown how Google search data can be a powerful predictive tool for economic indicators such as GDP, inflation, retail sales and many more (Koop et al., 2013; Guzman, 2011; Choi & Varian, 2012). Especially the estimation of unemployment figures has been a topic of interest across a multitude of developed countries, which resulted in methods that are able to put real-time search data into reliable projections. This procedure, dubbed as nowcasting the economy, led to successful econometric models in Germany, Belgium, the United Kingdom and Finland (Askitas & Zimmermann, 2009; Bughin, 2011; McLaren & Shanbhogue, 2011; Tuhkuri, 2014). Approximations of the development of the unemployment rate rely heavily on surveys and official reports of other government entities, which may not always be a reliable source. Especially in developing nations, institutions may not possess the accuracy that may be desired for a fundamental analysis of their economies. Unfortunately, previous research on the nowcasting of unemployment lacks the extension to non-western countries with weaker institutions. This paper will tackle the case of Brazil and attempt to broaden previous models to provide real-time estimates of unemployment in this South American country.

Currently, Brazil faces a desperate recession and a political crisis. This unfortunate situation will inevitably result in the loss of thousands if not millions of jobs. It would be of social interest to produce real-time estimates of the unemployment rate or to forecast these figures using alternative methods. The social and economic benefits of such a model range from being a useful tool in policy making to producing early warning signals of economic crises in the future. One simple and accessible source of data to facilitate the estimation of these indicators in a timely manner is the volume of Google keyword searches, which is available in real-time and sorted into fine-grained categories. To contribute to the subject, this paper will explore the following research question:

To what extent do Google keyword search volumes contain explanatory and predictive power for the monthly unemployment rate of Brazil?

Besides establishing the internal validity of the model, an approach must be taken to check the reliability of the predictions. By assumption, official government reports on unemployment lack optimal reliability. This is partly due to the extend of hidden unemployment in the country, which blurs the lines of how to define unemployment and retrieve the exact figures. As will be discussed in the theoretical framework, this phenomenon causes government estimates to be susceptible to adjustments. After a valid model to account for the issue has been established, past data could be analyzed to timely estimate the direction unemployment may be going in the future.

After the introduction, the theoretical framework will describe the economic landscape of Brazil, extend on the use of Google search data in past research and lay out the hypotheses of the paper. Then, the data and methodology sections will go into detail on the selection of appropriate sources, the cleansing of data and the use of models. In the results section, four specifications of our model will be presented, each containing a significant degree of predictive power. The final models can predict the direction in which the unemployment rate will move with around 80 percent accuracy halfway through the month. These techniques significantly improve baseline methods based on seasonal patterns, and thus can be deemed useful tools in decision making or fundamental analysis using forecasts. Finally, the conclusion will wrap up the paper with a summary of relevant outcomes and interpretations.

II. THEORETICAL FRAMEWORK

Over the past years, the massive Brazilian economy has slipped from its status of a rising super-power into a seemingly hopeless recession. After a decade of ever rising economic growth, 2015 ended an era with growth

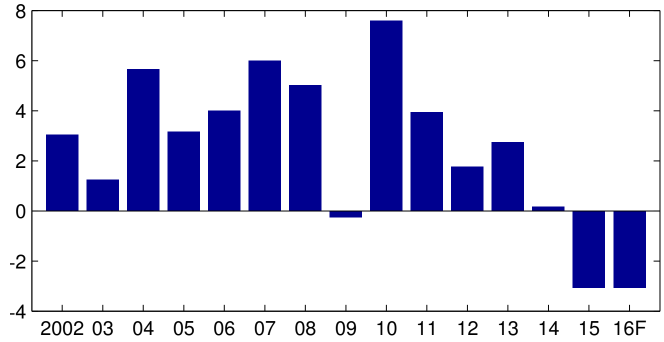


FIG. 1. Percentage change in Brazil's real GDP, 2002 - 2016 with expert forecast. Adapted from The Economist, *Brazilian Waxing And Waning*, Data Team, April 2016.

rates of around 4% per year and plunged the economy into a downturn, as can be seen in Figure 1. The causes for the decline are not clear-cut, but it is suspected that poor infrastructure and strong currency have rendered the Brazilian economy uncompetitive. The situation is worsened by Brazils insecure political environment, triggered by widespread corruption scandals. President Rousseff has been impeached by the Senate and politicians throughout all levels of government are suspected of accepting large bribes related to the state-owned oil giant Petrobras.

