Google as a tool for nowcasting household consumption: estimations on Hungarian data

Istvan Janos Toth[+], Miklós Hajdu[*]

Institute for Eonomic and Enterprise Research - HCCI

Abstract

In this paper we attempt to examine the applicability of Google Insights for Search in terms of household consumption, retail sale trends, and the nowcasting of car sales in Hungary. There have been several successful attempts to use Google data in a similar way, usually in connection with U.S. economic indicators. Our goal, however, is an analysis of consumption indicators on the basis of Internet search in a country where Internet penetration is lagging behind that of the U.S. and Western Europe.

The results show that Google can also be a useful tool for nowcasting consumption in a country where Internet use is at a lower level compared to developed countries. For retail sales, car sales, and household consumption, the extended models containing Google-based calculated indicators are more effective in nowcasting economic time series.

Keywords: Google, nowcast, Hungary, household consumption, car sales, retail trade

JEL Classification: D12, E21, E27

+ senior research fellow IE HAS: email : tothij@econ.core.hu

** research fellow IEER HCCI: email : miklos.hajdu@gvi.hu*

Introduction

In this paper, we attempt to examine the application of Google Insights for Search (GIS) in terms of household consumption, the development of retail sales, and the short-term forecasts (nowcast) of car sales in Hungary. In our endeavour we track the works of Nikolaos Askitas and Klaus F. Zimmermann (2009), Choi and Hal Varian Hyunyong (2009), Torsten Schmidt and Simeon Vose (2009a, 2009b) and Kholodilin et al. (2010) in order to explore the potential of GIS in a country where Internet use is less prevalent than in Western Europe or the USA. Accordingly, content available online has a much narrower consumer audience in Hungary; both producers and service companies also use only a fraction of the potential offered by the Internet. The question we would like to answer is in such an environment would we nevertheless get additional information if we use Google tools when we estimate consumption.

The first part of this paper summarizes, a priori, why Google would be useful instead for nowcasting and why it may be less suitable for forecasting. Subsequently, we will look into Hungarian and international data regarding Internet penetration and describe the method by which GIS can be used to create a chronological set of search indicators.

In the following section, this paper describes the use of models, after which there will be a discussion of the empirical results. At the conclusion of this study, we will draw an evaluation from the results.

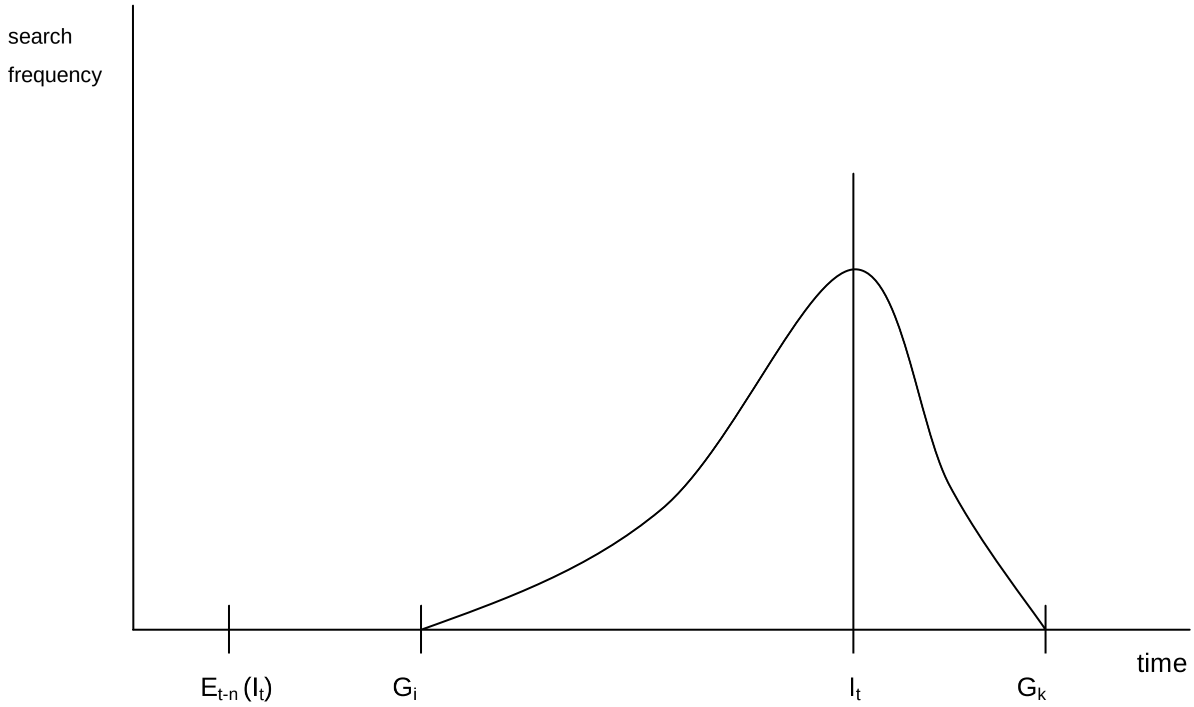
Estimates with the help of Google: nowcast or forecast?

Analysis of economic processes based on information provided by Google is founded on the assumption of whether the expected activity of consumers or entrepreneurs (product or service purchase) is present or preceded by an Internet search procedure: the search for product or service of a given economic actor or for a manufacturer (1) or the procedure of purchasing through the Internet, including the use of Google (2). Of course, not every Google search ends in a transaction - however, a positive relationship, *a priori*, between the two factors can be assumed. The association between a search and a transaction can be put down to a stochastic process.

This prediction model is fundamentally different from the expectations of economic agents and that of business tendency surveys. The latter occurs at t followed by I an economic decision (I_t) able to predict the economic player for this $t\text{-}n$ time surveyed intentions and based on expectations ($E_{t-n}(I_t)$). Google search results are based on usability *a priori*, but on the assumption that before making a purchase a consumer starts a search on Google (G_i), which falls within a maximum value directly before the economic transaction (purchase) period I_t, as well as after the purchase (G_k). Whether this latter component can be still considered a search-related activity of the purchase (i.e., the purchased product / service related information) is negligible, and is soon done after the purchase is over G_k. (see Figure 1). We do not know for certain, but it's intuitively obvious, that the intention for an economic transaction should take place first – and in this way is soon observable – before the actual search procedure for the expected transaction takes place ($t\text{-}n < i$). On the other hand, economic transactions associated with non-sense intentions analyse retrospectively whether it has actually occurred. Consequently, a survey-based analysis is more appropriate for forecasts, while results based on Google may be more useful for nowcasts.

It's another question when we take into consideration the preparation of the two analytical methods. Along these lines a Google based study has significant benefits since after the end of a given period of time (week, month, quarter) it can be prepared and performance verified in only a few days, while a traditional questionnaire-based surveys needs 2-3 weeks for preparation and all the real indicators of the transaction for the given period (e.g., a product sales, retail sales, consumption index monthly data, e.t.c.) is 30-40 days after a given period (month, quarter).

Figure 1 Google search and expected / intended behaviour of economic agents



In line with the abovementioned theory, Google does not replace traditional methods of preparing forecasts (e.g., the expected and/or intended behaviour of economic agents, and the expected results of a questionnaire survey), since its use is more appropriate for nowcasts rather than forecatsts. On the other hand, the opportunities offered by Google should be addressed because it provides very quick access to information.

In recent years there has been a proliferation in international literature of the number of research analysts studying the use of Google to estimate economic trends.

Askitas and Zimmermann (2009) examined the relationship between the German unemployment rate and the data series from certain keywords using Google Insights for Search. During their work, they were interested in trying to get search results for job search activity and other employment related terms. Their results show that the GIS is a promising tool in estimating the unemployment rate, despite the fact that the bias generated by changes to employment policy was hard to handle.

