



**Procedia of
Economics and Business**

Administration

ISSN: 2392 – 8174,

ISSN-L: 2392 – 8166

Available online at

www.icesba.eu

International Conference on

Economic Sciences and Business Administration



24-25 October 2014, Bucharest, Romania

Nowcasting Credit Demand in Turkey with Google Trends Data

Ömer ZEYBEK¹, Erginbay UĞURLU²

¹ING Bank Turkey Analytic CRM Dept., Maslak, İstanbul, 34357, Turkey

Tel: +902123351340, Fax: +902123351000, Email:omer.zeybek@ingbank.com.tr

Abstract: Age of Big Data and internet has brought variety of opportunities for social researchers on identifying on-going social trends instantly. As internet user base grew exponentially, major internet content search companies have begun to offer data mining products which could extract attitude of on-going trends and identify new trends on web as well. Since 2009, as a pioneer on these web analytics solutions Google has launched Google Trends service, which enables to researchers to examine change of trend on specific keywords. We use weekly Google Trends Index of “General Purpose Loan” (GT) and total out-standing volume of Turkish banking system from the data period of first week of March 2011 to second week of September 2014. In this paper we test whether the Google Analytics search index series can be used as a consistent forecaster of national general purpose loan (GPL) demand in Turkey. We show how to use search engine data to forecast Turkish GPL demand. The results show that Google search query data is successful at nowcasting GPL demand.

Keywords: Nowcasting web analytics, forecasting, general purpose loan.

JEL classification: C53

1. Introduction

Public and private data providers periodically release indicators on level of economic and financial activities for various sectors. However, due to difficulties in data collection and statistical calculation procedures, announcement of figures could be lagged for a period of

¹ ING Bank Turkey Analytic CRM Dept., Maslak, İstanbul, 34357, Turkey, Tel: +902123351340, Fax: +902123351000, Email:omer.zeybek@ingbank.com.tr

² Assoc Prof. Hitit University,FEAS, Department of Economics, Çorum, 19030, Turkey, Tel: +903642257700, Tel: +903642257710, Fax: +902123351000, e-mail: erginbayugurlu@hitit.edu.tr

time. In addition to this, data such as GDP or IP are subsequently revised due to post publication collection of updated observations. Unless short-dated delays up to three months are tolerable for economic policy and trend analysis, on marketing strategy decisions and monetary interventions of central banks side, these delays could cause very serious short-comings in decision making processes. So, it became a common practice for researchers to forecast current conditions when the real data is delayed. This process, the forecasting of present time could be referred as now casting which is a contraction of now and forecasting.

[Hendry et al.,2013] defines now-casting as “any procedure that uses additional information when producing contemptuous aggregate data, beyond just cumulating observed disaggregates as now-casting”. Beside that definition, [Banbura et al.,2010] generalizes time interval of forecasts and defines “now” as near future, now and near past. Originally as a procedure generally used by meteorologist, ‘the now casting’ could be defined as predicting today’s or near future’s weather conditions, based on extensive past meteorology data. But considering economics side, it’s a fact that quality of recent data is more suspicious compared to definite meteorology data. For example GDP or IP indices employed as consistent estimators, could be updated after initial publication of data.

However, the idea made now-casting important was spill over of monetary economics and government’s tendency to react spontaneous actions on any disequilibrium occurred in market environment. As stated in [Evans,2005] although, individual customers or firms are aware of their contemptuous activities, information about current state of economic activity is widely dispersed among economic agents. So this phenomenon causes disintegration between economic reality and already published data, thus markets and trading environment falls short on information on current state of economics. For example, as a major indicator of progress in economic activity the GDP indices are the most vulnerable data releases to this adverse effect.

Accordingly, when we examine recent works on now-casting dominance of analysis on GDP growth draws attention. While now-casting, current-quarter GDP a typical method would be using small bridge equations containing monthly data to forecast current quarterly GDP via these small models. Consequently, [Baffigi et al.,2004] and [Runstler and Sedillot,2003] could be counted as an example bridge equation method.