The number of issues Brazil is facing is larger than this paper wishes to delve into, but as may be expected, the recent setbacks discredit the government and weaken incentives to invest in this once promising nation. Inevitably, reduced investments and weaker cash flows force industries to shut down and may cost the Brazilian work force many jobs. The approximation of unemployment is conducted by government authorities and released once per month. However, the turbulent state of affairs in Brazil may demand a more efficient and real-time estimation of these figures, especially because the developing nation has seen a high volatility in unemployment rates over the past couple of years. Therefore, this paper proposes an alternative method to effectively trace the volatile unemployment figures in Brazil, by making use of publically available Google search trends. The internet plays a growing role as the primary intermediary between employees and businesses. This gives plenty of reason to suspect that variation in search behavior associated with job searches may have a strong correlation with the actual unemployment. Hence, the following hypothesis is investigated.

Hypothesis 1: *The volume of Google keyword searches related to unemployment is correlated with the monthly unemployment rate in Brazil.*

Previous research in this field has already been conducted by Amuri and Marcucci (2012), who established that the volume of Google keyword searches provides explanatory power for the unemployment rate in the United States. In addition, Google keyword searches have been used to develop successful predictive models for the unemployment rate in Germany, Belgium, the United Kingdom and Finland (Askitas & Zimmermann, 2009; Bughin, 2011, McLaren & Shanbhogue, 2011; Tuhkuri, 2014). All of these models have succeeded in improving or even replacing other available methods of estimation. However, as different countries may have distinct forces that influence their economic environment, it is not possible to extrapolate European models to the Brazilian economy. Furthermore, non-western societies are more likely to face lower literacy rates and limited access to the internet, which may hinder previously proposed models. Therefore, this paper will attempt to fill the void in the existing literature by investigating the feasibility of a predictive model for the case of Brazil.

Hypothesis 2: *The volume of Google keyword searches related to unemployment contains predictive power for the monthly unemployment rate in Brazil.*

Since it is the aim to approximate real unemployment rates, reliable estimates of the actual figures will be needed to test the model against. Unfortunately, official government estimates have been prone to adjustment in the past (Reuters, 2014). This is largely due to the issue of hidden unemployment. Many Brazilians from the lower social classes have given up on their search for work in the controlled industries and recede into low-end jobs, such as street vendor activities, informal trading and low skill manual labor. As these people are not actively searching for employment, they are not accounted for in official government figures, even though their jobs are likely to be below their qualifications. In developing countries that face this scenario, it may be difficult to exactly define the boundary as to who qualifies as being unemployed. Because the unemployment rate is a leading indicator of an economic success, it should incorporate these cases as well. The adoption of new methods and further revision caused official unemployment estimates to be volatile, even many years after their initial publishing. Therefore, a special procedure will be applied to address the issue of unreliable government data. For the purpose of this paper, the range of the analysis will be split into period one, the time before Rousseff was elected in 2011 and period two, the years following her election. As the unemployment rates in period one have had sufficient time to be revised, the data will be fitted onto the official figures between 2004 and 2011. By making the assumption that the official government report error terms are not negatively or positively biased, the model will be used to reliably forecast the unemployment rate in period two.

If a reliable model to timely predict the unemployment rates can be established, it would be a great achievement to find an internally valid model from the data before 2011 that is able to detect the increase in unemployment observed in 2015. To compare the effectiveness of the model, it must be compared to alternative techniques of estimation. The methodology section will lay out two alternative baseline methods to test the last hypothesis.

Hypothesis 3: *The model for estimating Brazils unemployment rates by Google search volumes significantly improves baseline forecasting techniques.*

The use of models will mostly rely on the methods described in Econometric Methods with Applications in Business and Economics by Heij et al. (2004). Most previous literature on nowcasting by the use of Google Trends refer to the methods used by Choi and Varian (2012). At the core of their work in the prediction of economic indicators, they use what is called an autoregressive moving-average model. The methodology section will go further into the exact use of techniques in this paper. Eventually, the goal of the research is to come up with a viable model that is applicable in real life and which will provide relevant and reliable information with regards to the Brazilian economy.