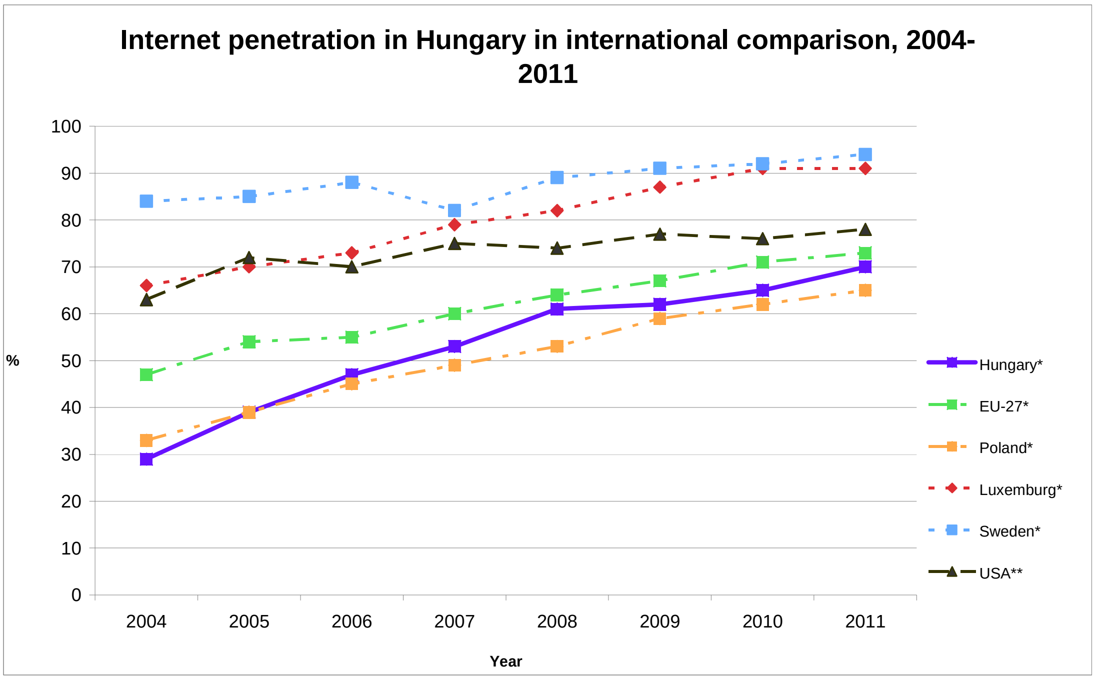
Varian and Choi (2009) were among the first to ask the question: "Can Google queries help predict economic activity?" By way of an answer, they tried to use Google Trends, a Google application like Google Insights for Search using a similar data series from the USA to tweak automotive, tourism and housing market data forecasts. The attempt was successful, the search data in all three sectors contributed significantly to improving the estimates.

Schmidt and Vose (2009a and 2009b) attempted to predict private consumption using Google Trends data series in the U.S. In their work, they compared the predictive ability of the search model and data lines they generated with two survey-based indicators (University of Michigan's Consumer Sentiment Index and the Conference Board's Consumer Confidence Index). The results show that the indicator calculated based on data from Google not only successfully predicts trends in consumption, but also is far more superior to that of the survey-based indicators.

Kholodilin et al (2010) tried to nowcast U.S. monthly private consumption with the help of a data series recorded from Google search activity. In their study they compared some survey-based indicators and financial data nowcasting models with the performance models built and supplemented with data from Google. According to his findings traditional data-based nowcasts can be further enhanced with Google data.

The success and accuracy of consumption forecasts based on Internet search trends has a major influence on the prevalence of Internet use. In addition to the potential demand-side characteristics of Internet use, the supply side factors of the Internet are also important in this regard. The next few paragraphs briefly describe the characteristics of Internet penetration in Hungary supplemented by a brief international outlook.

Figure 2 Internet penetration in Hungary in international comparison, 2004-2011



**Source: Eurostat*

***Source: Pew Internet & American Life Project*

According to Eurostat data, in 2011 the proportion of Internet users to the total Hungarian population was at about 70%. The data suggest that compared to the 29% in 2004 a continuous increase could be observed in Internet use. This rising trend of the proportion of Internet users in 2011 is just below the European Union's 27 member state average of 73%, whereas in previous years there was a serious difference of more than 10%. However, Hungary is still lagging behind compared to the Benelux and Scandinavian countries, where the proportion of Internet users is around 90% of the total population. In addition, among the post-socialist countries Hungary belongs to the middle range in terms of Internet use, ahead of some countries such as Poland. In the U.S., meanwhile, data from the Pew Internet & American Life Project estimates that 78% of the adult population used the Internet in 2011. This rate has been over 70% since 2005, but in 2004, it was only 63%.

We should note that the U.S. has a more limited definition of population than Eurostat, along with more stringent conditions of what constitutes an Internet user. Hence, the data is not comparable with each other, but nevertheless illustrating the differences between them is interesting (see Figure 2).

It is important to take into account the prevalence of corporate web sites as well, since a web presence is essential for production and service companies so that consumers can find them on the web. The Institute for Enterprise and Economic Research (IEER) conducted an empirical study in 2003 and found that among the 206 of Hungarian manufacturing companies geared for export, 107 of them (57%) reported operating their own website (Bacsko & Kollar, 2004). Subsequent to this empirical study, the authors performed content analysis on these websites. Key findings were that these companies regard websites as foremost as a tool to make and help maintain contact; the average corporate portal mostly provides basic information about their operation and main products. What is absent, however, is detailed information from on a firm’s operations and manufactured products, interactive marketing and communications solutions.

A report from the Hungarian Central Statistical Office (HCSO, 2011), which was published in October 2011, stated that 57% of businesses in Hungary had a website in 2010. This ratio is 10 percentage points lower than the EU average. Of all EU Member States, only Portugal, Cyprus, Latvia, Bulgaria and Romania had a smaller number of corporate websites than Hungary.

Data

The analysis uses a data series from the Google Insights for Search (GIS) application, the GKI Economic Research Co. (GKI) consumer confidence index, the Data House Ltd. automotive market information, and the Hungarian Central Statistical Office data series for household consumption and retail turnover. When collecting the data we tried to set the maximum length of time and the smallest unit of time to aggregate them. For all aspects in terms of spatial data, we used the entire territory of Hungary.[1]

The GIS services are available from January 1, 2004 in weekly segments. This application allows you to search for terms or categories in connection with a time series that generated from the search traffic in chronological order.

During our analysis, we used data sets generated from given categories, our assumption being that the resulting data sets are valid in representing the entire search traffic related to a consumer product group as compared to doing a search on some - arbitrary - key words and their combination.

From the Google data, we created each indicator in several steps. When selecting the categories we tried to cover all possible components of household consumption.

The first step, in our view, is to retrieve the weekly click-through data of categories related to the development of household consumption (see Table A1).

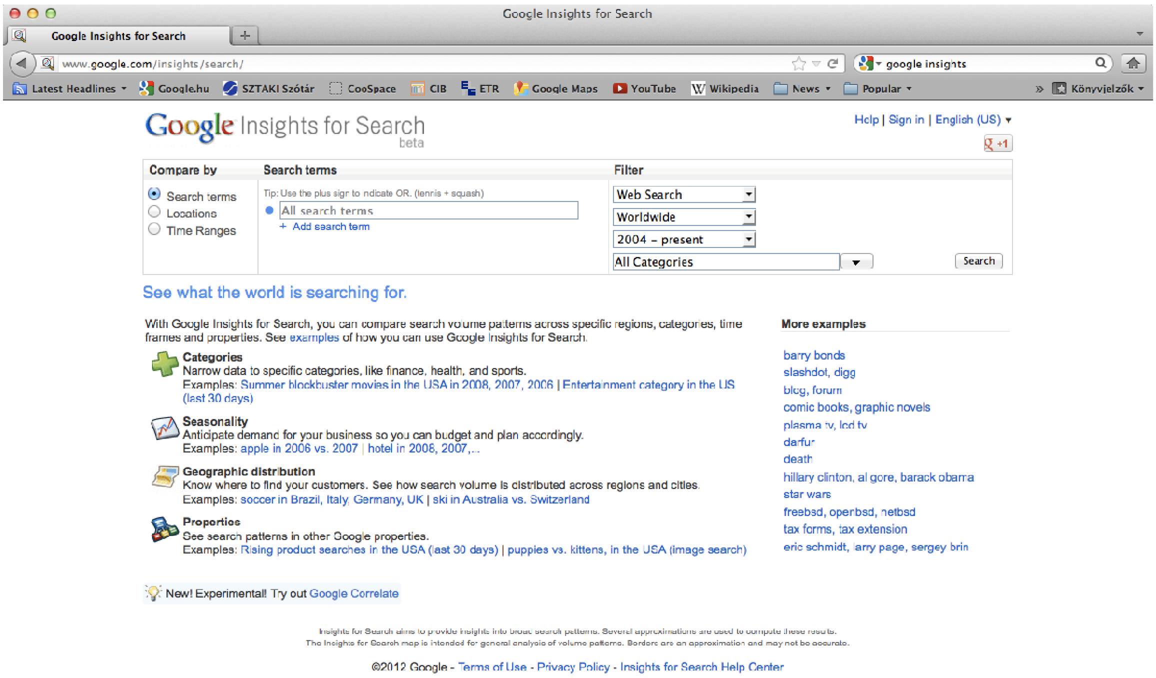
The second step is to convert the data into monthly and quarterly data segments.