Until recently predicting the present has been considered as a challenge for academic researchers working on specific issues like listed above or central bankers and other government agencies as a extension of their duty of managing level of economic activity. But in this paper we claim that contemporaneous forecasting of credit demand in high frequency could have crucial importance on operations of banking industry. We assume that by estimating weekly credit demand of retail side, banks could regulate their loanable funds stock more efficiently and place their campaigns as soon as it’s necessary. However it should be noted that a high frequency forecasting on weekly basis could require more specific econometric methods such as mixed model estimations or appliance of high level time series methods. On the other hand, because our goal in this paper is to familiarize readers on now-casting weekly outstanding credit figures using web searching trends we didn’t prefer complicated methods in this paper. However, we believe that using such valuable data in high frequency future researcher could create more advanced models.

The remainder of this article is organized as follows. Section 2 explains the outstanding credit conditions in Turkey. Section 3 investigates the previous literature and section 4 presents results of empirical application. Section 5 summarizes the paper.

2. Outstanding Credit Conditions in Turkey

As briefly introduced above, in this paper we investigate a rather different side of economic activity, personal finance demand of individuals in Turkey. As an emerging country, majority of the population comprises of youths with increasing disposable income, consumer spending in Turkey increases steadily for almost a decade³. Consequently, private income's contribution to 4.0% YoY increase in Turkish GDP came at 3.1% level (translating about 80% of all increase in GDP is explained by domestic demand) by the end of 2013. Moreover, total amount of personal finance in Turkey has reached approx. US\$250bn by the end of same year. The banking regulator Banking Regulatory and Supervision Agency of Turkey (BDDK)'s figures presents that nearly half of this figure (US\$141bn) is concentrated in general purpose loans⁴ segment, while rest of the personal debt is listed under mortgages, vehicle and overdraft/credit card loans. However because Turkish private consumption is largely consists of imported consumer products and local savings standing at very low-levels, the funds needed to stimulate this growing consumer economy needs to be financed from abroad. Consequently financing huge volume of funds from abroad creates extensive imbalances in Turkey. In order to regulate this unbalanced growth, from 2011 to 2013 government utilised different kinds of macro-prudential measures to down-down loan growth which reached to 40% YoY by 2010 to more moderate %15 levels in 2013.

As [Basci,2006] stated, the rapid growth in emerging countries bring problem of responding credit booms on policymakers table. On the other hand, one should not be miss out that any measure against credit boom could only be effective when the intervention comes at the right time, after annual credit growth reaching high levels and household debts hitting the top it would be extremely challenging to rebalance economic activity. On the contrary, when we leave macro ground and focus on micro company view; if you are a bank with a large stockpile of loanable funds it's very important to plan supply of this capital. Offering funds when the credit conditions are dull would face bank with unfavourable lending conditions. That's why monitoring real time credit demand has crucial importance for the banks either.

BDDK is responsible from publication of banking sector statistics. Each week the agency publishes weekly banking statistics with a one week of lag (as of 12nd Sept. We are able to reach 5th Sept's data). Regarding economic indicators, one week lag could be counted as an acceptable delay but if you are central banker commissioned with daily monitoring of credit conditions or a sales manager of a retail bank responsible from weekly planning of consumer loan campaigns, a 7 day of delay could cause some important problems. Respectively, in following sections we will try to build a reliable model to foresee level of outstanding credit volume both for central bankers and also both for commercial bankers. While seeking a reasonable proxy for estimating current outstanding GPL volume, as keeping in mind that this volume translates to current outstanding GPL credit demand, we preferred using web searches as a proxy of credit demand.

³ TURKSTAT <http://www.turkstat.gov.tr/HbGetirHTML.do?id=16193>

⁴ <http://ebulten.bddk.org.tr/haftalikbulten/index.aspx>

3. Literature Review

After emerge of internet as main source of information in late 1990s, first studies on the relation between web searches and macro-economic variables have started by early 2000s. [Kuhn and Mikal, 2002] argue that does web search has an positive impact on unemployment period of U.S. job seekers, and they were unable to find any significant statistical relation between being a job seeker using web search could shorten period of being unemployed. On the other hand, [Ettredge et al.,2005] was the first study which aims to now-cast unemployment rate in US, especially a reasonable time before the BLS report of unemployment subsequent year. However due to lack of data availability they prefer to make an indirect forecast of unemployment rate via using number of jobseekers on internet. Consequently, Google's launch of Google Insights in 2007 (Google Trends after 2012) enables researchers to employ strong proxy variable for forecasting any trend. [Choi and Varian,2009a,b] described how can Google Trends helps to predict several economic indicators. However widespread recognition of web search data as an indicator of present came up with 2009's H1N1 Flu Pandemic. Flu Trends a service by Google has already been launched in 2009 and a model behind the system was continuously scoring a flu activity probability of a specific region by using 50 million queries entered search engine every day. [Polgreen et al.,2008], [Ray and Brownstein, 2013] and [Cook et al. , 2011] has confirmed Google's claim that search data is capable for estimating H1N1 pandemic activity which out broke in 2009. When we checked economics area, [Guzman, 2011] examined Google data as an estimator of Israeli inflation. As we stated earlier because now casting current economic activity is central banker's main challenge, we could also see several works by central bankers aiming to estimate current activity, such as [Swallow and Labbe, 2011] studied on now-casting car purchases in Chile.