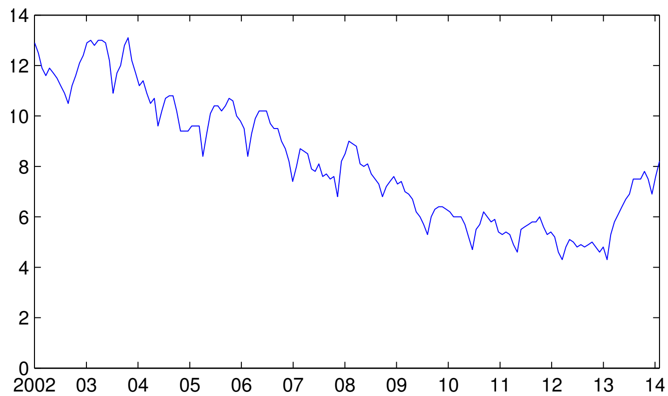


FIG. 2. The development of Brazil's unemployment rate (%) during the period January 2004 until February 2016

III. DATA

As a starting point, data regarding the Brazilian unemployment rate were obtained from the Brazilian Institute of Geography and Statistics, Instituto Brasileiro de Geografia e Estatstica (IBGE). The IBGE conducts a monthly employment survey (pesquisa mensal de emprego) which produces indicators regarding the country's labor force. The survey covers only the six largest metropolitan areas in Brazil (Recife, Salvador, Belo Horizonte, Rio de Janeiro, So Paulo and Porto Alegre), which together comprise roughly a quarter of the nation's population. It will be assumed that the unemployment rate in these cities is representative for the unemployment level in Brazil. The data is split between the areas, but total figures are also published, which shall be used in this research. To answer the research question of this paper, the published monthly "unemployment rate" will be utilized. Although unemployment rates are often seasonally adjusted to account for hiring and layoff patterns during the calendar year, the data will not be adjusted in order to reflect the real unemployment. Besides, it is especially interesting to analyze whether Google search volumes follow these seasonal patterns. To match the Google search data, the Brazilian unemployment rates from January 2004 to February 2016 will be considered. Summary statistics are given in Table I and are visualized in Figure 2.

TABLE I. Summary descriptive statistics Brazilian unemployment rate from January 2004 till February 2016

Obs.	**Mean**	**Median**	**Min.**	**Max.**	**Std.**
146	7.65	7.5	4.3†	13.1§	2.50

† (Dec., 2013)
§ (Apr., 2004)

The independent variables for the models were obtained using the free database "Google Trends", provided by Google Inc. By using this service, one is able to download the relative interest for particular Google search keywords over a specified range of time in customized geographic areas. It is important to note that keyword interest is measured in terms of index values, meaning that the obtained values have been normalized to an interest factor between a value of 0 and 100. Hence, the absolute search volume cannot be inferred from the data. In addition to data on separate keywords, Google compiles its own indexes regarding different themes, such as government affairs or economic indicators, by using a proprietary algorithm. Other than the keyword searches, the indexes reflect a general interest in these particular topics which allows it to take on negative values.

The search data can be downloaded on a weekly basis only, with specified beginning and start dates. However, from the Google trends website, there is a possibility for obtaining approximated monthly values from their graphs. For this research, a multitude of keyword interest series related to unemployment has been downloaded on both a weekly and monthly level. Next to using their absolute values, two types of transformations will be used to investigate potential non-linear relationships between the data and the unemployment rates. To begin with, first differences will be taken from the absolute level. This creates a set of values containing the changes in relative interest values from one month to another. Then, another variable will be extracted by seasonal adjustment of the search data. The seasonal adjustment eliminates patterns throughout the data and transforms undulations into steady trends.

The keywords have been chosen on the basis of theory and other research as laid out in previous sections. For a list of the investigated keywords, please refer to table II. Included in this table are the Google indexes on the themes "unemployment and social benefits" (U&SB) and "Job vacancies". A visual graph to reflect the relative search volumes is given in the appendix in Figure 3. More recent data is available, but has been truncated because the unemployment data of Brazil is available only until February 2016.