Each category of the Google search system calculates an increasing percentage relative to the time of the first series. The starting point of the series is, mutatis mutandis, 0%, while for the other periods there is a percentage change recorded. The weekly Google homepage allows you to save the series, but the car market and retail data are available in monthly segments while household consumption data is quarterly. There is an aggregation of the search data with a greater time unit for comparison purposes: from the average weekly data for each month, we calculated the monthly average. For those weeks in which a new month begins, for practical reasons those weeks belong to those new months.

After that in the third step we choose the component wich were significantly linked ($p < 0.05$) with the reference time series (the retail trade, car sales, and household consumption). For our analysis we kept those which showed significant correlation with the reference time series (see A1 table).

[1] You can download the analysed original dataset from http://www.wargo.hu/tij/kutatas/ciret_2012/ciret_2012_tij_hm_data_120415.zip

Figure 3 The main page of Google Insights for Search (GIS)



Source: www.google.com/insights/search

Models

We examined the Google usability problem using a common research strategy. The development of retail turnover and car sales we used monthly data. In the case of household consumption, quarterly data were available. The basic model (baseline model) is simply a seasonal autoregressive model (SAR):

Model 1: $$C_t = \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-k} + u_t \tag{1a}$$

In case of household consumption we use the simple autoregressive model AR(1):

Model 1: $$C_t = \beta_0 + \beta_1 C_{t-1} + \beta_2 G_t + u_t \tag{1b}$$

Where *t-1* is the preceding month or quarter, and *t-k* is the same period the previous year.

The second, exclusively based on information from Google search, these models contain:

Model 2: $$C_t = \beta_0 + \beta_1 G_t + u_t \tag{2}$$

Where $\mathbf{G_t}$ is the variables (components) derived by the Google search categories ($g_1 \ldots g_n$).

Third, we also take into consideration an extended model, which includes the factors of #1 and #2 models:

Model 3: $$C_t = \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-k} + \beta_3 G_t + u_t \tag{3a}$$

In case of household consumption we use the following model:

Model 3: $$C_t = \beta_0 + \beta_1 C_{t-1} + \beta_2 G_t + u_t \quad (3b)$$

With this extended model, we examine whether it contributes to a more accurate forecast of retail sales if we take into account Google information searches with autoregressive effects. In other words, is it worth observing whether Google search does a better nowcast?

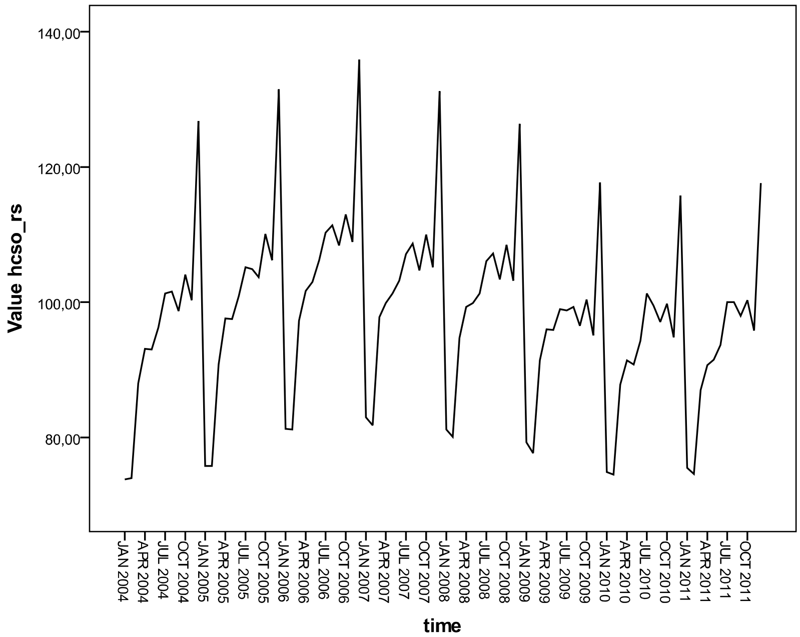
Empirical results

For the search categories in Table A1 we selected separately those categories in retail trade, car sales and household consumption with a statistically significant relationship and reference timeline. Next, a principal component analysis was performed to filter for multicollinearity, then we analysed the reference timeline with the help of each component containing information on Google searches. We then performed the analysis first using the baseline model, and then the model based on Google, and finally the extended model. Our results of the following three indicators are described separately.

a) Retail trade

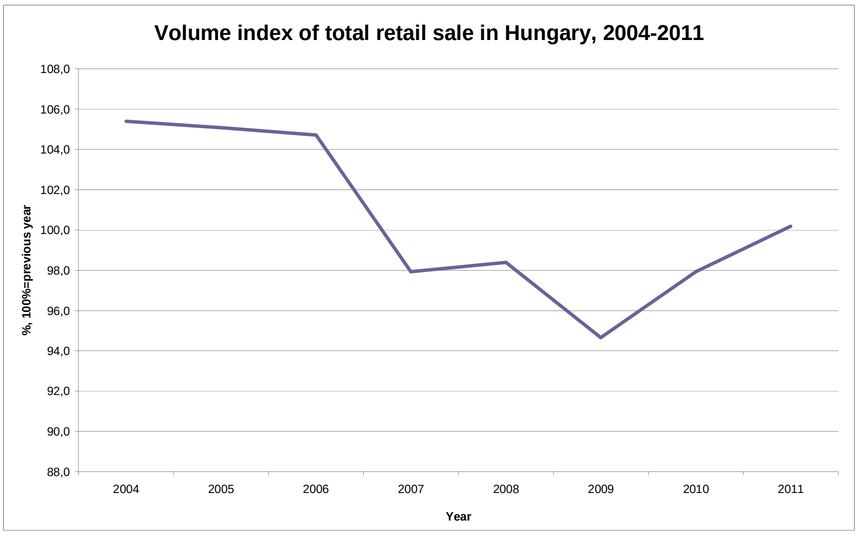
In the case of retail sales it is evident that the calendar-adjusted volume indices also show a high degree of seasonality. In addition, the effect of the crisis can be observed – after 2008 there was a significant decrease in retail sales in Hungary (see Figure 4 and Figure 5).