Although studies on prediction power of "Google Trends" have become popular around the world, studies on Turkish case remained limited until now. As far as we know the only study [Chadwick and Sengul, 2012] examined forecasting unemployment rate in Turkey which is announced with a 3 months of lag.

4. Empirical Application

Google Trends is a service which provides time series index of volume of searches for a specific keyword at predefined time interval. Introduced in 2007, Google Trends is able to draw, web search data since 1 Jan 2004 to present. Data are delivered in weekly frequency and users served with an index normalized by highest value, fluctuating between 0-100. Due to normalization procedure the date with highest score represents with 100 value, while 0 means there are not enough searches during that particular week. It should be noted that these values are calculating based on a specific keyword provided by the user.

Although it's possible to receive index values on country level, user could also reach regional volume indexes. Finally, as a last remark, due to performance issues Google Trends data compute by using a sampling method and that's why results could vary day to day. It's clear that important characteristic of the data could cause measurement error in the model, but [Swallow and Labbe, 2011] test the variation of the indices by downloading index data on a keyword 50 times in 50 day and they stated that measurement error is as source of concern because it weakens information contents of Google data thus it makes more difficult to reject the null hypothesis tested. Flowingly, because researchers have no access to raw data, they suggest Google to provide cleaner data available in the future.

We collected data from both Google Trends Index and BDDK's Weekly Outstanding General Purpose Loan figures. Google Trends Index is calculated based on key word "İhtiyaç Kredisi" which means "General Purpose Loan (GPL)" in Turkish that is downloaded from Google Trends web page at 20 September 2014. Because of sampling procedure on data which is argued above, the exact timing of data collection is crucial in analysis. Regarding credit demand indicator, the total out-standing GPL volume of Turkish Banking System is downloaded from BDDK's statistics web page in weekly frequency. It should be noted that a flow data for credit stock is only available for quarterly frequency, so for a high frequency analysis such like this one, researchers have to use outstanding figures. Each row constitute volume of GPLs in TRY by the end of week while Google Trends data prints weekly search performance for specific keyword.

We use weekly data start from first week of March 2011 to second week of September 2014 thus we have 185 observations. Figure 1 shows the time series graphs of Google Trends Index of "General Purpose Loan" (GT) and total out-standing volume of Turkish Banking System (BK).

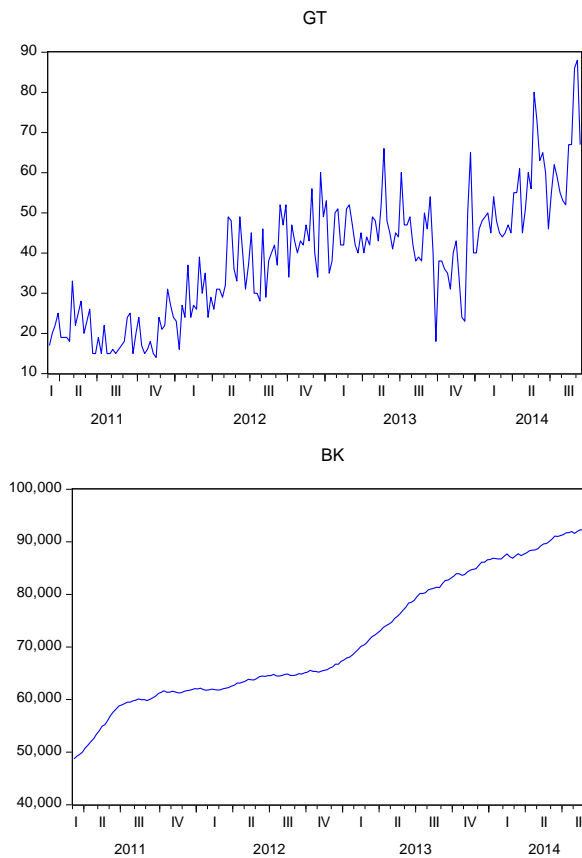


Figure 1: Graph of GT and BK Series

Source: Authors

[Choi and Varian,2009] used a simple AR process to forecast time series. So they defined y_t the log transformation of the observation at time t. Because they used monthly data, AR-1 model is defined as follows:

$$y_t = b_1 y_{t-1} - b_{12} y_{t-12} + e_t \quad (1)$$

We use weekly data hence we take a first and fourth lag of the log of the BK (LBK) to show effect of first week and first month. To estimate the model we use in sample period, 3 May 2011- 08 August 2014 period. The estimated model is below:

$$LBK_t = \beta_0 + \beta_1 LBK_{t-1} + \beta_2 LBK_{t-4} + \varepsilon_t \quad (2)$$

Table 1 shows the result of the estimated model of equation 2. Except intercept all coefficients are statistically significant in 1% level and the model is statistically significant according to F statistics in 1% level.

Table 1: Estimation Results of Baseline Model

Dependent Variable: LBK		
Variable	Coefficient	Std. Error
C	0.025893	0.018953
LBK(-1)	1.217201***	0.026902
LBK(-4)	-0.219420***	0.026459
R-squared: 0.999527		
Adjusted R-squared: 0.999521		
F-statistic: 182597.6***		

*** indicates significance at the 1% level

Source: Authors Calculation

To see whether google trend (GT) variable has impact or not on BT, it is added to model. Table 2 shows the model with GT variable. However, after GT variable added to model significance of the intercept and value of R-squared increased, the significance of the GT variable is low (10%). Regarding to this model we can say that the google trend data can be used as a forecaster of BK series with positive effect on it.

Table 2: Estimation Results of Model with Google Trend Data

Dependent Variable: LBK		
Variable	Coefficient	Std. Error
C	0.055034**	0.025749
LBK(-1)	1.212179***	0.026707
LBK(-4)	-0.217158***	0.026391
GT	4.44X10 ⁻⁵ *	2.67X10 ⁻⁵ *
R-squared: 0.999534		
Adjusted R-squared: 0.999526		
F-statistic: 122972.9***		

*,** and *** indicate significance at the 10%,5% and 1% level.

Source: Authors Calculation

After the forecaster variable is defined other important question would be whether the GT variable improves forecasting performance or not. To check this, we use forecasting evaluation criteria to compare the forecasting⁵ performance of these two models. The used criteria are Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), Root Mean Squared error (RMSE) and Theil Inequality Coefficient. The model which has smaller MAE, MAPE and RMSE has a better forecasting performance comparing to other one. On the other hand the model whose, Theil inequality coefficient is close to zero would has better forecasting performance than the other one.

⁵ We use static forecast.

Table 3: Forecasting Performance of Two Models

	Baseline Model	Model with GT
MAE	0.002823	0.002776
MAPE	0.025294	0.024872
RMSE	0.003487	0.003459
Theil Inequality Coefficient	0.000155	0.000155

Source: Authors Calculation

Table 3 shows that all criteria of model with google trend data have smaller values than baseline model. Thus we can say that google trends variable has a positive impact on forecasting performance of total out-standing volume of Turkish banking system.

The final step of the empirical application is calculating out of sample forecast performance for the 8/15/2014- 9/12/2014 period. We employ the following equation for the calculations.

$$\widehat{LBK}_t = \beta_0 + \beta_1 LBK_{t-1} - \beta_2 LBK_{t-4} + \beta_3 GT_t \quad (3)$$

Forecasted values were calculated using actual data of the variables. To see the performance of the model we calculate proportion of absolute deviation from the actual values of forecasts.

Table 4: Forecasting Performance of Out of Sample Forecast

Term	\widehat{LBK}_t	LBK_t	\widehat{BK}_t	BK_t	$ (\widehat{BK}_t - BK_t)/BK_t $
8/15/2014	11.43098	11.43245	92132.22	92268	0.001472
8/22/2014	11.43431	11.43243	92439.53	92266	0.001881
8/29/2014	11.43592	11.43383	92588.48	92395	0.002094
9/05/2014	11.43681	11.4384	92670.92	92818	0.001585
9/12/2014	11.44074	11.44503	93035.83	93436	0.004283

Source: Authors Calculation

Table 4 shows the absolute deviation. Maximum value of the deviation is 0.004283 and the mean of the deviation is 0.002263. We have also had success with out of sample period.