TABLE II. Description of used keywords

Keyword	**Description**
Décimo terceiro salário (DTS)	Social security programme
Empregos	Jobs
FGTS	Severance Indemnity Fund
INSS	National Institute for Social Security
Seguro desemprego	Employment Insurance
U&SB (index)	General interest in unemployment
Job vacancies (index)	General interest in job searches

IV. METHODOLOGY

Since the Google search data is provided in both weekly and monthly intervals, whereas the Brazilian unemployment rate is provided only in monthly intervals, there is a matching problem to which multiple solutions are possible. One can of course utilize the monthly data only, but this will inevitably result in a loss of information. Askitas & Zimmermann (2009) proposed to combine the first two weeks of a month and the last two weeks of the month, thereby approximately halving the total number of observations and creating two separate series. In this research paper, both approaches are investigated, albeit with a slight modification. That is, search data will be aggregated on a monthly level and on a semi-monthly level. For the semi-monthly data, only the first 15 days of the month will be examined, in order to investigate whether this data can be used to forecast the unemployment figure of the entire month. The search interest for the first 15 days is calculated by allocating the weekly data to their respective individual dates and taking the average of the first 15 days of the month. Next, the concept of the linear regression used in this paper will be explained. The unemployment rate at the end of a certain month is defined as coU_t. In addition, the search level of keyword k in month t is defined as $S_{k,t}$. $\Delta S_{k,t}$ and ΔU_t are defined as the change in the search level of the keyword k and in the unemployment rate between month t and month $t-1$ respectively. Finally, the coefficients are denoted by β_i, where β_0 represents the constant.

Two models will be investigated, the first being a standard linear model with a combination of search levels and first differences:

$$\Delta U_t = \beta_0 + \beta_1 S_{1,t} + \ldots + \beta_k S_{k,t} + \gamma_1 \Delta S_{1,t} + \ldots + \gamma_p \Delta S_{p,t} + \epsilon_t \quad (1)$$

In addition, it will be tested whether an autoregressive part as component of the first model will provide an improvement of the results:

$$\Delta U_t = \beta_0 + \beta_1 S_{1,t} + \ldots + \beta_k S_{k,t} + \gamma_1 \Delta S_{1,t} + \ldots + \gamma_p \Delta S_{p,t} + \xi \Delta U_{t-1} + \epsilon_t \quad (2)$$

Each model will be subjected to four different approaches of independent variable. First, note that $S_{k,t}$ can stand for both the monthly and the biweekly data. Also, the models will distinguish between mixed and unmixed choices of variables. The mixed data will use both seasonally adjusted and non adjusted data whereas unmixed will use non adjusted data only. This technique is adopted to examine whether it is necessary to include adjusted and non adjusted data of the same source to come to an effective model.

The first challenge into constructing an efficient model entails finding the right combination of keywords as independent variable. To avoid over-fitting the data by an excessive number of independent variables, it is the aim to find a model with the five most significant explanatory variables (excluding the constant). To determine the optimal combination of keywords, permutation selection will be utilized, by using the explained variance (R-squared) as indicator of model performance. A high R-squared with significant coefficients for the independent variables would be evidence for the existence of a correlation between Google search volumes and the Brazilian unemployment in the case of monthly data and signify predictive power when using biweekly data. The combinatorial regression will be performed in-sample over the period of January 2004 until December 2010 (period one). Once an optimal combination of keywords has been found, model performance will be evaluated using both in-sample (period one) and out-of-sample (period two) statistics.

Finally, to address hypothesis three, the models will be tested against two baseline scenarios. The first one using the change in the previous period as a current estimate for the change in unemployment. It is reasonable to assume that when unemployment increases/decreases in one month, it will follow the same direction in the next one.

Baseline method 1:

$$\Delta U_t = \Delta U_{t-1}$$

The second baseline method makes use of yearly seasonal patterns to establish an educated guess on current unemployment rate. It uses the exact same rate as 12 months earlier as an estimation for current figures.

Baseline method 2:

$$\Delta U_t = \Delta U_{t-12}$$

By establishing these baselines, root mean square errors (RMSE) and hit rates can be compared with the search-based models in order to examine whether they provide any predictive power. RMSE is a measurement that computes the average distance from the forecasts and reported rates and the hit rate tallies the number of times the forecast took the same direction (increase or decrease) as the reported rate.

TABLE III. Performance measures of different models

	In sample			Out of sample		2015	
	R^2	RMSE	Hit rate†	RMSE	Hit rate	RMSE	Hit rate
Monthly unmixed	0.63	0.29	0.83	0.46	0.68	**0.32**	**0.86****
Monthly mixed	**0.73**	**0.25**	**0.87**	0.32	0.76**	0.38	0.79**
Biweekly unmixed	0.62	0.30	0.80	0.42	0.66	0.36	0.79**
Biweekly mixed	0.70	0.27	0.82	0.31	**0.79****	0.45	**0.86****
Baseline 1‡	0.00	0.67	0.62	0.49	0.56	0.62	0.64
Baseline 2§	0.45	0.40	0.72	**0.28**	0.71	0.35	0.57

† The hit rate is defined as the fraction of correctly predicted directions of unemployment change; that is, a positively predicted change corresponds to a positive change in measured unemployment and a negatively predicted change to a negative measurement.
‡ Baseline 1 uses the previous change in the unemployment rate, such that $\Delta U_t = \Delta U_{t-1}$
§ Baseline 2 consists of the rate of change of the same month, one year ago, such that $\Delta U_t = \Delta U_{t-12}$
** signifies significantly better performance ($p < 0.01$) than baseline 2. Bold entries were the top performers in their respective periods.