Figure 4 Retail trade in Hungary at 2005 constant price, (2004m1 – 2011m12, t= 96)



Source: HCSO

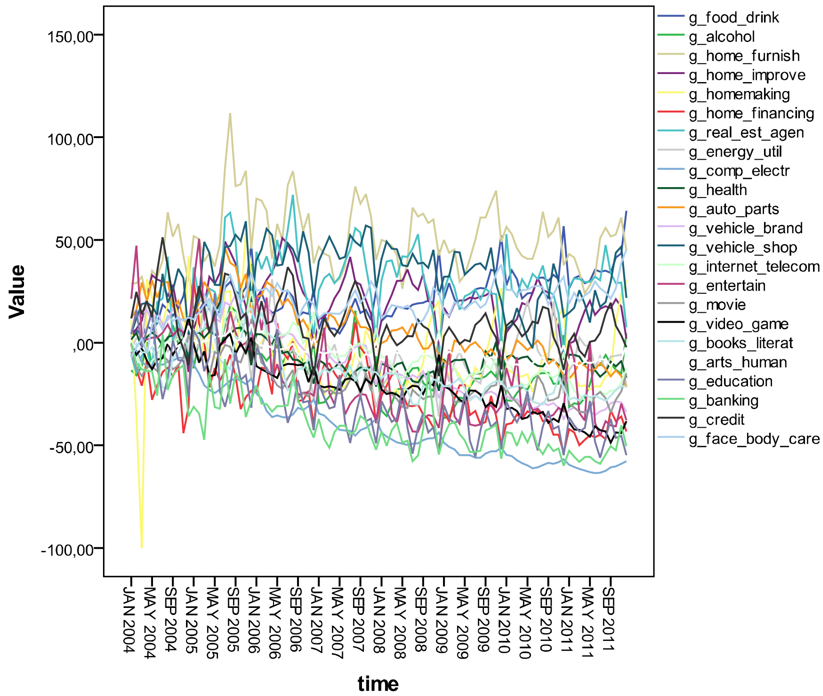
Figure 5 Volume index of total retail trade sale in Hungary, 2004-2011



Source: HCSO

With the help of Google Insights for Search our search categories selected in the first step (see Table A1) shows a more or less uniform picture. One the one hand, some seasonality can be observed; on the other hand, it is clearly the impact of the economic crisis – after 2008 there were fewer hits in most of the categories. (See Figure 6)

Figure 6 Selected Google data for Hungary, (2004m1 – 2011m12, t=96)



Source: own calculation based on GIS

For the reference time series (retail sales) 14 Google search categories were significantly correlated (see Table A1). Accordingly, we arrived at 14 components uncorrelated with one another using principal component analysis, in which Model #2 served to estimate retail sales.

According to the baseline model (Model #1) we can see a high degree of seasonality in the context of the values of autoregressive effect from the previous year, while insignificant for the constants and the last month effects. The reference time series is non-stationary and contain unit root (See Table A5). The strong autoregressive effects and extremely high R^2 value are obtained (R^2 = 0.918).

Table 1 Main result of estimations – retail trade (2004m1 – 2011m12)[*]

Factors	Baseline model (model #1)	Google (model #2)	Extended model (model #3)
Constant	1.189 (0.289)	98.423 (185.096)	4.302 (1.771)
Lag(1)	0.053 (1.653)	-	-
Lag (12)	0.932 (29.366)	-	0.958 (39.381)
F2_1	-	9.341 (17.475)	-
F7_1	-	-5.714 (-10.691)	-
F3_1	-	-3.074 (-5.752)	-
F8_1	-	-2.789 (-5.217)	-
F4_1	-	1.835 (3.433)	-
F12_1	-	1.482 (2.772)	-
F1_1	-	-	2.522 (7.355)
F14_1	-	-	-0.661 (-2.113)
R^2	0.918	0.849	0.952
Adj. R^2	0.916	0.839	0.950
Durbin-Watson	0.259	1.820	0.829
RMSE	3.670	5.016	2.333
T	84	96	84

**t value in brackets*

From the Google categories we choosed 14 item and we run a principal component analysis. The relationships between Google search categories and components are demonstrated in the A2 Table. In accordance with the analysis of the results from the second model the six component obtained from Google search categories had significant effect (F2_1, F3_1, F4_1, F7_1 F8_1 F12_1). For the detailed results see table 1. It is estimated that the Google model is based on effects worse than the baseline model (R^2 = 0.849). Here also the Google time series is not stationary and the co-integration test for reference and Google timeseries is ambiguous (see Table A5). Thus, according to these results information provided by Google (the second model) gives an approximation as inferior as the baseline model, that is, it does not provide additional information, if we regard this process as an seasonal autoregressive one. The cross-correlations show that the estimation based by Google data and the reference time series are similar with no lag or lead (see Figure 7.)

In addition to autoregressive factors, the extended model (Model #3) contains information supplied by Google. The extended model fits slightly better than the baseline model (R^2 = 0.952). The outcome of the analysis suggests that the factors calculated with the help of Google had a more significant impact on the results (F1_1, F14_1). Thus Google search results ultimately contributed to an improvement in the accuracy of the estimate. According to the RMSE values, the extended model which also contains information from Google fits better than the basic model.

Figure 7 Cross-correlation of retail trade with estimation based on Google data (ge_hcso_rs), (2004m1 – 2011m12, t=96)

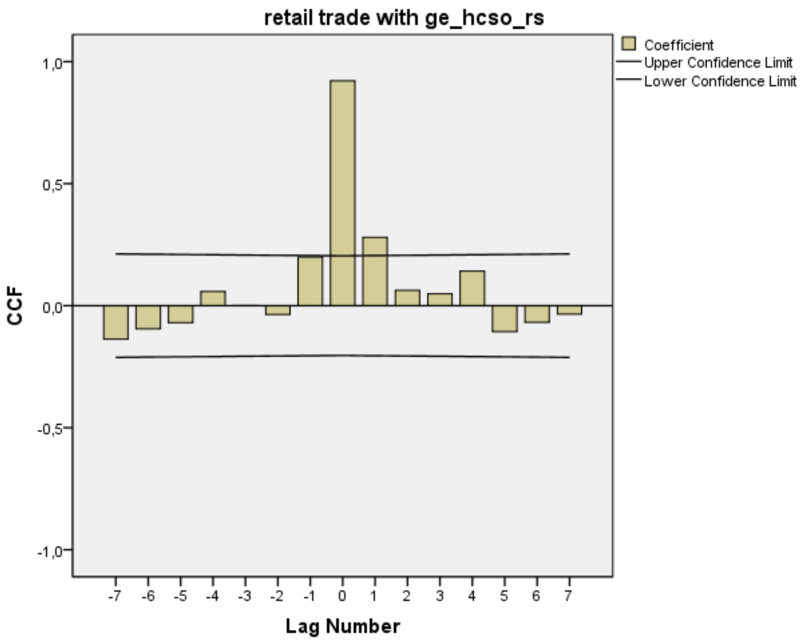
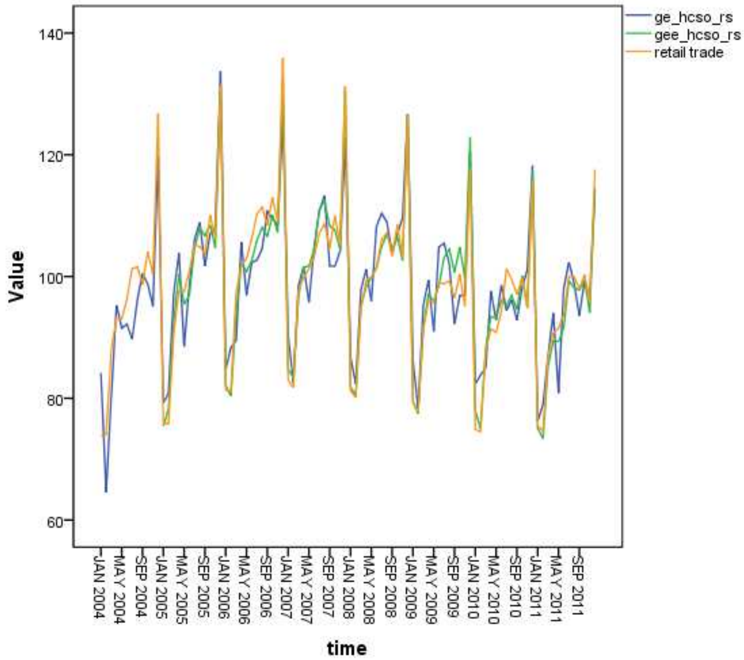


Figure 8 Retail trade and the estimated time series (2004m1 – 2011m12, t=96)



*: *ge_hcso_rs: estimation by Google data*
gee_hcso_rs: estimation by extended model (using Google data and autoregressive factors)

b) Car sales

For car sales (new and used cars) there was a significant downturn in 2008 and only from the third and fourth quarters of September 2011 there was a slight improvement. The GKI consumer confidence index, Hungary's only available long-term time series for consumer sentiment, moved more or less in line with the car sales time series ($r = 0.699 \quad p < 0.000$), but then breaks after May 2009[2] ($r=-0.252$, $p < 0.172$) (see Figure 9).