Conclusion

This paper provides empirical evidence for using Google Insights Search data to now cast total out-standing volume of personal finance among banking system in Turkey. We use weekly Outstanding General Purpose Loan and Google Trends Index data for the is calculated based on key word "İhtiyaç Kredisi" which means "General Purpose Loan (GPL)" for the 3 May 2011- 08 August 2014 period.

Our findings show that Google Trends data has significant explanatory power on forecasting of credit demand variable. So it could be employed to monitor nearly - real time developments in personal credit demand. However it is obvious that using lagged variable of first week and first month are very naïve forecasting method and indigenous lagged variables could absorb much of exogenous variables forecasting performance. However because our goal in this paper was introducing google trends as a capable forecasting instrument in high frequency economic activity data we didn't employed more complex forecasting methods and also we didn't test whether GT is the perfect forecasting instrument among other exogenous variables. As we proved that the series has a significant explanatory power these question would be subject of following studies.

References

- [1] Baffigi, A., R. Golinelli and G. Parigi (2004), "Bridge Models to Forecast the Euro Area GDP" International Journal of Forecasting, forthcoming
- [2] Banbura, Marta & Giannone, Domenico & Reichlin, Lucrezia, 2010. "Nowcasting," CEPR Discussion Papers 7883, C.E.P.R. Discussion Papers
- [3] Başı, E 2006,. "Credit growth in Turkey: drivers and challenges." BIS Papers, 28 The banking system in emerging economies: how much progress has been made (2006): 363-375.
- [4] Brownstein, John S., Clark C. and Lawrence C, 2009 "Digital disease detection—harnessing the Web for public health surveillance." New England Journal of Medicine 360.21 (2009): 2153-2157.
- [5] Carrière-Swallow, Yan, and Felipe Labbé., 2011 "Nowcasting with Google Trends in an emerging market." Journal of Forecasting 32.4: 289-298.
- [6] Chadwick, M. G., & Sengul, G. 2012. Nowcasting unemployment rate in Turkey: Let's ask Google. Central Bank of the Republic of Turkey, Ankara.
- [7] Choi H. Varian H. 2009b Predicting initial Claims for unemployment benefits, Economics Research Group, Google
- [8] Cook S, Conrad C, Fowlkes AL, Mohebbi MH (2011) Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic. PLoS ONE 6(8): e23610. doi:10.1371/journal.pone.0023610
- [9] Etterdge, M., Gerdes, J. and Karuga, G. (2005) "Using web based search data to predict macro economic statistics' Communications of the ACM, 48(11), 87 -92.
- [10] Euro Surveillance : Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin [2009, 14(44):429-433]
- [11] Evans, Martin D.D., and Richard K. Lyons. 2004a. "A New Micro Model of Exchange Rate Dynamics." NBER Working Paper No.10379 (March).
- [12] Evans, Martin D.D., 2005 "Where Are We Now? Real-Time Estimates of the Macroeconomy", International Journal of Central Banking, Vol.1, No.2, September.
- [13] Guzman, G., 2011 Internet search behavior as an economic forecasting tool: The case of inflation expectations. Journal of economic and social measurement, 36(3), 119-167.
- [14] Hendry F, Castle L, Kitov I. Forecasting and Nowcasting Macroeconomic Variables : A Methodological Overview, Oxford University Department of Economics Series Working Papers, No.674
- [15] Kuhn, Peter; Skuterud, Mikal (2002) : Internet Job Search and Unemployment Durations, IZA Discussion paper series, No. 613
Choi H. Varian H. 2009a Predicting the present with Google Trends, Economic Research Group, Google
- [16] Polgreen, P. M., Chen, Y., Pennock, D. M. & Forrest, N. D, 2008 Using internet searches for influenza surveillance. Clinical Infectious Diseases 47, 1443–1448).
- [17] Rünstler, Gerhard & Sédillot, Franck, 2003. "Short-term estimates of euro area real GDP by means of monthly data," Working Paper Series 0276, European Central Bank.
- [18] Wilson N, Mason K, Tobias M, Peacey M, Huang QS, Baker M Department of Public Health, University of Otago, Wellington, New Zealand. nick.wilson@otago.ac.nz