V. RESULTS

The results of model one have been determined in four different specifications. The first specification uses the monthly search data with seasonally unadjusted variables ('monthly unmixed'). The second specification is based on the monthly search data with seasonally adjusted and unadjusted variables ('monthly mixed'). The third specification contains semi-monthly search data with seasonally unadjusted variables ('biweekly unmixed'). The fourth and last specification uses the semi-monthly search data with a combination of seasonal adjusted and unadjusted data ('biweekly mixed'). With regards to the first two hypotheses, it is clear from the OLS results that all search variables are highly correlated with the unemployment of Brazil. Hence, search behavior seems to exhibit patterns similar to the unemployment rate in the country. The power of the biweekly models is their ability to efficiently approximate the unemployment rate halfway through the month already. After stepwise regressions on the basis of permutations were performed, the autoregressive term postulated in model two of the methodology provided no significance in any specification. Hence, model two shall be disregarded in the further discussion of the results. The estimated parameters for the four scenarios are given in Table IV. Seasonally unadjusted variables are indicated by "(U)" and seasonally adjusted variables by "(A)". All five search data variables in each specification were found to be significant, with the exception of the first difference of the unemployment index in the biweekly mixed model. Graphs containing the forecast and actual values can be found in the appendix in Figures 4 and 5.

Judged by in-sample performance (Table III), it is clear that the monthly mixed model performs best on all metrics. The biweekly mixed also provides a higher in-sample variance explanation than the rest of the models.

When examining out-of-sample performance, it must first be noted that the second baseline (lagged by one year) has the lowest root mean squared error (RMSE). Hence, there seems to be a seasonal component in play. Moreover, the null hypothesis of equal RMSE is rejected for all other models (with $p < 0.01$), except for the biweekly mixed model (using a 5% significance level).[1] Subsequently, there is no reason to assume that the biweekly mixed model provides worse forecasts than baseline 2 in terms of RMSE out of sample. In addition, the biweekly mixed model also proves to behave significantly better when judged by its out-of-sample hit rate of 0.76, which measures the fraction of times the model was able to forecast the correct direction of the move (up or down). Using a two-proportion z-test[2], all models were found to provide better forecasts than baseline model 1 in terms of hit rates. Again, when looking at hit rates, only the monthly mixed and biweekly mixed models were able to provide statistically significantly better results than baseline model 2.

In addition to the examination of the entire out-of-sample period, it is of interest to look at the period January 2015 until February 2016, since the unemployment data started to exhibit a deviation from its regular pattern as the crisis kicked in. Hence, baseline methods relying on historical data prove to be less informative during these periods. Indeed, it is observed that all models perform significantly better in terms of hit rate during this period than baseline model 2. A test of difference in RMSE between the models and baseline 2 leads to no rejection of the null hypothesis of equal RMSE, but these results must be interpreted cautiously because of the small sample size during this period ($n = 14$). Striking is the stellar performance of the monthly unmixed

[1] This test is based on the assumption that $(MSE_1 - MSE_2) \sim N\left(0, s_1^2/N_1 + s_2^2/N_2\right)$ and hence must be exercised with caution because of its strong assumptions. However, since the sample size is 62, it is reasonable to argue that the Central Limit Theorem applies in this case.

[2] $z = (\hat{p}_1 - \hat{p}_2)/\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}$ with $\hat{p} = (\hat{p}_1 n_1 + \hat{p}_2 n_2)/(n_1 + n_2)$

model, which provided poor estimates during the rest of the out-of-sample period.