According to the baseline model (Model #1), previous month sales have a strong impact on monthly car sales, and the previous year's sales figures are also significant, while the constant is insignificant. Because of the strong autoregressive effects a high R^2 value is obtained ($R^2 = 0.780$).

For the car sales time series, Google showed a significant correlation in 19 categories according to paired correlations (see Table A1 and A3). Five components showed a significant effect when using the second model for estimations, which takes into account Google's data (F1_2, F3_2 F4_2 F6_2 F14_2). For the detailed results, see Table 2. It is estimated that the Google-based models have a better effect than the baseline model ($R^2 = 0.853$). The Google data is non-stationary time series and the reference time series with Google time series are co-integrated (See Table A5). Thus, according to the result of this assessment, information provided by Google (Model #2) gives better estimates than the basic model, for it contains additional information rather than the process being regarded as simply

[2] A possible reason of this phenomenon is that in spring 2010 elections were held in Hungary, and the expectations regarding the new government could fundamentally affect the consumer confidence. This hypothesis should be a subject of further research.

a seasonal autoregressive process. The cross-correlation data suggest that the Google estimated timeline is simultaneous with the reference timeline, i.e., Google is more appropriate for nowcasting (see figure 10).

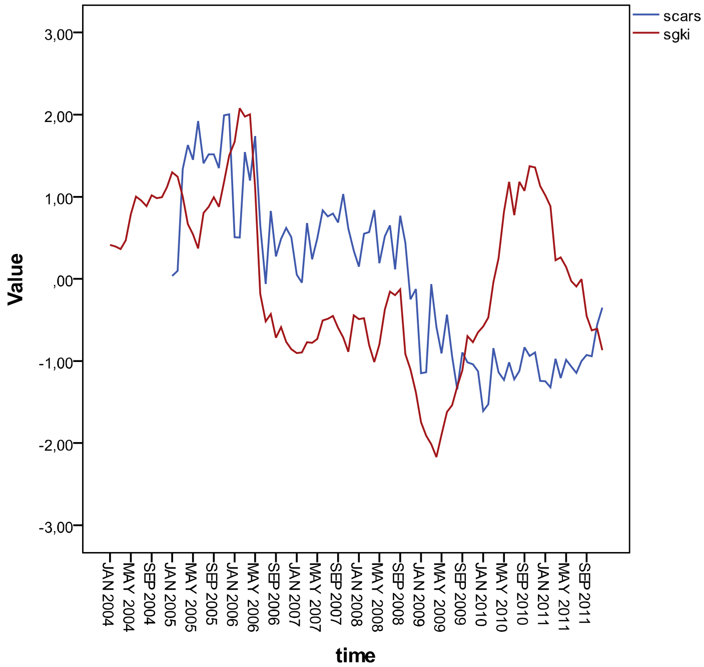
For the extended model (Model #3) the autoregressive members also contain information supplied by Google. The extended model is estimated by a non-stationary time series (See Table A5). The extended model fits slightly better than the basic model, and also than the second model (R^2 = 0.952). The estimation results suggest that the greater the Google calculated factors, the more significant the effect on the estimate (F1_2, F4_2, F6_2, F14_2). Thus the Google search results contributed significantly to improving the accuracy of the estimate. According to the RMSE values, models containing information from Google and autoregressive effect fit better than the basic model.

Table 2 Main result of estimations – car sales (2004m1 – 2012m12)[*]

Factors	Baseline model (model #1)	Google (model #2)	Extended model (model #3)
Constant	4755.364 (1.583)	53949.327 (93.501)	35048.749 (4.387)
Lag(1)	0.644 (8.339)	-	0.301 (2.860)
Lag (12)	0.233 (3.295)	-	0.045 (0.534)
F1_2	-	98899.258 (16.227)	6548.026 (3.919)
F4_2	-	-3848.312 (-7.109)	-2338.657 (-2.692)
F6_2	-	2646.544 (4.805)	1342.506 (2.025)
F3_2	-	2613.504 (3.968)	1287.287 (1.638)
F14_2	-	-1317.808 (-2.400)	-1863.008 (-3.175)
R^2	0.780	0.853	0.952
Adj. R^2	0.774	0.844	0.950
Durbin-Watson	2.008	1.412	1.811
RMSE	5084.184	4825.501	4311.926
T	72	84	72

**t value in brackets*

Figure 9 Car sales and the GKI Consumer Confidence Index, standardized data, (2004m1 – 2011m12, t=96)



**: scars – car sales*

sgki – GKI consumer confidence index

Figure 10 Cross-correlation of car sales with estimation based on Google data (2004m1 – 2011m12, t=96)

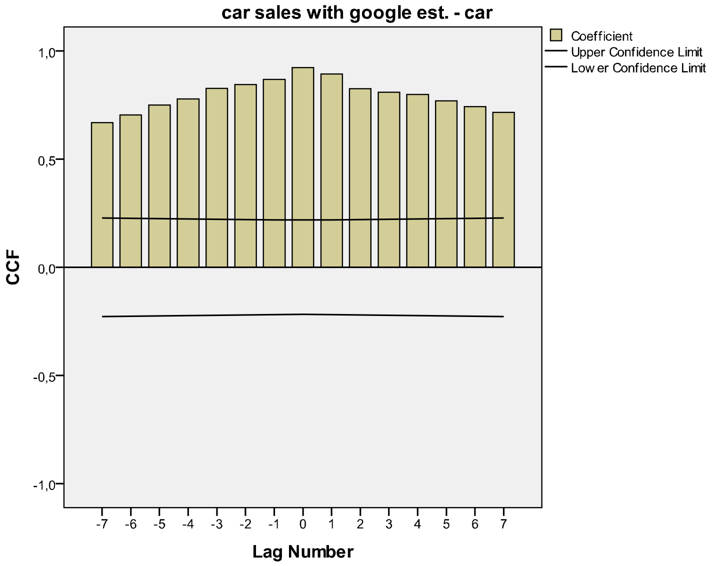
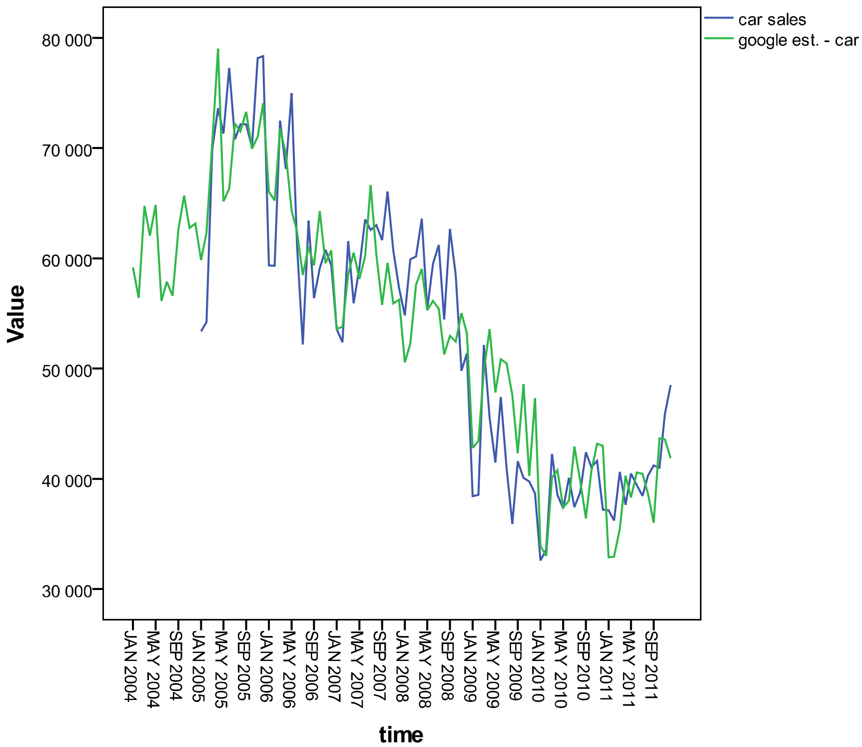