Summarizing, it can be concluded that, when examining both in-sample and out-of-sample performance, all models exhibit significant predictive power. However, especially the biweekly mixed method seems to improve current baseline methods on all metrics, especially when judged by its hit rate. Hence, there is no evidence to reject the third hypothesis of the paper. Furthermore, it appears that the mixed specifications are generally superior to the unmixed ones. From Table IV, one can observe that the parameters of the keywords “Empregos” and “Décimo terceiro salário” and their seasonally adjusted versions to (partly) cancel each other out. The remainder of this equation is the seasonal patterns that appears to be a strong predictor of unemployment. Therefore, it can be concluded that a large part of the explained variation is the prediction of seasonal patterns. Still, this outcome is very interesting, as it provides strong evidence that Google searches indeed directly correlate with real unemployment. Finally, their predictive power easily exceeds the forecasting of seasonal patterns, because the methods were very successful into forecasting correct estimations in 2015 when most seasonal pattern were disrupted due to worsening economic conditions. For these reasons, we propose the mixed monthly and mixed biweekly model as explanatory and predictive models for the Brazilian unemployment. The accuracy of their forecasts based on the parameters in Table IV is visually represented by Figure 6 in the appendix.

VI. CONCLUSION

The era of big data has just begun, and once more we can catch a glimpse of its immense power. As proven by this paper, it is possible to predict the direction of the monthly Brazilian unemployment rate halfway through the month with more than 80 percent accuracy, solely based on Google search volumes. Surprisingly, the results found in this paper do not significantly differ from previous research in western nations. The proposed models decreased the root mean squared error, which quantifies the average distance from the forecast to the actual rates, from 0.35 in baseline methods based on seasonal patterns to 0.29, improving performance by approximately 18 percent. It is a worthy addition to this line of research to conclude that also in a developing country such as Brazil, predictive Google-based models are achievable. The implications of models containing such predictive power entail an efficient and simple way of measuring the economic climate of Brazil. Further research could strive to optimize the techniques and models as prescribed in this paper by examining a larger pool of potential keywords or improved models to extract more information from the data. For instance, the models used in this paper were based on data from 2004 to 2011, but as time passes, one could increase the time range and re-calibrate the models. In addition, the possibility of adaptive calibration could be further investigated. Moreover, the methods could be altered to facilitate forecasting multiple months ahead.

Eventually, because these models proved to be especially effective during the economic downturn in 2015, these techniques could be extended into useful tools to create early warning signals for volatile unemployment rates and recessions. In any case, it is clear that Google search data reflects real economic behavior and could be a valuable device in the future of econometrics.

VII. REFERENCES

Askitas, N., & Zimmermann, K.F. (2009). Google econometrics and unemployment forecasting. *German Council for Social and Economic Data (RatSWD) Research Notes, 41.*

Bughin, J.R. (2011). ’Nowcasting’ the Belgian Economy. *Available at SSRN 1903791.*

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

d’Amuri, F., & Marcucci, J. (2012). The predictive power of Google searches in forecasting unemployment. *Bank of Italy Temi di Discussione (Working Paper), 891.*

Engle, R.F. & Granger, C.W.J. (1987): Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica, 55*, 251-276.

Guzman G. (2011) Internet search behavior as an economic forecasting tool: the case of inflation expectations. *The Journal of Economic and Social Measurement.*

Heij, C., de Boer, P., Franses, P.H., Kloek, T., & van Dijk, H.K. (2004). Econometric Methods with Applications in Business and Economics. *Oxford university press.*

Koop, G., & Onorante, L. (2013). Macroeconomic nowcasting using Google probabilities. University of Strathclyde.

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

Reuters (2014). Brazil’s jobless rate revised up under new methodology. *Thomson Reuters.*

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

Wu, L., & Brynjolfsson, E. (2009). The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales.

TABLE IV. Estimated model parameters for calibration period January 2004 - December 2010

	Monthly unmixed	**Monthly mixed**	**Biweekly unmixed**	**Biweekly mixed**
Constant	−0.8171	−0.0557	0.5321*	−0.1178
	(0.6025)	(0.2234)	(0.2652)	(0.2267)
Empregos (U)	0.0185**	0.0520**		0.0581**
	(0.0066)	(0.0041)		(0.0046)
DTS (U)(D)	0.0079**	0.0139**		0.0078**
	(0.0023)	(0.0025)		(0.0025)
Empregos (A)		−0.0521**		−0.0568**
		(0.0059)		(0.0067)
DTS (A)(D)		−0.0070**		
		(0.0026)		
Vacancy Index (U)(D)		0.0124**	−0.0298**	
		(0.0043)	(0.0071)	
Empregos (U)(D)	0.0081*		0.0325**	
	(0.0039)		(0.0068)	
Vacancy Index (U)	0.0188*		0.0412**	
	(0.0082)		(0.0071)	
FGTS (U)			0.0412*	
			(0.0063)	
Unemp. Index (U)(D)			−0.0065*	−0.0050
			(0.0029)	(0.0026)
Empregos (A)(D)				0.0460**
				(0.0152)
Unemp. Index (U)	0.0095**			
	(0.0026)			

* signifies $p < 0.05$, ** signifies $p < 0.01$. (U) signifies seasonally unadjusted data, (A) signifies seasonally adjusted data and (D) signifies the first level difference of a series.