Figure 11 Car sales and estimated car sales by Google (2004m1 – 2011m12, t=96)

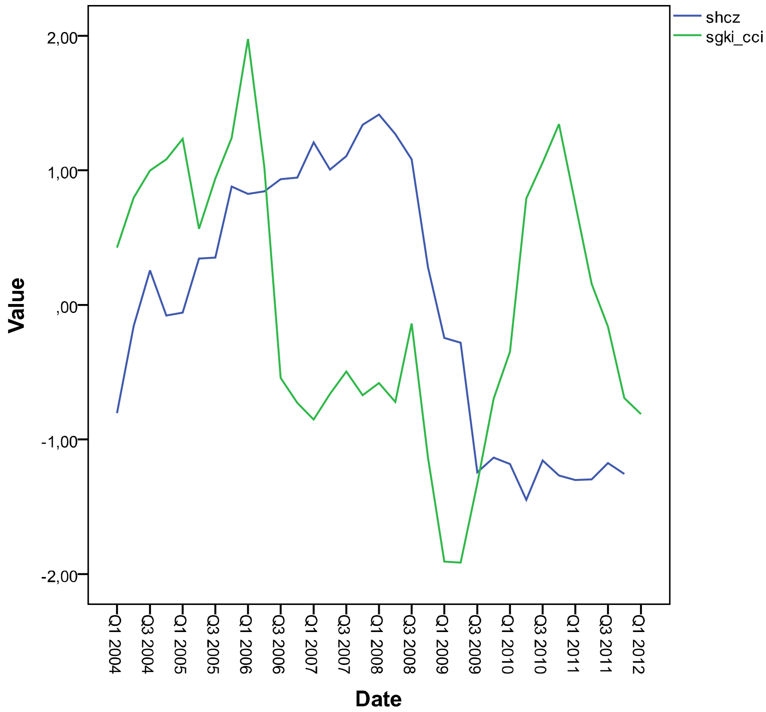


c) Household consumption

Quarterly data for household consumption are only available. The estimates are based on a total of 32 observations, so the calculated results can be regarded as a preliminary result. The longer the time series, the more likely the subsequent results will be valid. Nevertheless, it is still worthwhile to perform an analysis for this reason: in the estimation of household consumption, can some kind of role be detected with the information obtained from Google.

In Hungary, the household consumption clearly illustrates the effect of the economic crisis: after the second quarter in 2008 a sharp decline was observed until the third quarter in 2009, which was then followed by stagnation along low consumption levels (see Figure 12. below). The transformed timeline of the GKI consumer confidence index of the quarterly data set was not in sync with the progression of household consumption - probably the sensitivity of the Hungarian population to political factors played a role in this.

Figure 12 Household consumption and GKI Consumer Confidence Index, standardized data (2004q1 – 2011q4, t=32)



*: *shcz – household consumption*
sgki – GKI consumer confidence index

According to the baseline model (Model #1) consumption in the previous period has a strong effect on household consumption in the current quarter, while the effect of constant term is insignificant. In the estimation we obtain a high R^2 value ($R^2 = 0.888$).

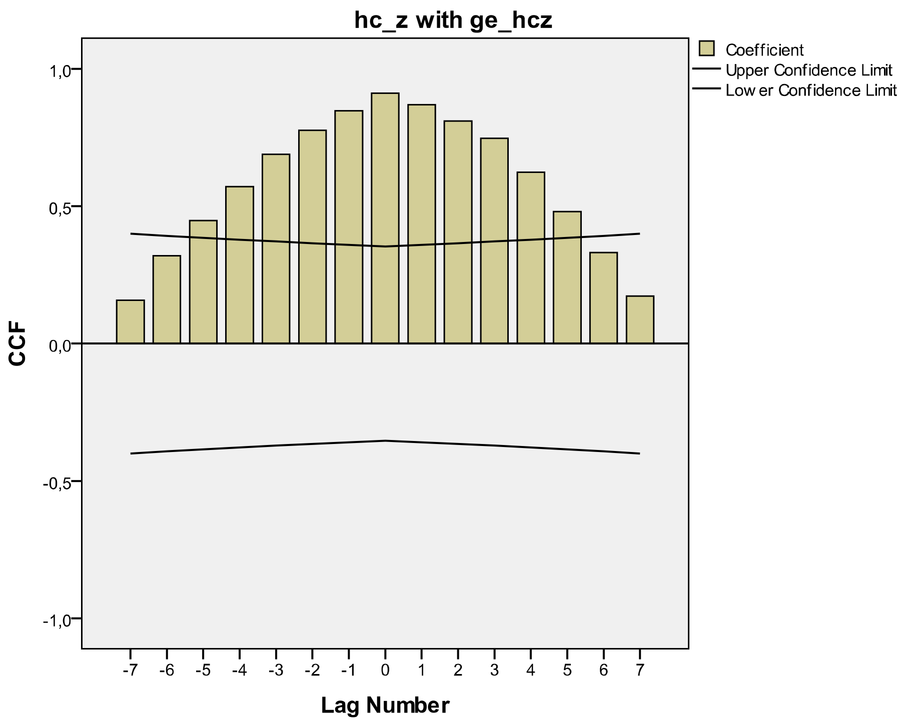
The ten Google categories for household consumption show a significant correlation between the paired correlations (see Table A1 and A4.). Estimates from the second model which take into account data from Google showed a significant effect in three components (F1_1, F3_1 F4_1). For the detailed results, see table 3. It is estimated that the effect of the Google based model is less than that of the basic model (R^2 = 0.849). The Google time series is non-stationary and the combined reference time series with Google time series are co-integrated (See Table A5). Thus, according to the results of the assessment, information provided by Google (model #2) gives results almost as good as the basic model. The cross-correlation data suggest that the Google estimated time series is simultaneous with the household consumption. The Google is more appropriate for nowcasting (see Figure 13).

Table 3 Main result of estimations – household consumption (2004q1 – 2012q4)*

Factors	Baseline model (model #1)	Google (model #2)	Extended model (model #3)
Constant	120389.828 (0.659)	2886361.542 (339.529)	515823.938 (2.596)
Lag(1)	0.958 (15.199)	-	0.821 (11.1979
F1_1	-	79805.858 (9.040)	24972.246 (3.282)
F4_1	-	-49085.120 (-4.388)	-
F2_1	-	42951.496 (5.093)	-
F3_1		23699.449 (2.750)	-
F5_1		-21364.469 (-2.240)	-
R^2	0.888	0.849	0.919
Adj. R^2	0.885	0.819	0.914
Durbin-Watson	1.538	1.117	1.870
RMSE	36653.366	42502.495	31148.167
T	31	32	31