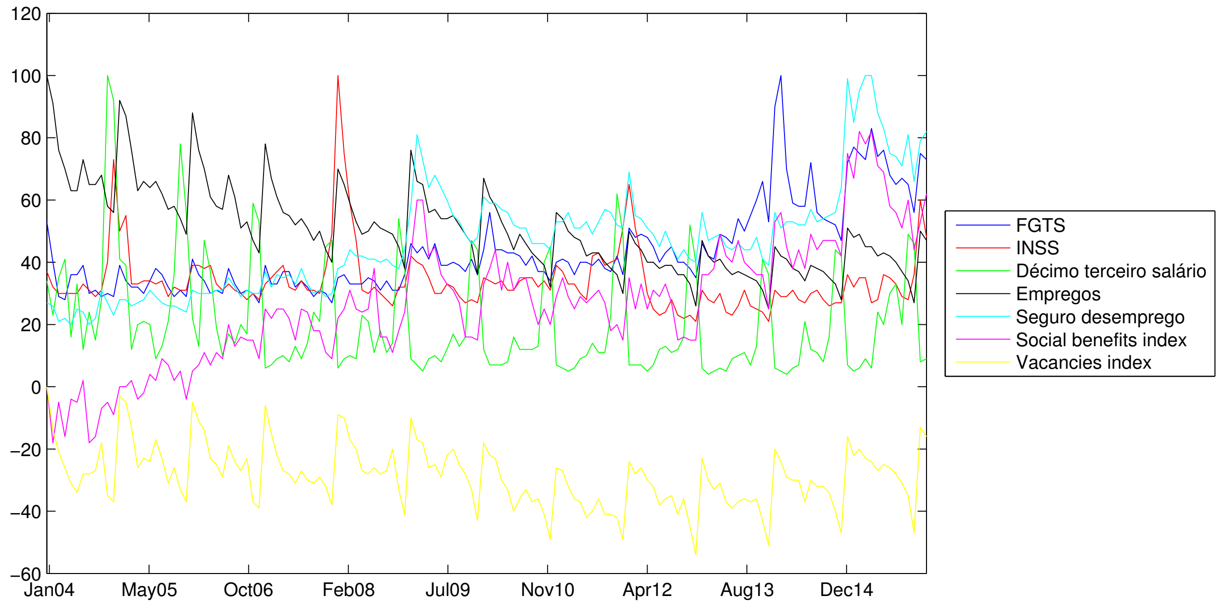
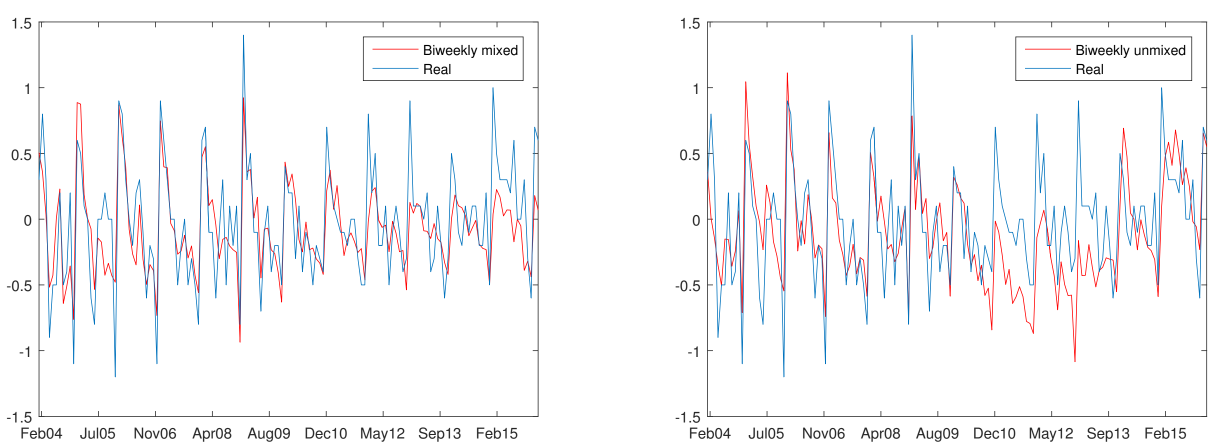


FIG. 3. Google search volume for the investigated keywords during the period January 2004 until February 2016

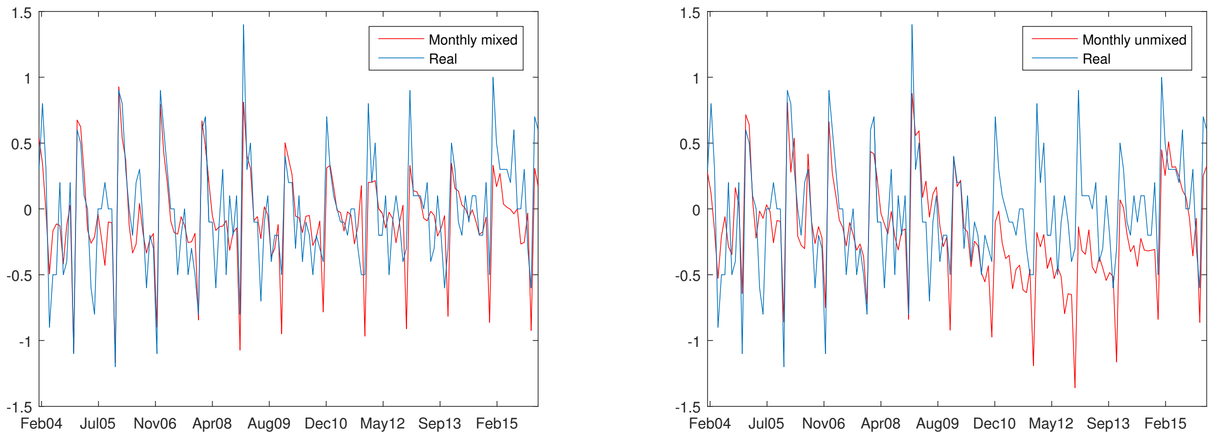
FIG. 4. Biweekly forecast versus actual unemployment rate change



(a) Mixed

(b) Unmixed

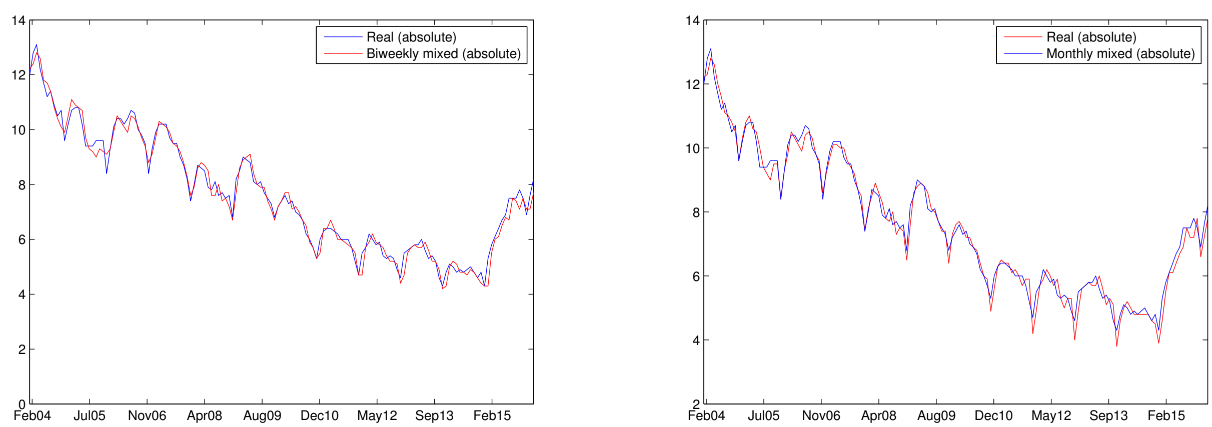
FIG. 5. Monthly unmixed forecast versus actual unemployment rate change



(a) Mixed

(b) Unmixed

FIG. 6. Mixed model forecasts versus actual unemployment rate level



(a) Mixed biweekly

(b) Mixed monthly