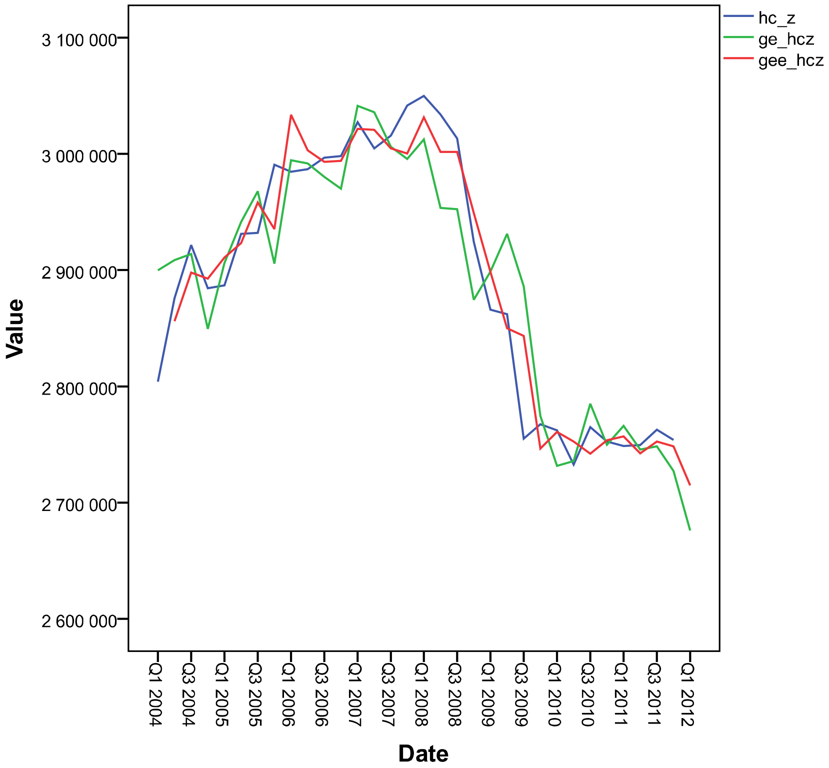
**t value in brackets*

Figure 13 Cross-correlation of household consumption (hc_z) with estimation based on Google data (ge_hcz), (2004q1 – 2011q12, t=32)



For the extended model (Model #3) also contain information supplied by Google. The extended model is estimated by a non-stationary time series (See Table A5). The extended model fits better than the basic model, and better than the second model (R^2 = 0.919). The estimation results suggest that a factor calculated using Google still has a significant effect on the estimate (F1_1). Thus, the Google search results here also contributed significantly to improving the accuracy of the estimate (see Figure 14). According to the RMSE values, models containing information from Google and autoregressive factor fit better than the basic model.

Figure 14 Household consumption and estimiations by model #2 and model #3 (2004q1 – 2011q12, t=32)



*: *hc_z: household consumption (million HUF, at constant price 2005)*
ge_hcz: estimation by Google data
gee_hcz: estimation by extended model (using Google data and autoregressive factor)

Conclusions

In analyzing the effectiveness of information provided by Google, the consumption estimates we studied were from a country with relatively low Internet penetration, characterized in terms of Internet use of both consumers and entrepreneurs. The question we put forward was that in such an environment can the same sort of results be shown as in developed countries, mainly in the USA: that is, can information from Google effectively contribute to more accurate consumption nowcasts.

Hungarian data calculations essentially show that information from Google, if not by itself, but taken into account with other factors (i.e. autoregressive effects), can help to ensure that household consumption, and its various aspects, are more accurately estimated. For retail sales and household consumption, data from Google alone were not effective, but with the autoregressive factors, forecast accuracy was improved. For car sales, data from Google already significantly contributed to a more accurate estimate.

Although these were about nowcasts, in reality they can represent a 30 to 40 day forecast, for in any countries reference time series data are available much later than the data we currently estimated with using Google.

The results suggest that it is worth experimenting with the data provided by Google (taking into account the different categories and different methods of these components to build an aggregate) even in countries characterized by low Internet penetration. Survey data, together with macro indicators as well, need to also be taken into account when nowcasting consumption.

References

Askitas, Nikolaous - Zimmermann, Klaus. F. (2009): Google Econometrics and Unemployment Forecasting. IZA. Bonn. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1415585 (acc.: April 5 2012)

Bacskó, Gergely - Kollár, Tamás. (2004): Info-kommunikációs technológiák a vállalati szektorban. Institute for Economic and Enterprise Research. Budapest. http://gvi.hu/data/papers/kfuz_2004_2.pdf (acc.: March 22 2012)

Choi, Hyunyoung - Varian, Hal. (2009): Predicting the Present with Google trends. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1659302 (acc.: March 20 2012)

GKI Economic Research Co. (2012): A tavaly nyár végi szintre emelkedett a GKI-Erste konjunktúraindex. Budapest. http://www.gki.hu/sites/default/files/users/Petz_Raymund/konj_1203.pdf (acc.: April 2 2012)

Hungarian Central Statistical Office (2011): Az információs és kommunikációs technológiák használata az üzleti szférában, 2010. http:// www.ksh.hu/docs/hun/xftp/idoszaki/ikeszk/ikeszk10.pdf (acc.: April 10 2012)

Kholodilin, Konstantin A. - Maximilian Podstawski - Boriss Siliverstovs (2010): Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US, Discussion Papers – 997, DIW Berlin. http://www.diw.de/documents/publikationen/73/diw_01.c.356220.de/dp997.pdf (acc.: March 28 2012.)

Schmidt, Torsten - Vosen, Simeon (2009a): Forecasting Private Consumption: Survey-based Sentiment Indicators vs. Google Trends, CIRET/KOF/GKI-Workshop 2009, Budapest

Schmidt, Torsten - Vosen, Simeon. (2009b): Forecasting Private Consumption: Survey-based Indicators vs. Google Trends. Ruhr Economic Papers, 155. http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1514369_code343237.pdf?abstractid=1514369&mirid=1 (acc.: March 14 2012)

Appendix

Table A1 The correlations of categories of Google Insights for Search which were analysed

	retail trade			car sales			household consumption		
	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N	Pearson Correlation	Sig. (2-tailed)	N
g_food_drink	0.311	0.002	96	-0.453	0.000	84	-0.610	0.000	32
g_alcohol	0.573	0.000	96	0.559	0.000	84	0.182	0.318	32
g_home_furnish	0.188	0.067	96	0.326	0.003	84	0.251	0.166	32
g_home_improve	0.035	0.736	96	0.466	0.000	84	0.424	0.016	32
g_homemaking	0.416	0.000	96	0.270	0.013	84	0.045	0.806	32
g_home_financing	-0.126	0.221	96	0.627	0.000	84	0.595	0.000	32
g_real_est_agen	-0.112	0.277	96	0.197	0.073	84	0.303	0.092	32
g_energy_util	-0.319	0.002	96	0.203	0.064	84	-0.207	0.256	32
g_comp_electr	0.074	0.474	96	0.819	0.000	84	0.419	0.017	32
g_health	-0.389	0.000	96	0.374	0.000	84	-0.002	0.990	32
g_auto_parts	0.094	0.363	96	0.831	0.000	84	0.556	0.001	32
g_vehicle_brand	-0.005	0.961	96	0.848	0.000	84	0.568	0.001	32
g_vehicle_shop	0.210	0.040	96	0.623	0.000	84	0.763	0.000	32
g_internet_telecom	0.399	0.000	96	0.717	0.000	84	0.647	0.000	32
g_entertain	-0.372	0.000	96	0.425	0.000	84	0.137	0.455	32
g_movie	0.250	0.014	96	0.580	0.000	84	0.284	0.115	32
g_video_game	0.263	0.009	96	0.781	0.000	84	0.557	0.001	32
g_books_literat	0.030	0.768	96	0.734	0.000	84	0.437	0.012	32
g_arts_human	0.444	0.000	96	0.701	0.000	84	0.500	0.004	32
g_education	-0.336	0.001	96	0.524	0.000	84	0.249	0.169	32
g_banking	-0.044	0.673	96	0.648	0.000	84	0.301	0.094	32
g_credit	-0.202	0.048	96	0.334	0.002	84	0.176	0.334	32
g_face_body_care	0.317	0.002	96	-0.168	0.127	84	-0.257	0.156	32

Table A2 The component matrix of principal component analysis for retail trade

Component Matrix[a]

	Component													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
g_food_drink	-.355	.728	.491	-.089	-.066	.052	.081	.180	.018	.185	.051	-.035	.095	.037
g_alcohol	.670	.526	.206	.087	-.201	.148	-.287	-.086	.240	.064	-.048	.049	-.074	-.018
g_homemaking	.406	.626	.267	.063	.101	-.592	-.028	-.038	-.011	-.058	-.026	.027	.003	.004
g_energy_util	.403	-.486	.585	.270	.188	.092	-.302	.027	-.190	.017	-.108	-.002	.048	.006
g_health	.747	-.397	.421	.099	.156	-.026	.109	.037	-.004	.067	.206	.009	-.084	-.068
g_vehicle_shop	.115	.059	-.447	.824	.244	-.023	.037	.164	.111	.044	-.024	-.025	-.006	.017
g_internet_telecom	.775	.398	-.310	.090	.058	.053	.154	-.242	-.119	.150	-.055	-.040	.064	-.059
g_entertain	.819	-.262	.174	-.219	.048	.040	.340	.128	.055	.016	-.173	.109	-.025	.010
g_movie	.858	.393	.020	-.149	-.054	.074	.025	.119	-.062	-.092	-.043	-.207	-.083	.008
g_video_game	.889	.124	-.349	-.075	-.024	.023	-.083	-.011	-.152	.065	.092	.081	-.050	.117
g_arts_human	.778	.461	-.247	-.109	.054	.138	-.102	.157	-.016	-.151	.069	.081	.118	-.057
g_education	.776	-.490	.125	-.124	.181	-.002	.027	-.137	.219	-.044	.050	-.078	.111	.065
g_credit	.478	-.352	.130	.445	-.639	-.070	.112	.003	-.037	-.046	.023	-.003	.054	.005
g_face_body_care	-.341	.630	.429	.353	.130	.252	.211	-.156	-.041	-.151	.028	.040	-.026	.036

Extraction Method: Principal Component Analysis.

a. 14 components extracted.

Table A3 The component matrix of principal component analysis for car sales

Component Matrix[a]

	Component																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
g_food_drink	-.464	-.573	.381	.474	.136	.088	.077	.107	.088	.105	.041	-.033	.106	.037	.023	.017	.090	.020	.008
g_alcohol	.618	-.411	.389	.296	-.127	.318	.128	-.163	.090	-.163	-.027	.047	.044	-.060	-.049	-.001	-.048	-.030	-.006
g_home_furnish	.256	.229	.737	.303	.292	-.323	-.110	-.136	-.075	-.002	-.098	.041	-.053	-.080	.037	-.001	.011	.007	.004
g_home_improve	.311	.773	.321	.113	-.147	.312	-.042	-.028	-.157	.166	.084	.040	.016	-.018	-.030	-.069	.015	.010	.019
g_home_financing	.717	.383	-.010	-.140	-.123	-.177	.451	-.082	.152	.154	-.098	-.044	.060	-.023	-.012	.008	.001	.006	.005
g_comp_electr	.931	-.172	-.169	-.077	-.116	-.054	-.122	.002	.059	-.116	-.048	.030	.055	.001	.022	-.052	.013	.089	.064
g_health	.739	.163	-.343	.365	.329	.083	-.077	.067	-.061	.047	-.131	-.069	.082	.090	-.003	-.024	-.085	.000	-.011
g_auto_parts	.916	.200	.113	-.037	-.076	.143	-.046	.056	-.146	-.062	-.051	-.168	-.020	-.038	-.029	.119	.039	.007	.019
g_vehicle_brand	.975	.131	-.014	-.070	.031	.037	.023	.015	-.005	-.084	.018	-.008	.006	-.014	.003	-.037	.057	.062	-.089
g_vehicle_shop	.230	.456	.592	-.433	.333	.084	.001	.208	.154	-.084	-.001	.069	.035	.038	-.042	.010	-.003	-.018	.010
g_internet_telecom	.771	-.292	.384	-.177	.030	-.214	-.050	-.161	-.066	.035	.195	-.045	.065	.108	-.046	.027	-.046	.016	-.002
g_entertain	.786	-.112	-.402	.190	.164	-.195	.132	.178	-.093	-.023	.141	.043	.045	-.148	-.053	-.010	-.008	-.023	.011
g_movie	.792	-.494	.160	.109	-.076	-.060	.081	.110	-.016	.057	-.064	.009	-.184	.082	-.109	-.045	.018	-.005	.006
g_video_game	.888	-.216	.105	-.205	-.200	-.109	-.149	.020	-.066	-.011	-.066	-.028	.099	.019	.046	-.058	.056	-.097	-.004
g_books_literat	.817	-.220	-.095	-.177	.362	.207	.121	-.089	.060	.021	.091	-.099	-.107	-.021	.107	-.041	.006	-.018	.019
g_arts_human	.739	-.482	.273	-.181	-.153	.088	.022	.150	-.089	.136	-.034	.121	-.009	-.030	.103	.051	-.061	.022	-.013
g_education	.780	.114	-.511	.053	.174	.071	.031	-.151	-.043	.010	-.012	.191	-.002	.082	-.003	.069	.067	-.019	.008
g_banking	.819	.089	-.166	.122	-.075	-.003	-.396	-.013	.296	.143	.037	-.009	-.028	-.062	-.027	.029	-.005	-.010	-.010
g_credit	.534	.578	.094	.462	-.269	-.121	.096	.108	.078	-.122	.099	.003	-.054	.089	.091	.015	-.015	-.018	.003

Extraction Method: Principal Component Analysis.

a. 19 components extracted.

Table A4 The component matrix of principal component analysis for household consumption

Component Matrix[a]

	Component									
	1	2	3	4	5	6	7	8	9	10
g_food_drink	-.646	-.120	.496	.529	.176	.089	.054	.005	.013	-.002
g_home_improve	.414	.785	-.199	.121	.391	-.037	.003	.040	-.046	.007
g_home_financing	.687	.258	-.327	.376	-.436	.143	-.041	.017	.009	.007
g_comp_electr	.902	-.363	-.083	-.014	.059	.109	.163	-.030	-.048	.049
g_auto_parts	.960	.029	-.067	.031	.229	-.027	-.032	-.054	.118	.019
g_vehicle_brand	.980	-.065	-.070	.099	.039	-.034	.070	-.087	-.024	-.066
g_vehicle_shop	.433	.614	.570	-.172	-.258	-.035	.112	-.009	.019	.005
g_internet_telecom	.863	-.046	.433	-.052	.048	.126	-.200	-.043	-.049	.009
g_video_game	.932	-.217	.103	-.130	.092	.164	.026	.142	.025	-.027
g_books_literat	.848	-.284	.140	.188	-.088	-.366	-.027	.057	-.017	.011

Extraction Method: Principal Component Analysis.

a. 10 components extracted.

Table A5 Test results for unit root and cointegration (Augmented Dickey-Fuller tests)

Variable name	Lags	T	Test statistics Z (t)	Interpolated Dickey-Fuller critical values			MacKinnon appr. P-value for Z(t)
				1%	5%	10%	
Retail Trade (HCSO_RET)	11	84	0.346	-3.532	-2.903	-2.586	0.9794
Estimation for model #2 (GE_HCSO)	11	84	-0.921	-3.532	-2.903	-2.586	0.7808
Estimation for model #3 (GEE_HCSO)	11	72	2.202	-3.549	-2.912	-2.591	0.9989
Residual for model #2 (GR_HCSO)	11	84	-2.677	-3.532	-2.586	-2.586	0.0781
Residual for model #3 (GER_HCSO)	11	72	-2.330	-3.549	-2.912	-2.591	0.1626
Car sales (CARS)	1	82	-1.672	-3.535	-2.904	-2.587	0.4457
Estimation for model #2 (GE_CARS)	1	94	-1.449	-3.518	-2.895	-2.582	0.5584
Estimation for model #3 (GEE_CARS)	1	70	-1.538	-3.552	-2.914	-2.592	0.5144
Residual for model #2 (GR_CARS)	1	82	-5.114	-3.535	-2.904	-2.587	0.0000
Residual for model #3 (GER_CARS)	1	70	-5.537	-3.553	-2.914	-2.592	0.0000
Household consumption (HC_Z)	1	30	-0.564	-3.716	-2.986	-2.624	0.8789
Estimation for model #2 (GE_HCZ)	1	31	-0.042	-3.709	-2.983	-2.623	0.9549
Estimation for model #3 (GEE_HCZ)	1	30	-0.293	-3.716	-2.986	-2.624	0.9265
Residual for model #2 (GR_HCZ)	1	30	-3.114	-3,716	-2.986	-2.624	0.0256
Residual for model #3 (GER_HCZ)	1	29	-3.244	-3.723	-2.989	-2.625	0.0176