Google Data in Bridge Equation Models for German GDP

Thomas B. Götz*
Deutsche Bundesbank, Frankfurt am Main, Germany - thomas.goetz@bundesbank.de

Thomas A. Knetsch
Deutsche Bundesbank, Frankfurt am Main, Germany - thomas.knetsch@bundesbank.de

Abstract

There has been increased interest in the use of "big data" when it comes to forecasting macroeconomic time series such as private consumption or unemployment. However, applications on forecasting GDP are rather rare. In this paper we incorporate Google search data into a Bridge Equation Model, a version of which usually belongs to the suite of forecasting models at central banks. As the choice of which Google search terms to add to which equation is crucial, we compare different approaches (among others factor and shrinkage methods) in terms of their out-of-sample forecast performance in pseudo-real time. We find that there are indeed sizeable gains possible from using Google search data, at least when replacing survey by Google variables and when considering a more recent evaluation period.

Keywords: Big Data; Bridge Equation Models; Forecasting; Variable Selection.

1. Introduction

Given the widespread use of the internet,[1] the question arises whether we are able to generate knowledge for macroeconomic activity from internet data. Advances in computer technology now enable researchers to not only generate vast amounts of data, but also process non-standard, rather unstructured ones emerging on, for example, the internet. For such data we use the term "big data" here.[2] In this paper we investigate whether such big data - more specifically, data derived from them - lead to forecast accuracy improvements as far as macroeconomic quantities in general and, due to the omnipresence of the web, the gross domestic product (GDP) in particular are concerned. To be more precise, we employ Google Search data, a proxy for internet usage behavior, for Germany.

To the best of our knowledge, almost all related studies focused on a specific macroeconomic indicator, usually sampled at the monthly frequency, instead of economic activity as a whole, i.e., GDP growth. In this paper we intend to fill this gap in the literature by incorporating Google search data into a Bridge Equation Model (BEM), one of the workhorse models used for short-term GDP forecasting in many central banks (see, e.g., ECB, 2008, Bell et al., 2014 or Bundesbank, 2013). Indeed, the model's simplicity, transparency and structure lend themselves eminently to such an analysis. Furthermore, as the choice of which Google search terms enter a given model often turns out to be crucial for the forecast performance in the end, we investigate different selection procedures in terms of their out-of-sample forecast performance: principal components analysis (PCA), partial least squares (PLS), the least absolute shrinkage and selection operator (LASSO), Boosting and a couple of subjective (ad-hoc) methods. Finally, we also pay attention to the specific Google search terms actually chosen over time.

2. Model and Data

BEMs were introduced by Klein and Sojo (1989) as a regression-based system for GDP growth forecasting, whereby the different GDP components of the National Accounts (NAs) are modeled individually, and are

[1] In Germany, e.g., nearly everyone below the age of 45, about 90% of people aged 45-64 years and almost 50% of people over the age of 65 used the internet in 2015 (Destatis, 2015).

[2] It is common to characterize big data using the so-called "4 V's definition" (volume, velocity, variety, veracity). Varian (2014) provides an interesting overview of tools to manipulate and analyze big data in general.

usually augmented with short-term indicators tailored to the specific equation in question. Technically, a BEM is characterized by dynamic linear equations, whereby GDP growth or a component thereof represents the (low-frequency) dependent variable. Apart from low-frequency lags, the regressor set may contain time-aggregated short-term (high-frequency) indicators, e.g., industrial production. The latter may be predicted using survey indicators to take advantage of the timeliness of such indicators.
Letting y_t denote the quarterly growth rate of GDP (or one of its components) in period $t(= 1, \dots, T)$, x_t a short-term monthly, usually "hard" indicator and z_t a timely, "soft" survey indicator, our benchmark BEM can be summarized by the following three equations or steps:

$$\begin{aligned}
(1) \quad z_t^m &= \mu_z + \rho_z(L^{1/3}) z_{t-1/3}^m + \epsilon_t^z, \\
(2) \quad x_t^m &= \mu_x + \rho_x(L^{1/3}) x_{t-1/3}^m + \delta_x(L^{1/3}) z_t^m + \epsilon_t^x, \\
(3) \quad y_t &= \mu_y + \rho_y(L) y_{t-1} + \beta(L) x_t^q + \epsilon_t^y,
\end{aligned}$$

where ρ, β, δ are lag polynomials in L (quarterly frequency, i.e., $L^i y_t = y_{t-i}$) or $L^{1/3}$ (monthly frequency, i.e., $L^{i/3} x_t^m = x_{t-i/3}^m$). Between steps (2) and (3), the monthly indicators get temporally aggregated using $x_t^q = w(L^{1/3}) x_t^m = \sum_{i=0}^{2} w_i L^{i/3} x_t^m$, whereby the weights w are determined by the stock-/flow-nature of the variable in question (see, e.g., Silvestrini and Veredas, 2008). Hence, we estimate – with ordinary least squares (OLS) – each equation in turn over $t = 1, \dots, T$ and forecast the respective variable until period $T + h$ using the thusly predicted values of eventual regressors from a previous step. Finally, GDP growth is obtained as weighted average of the various GDP components according to their share in the NAs.
In this paper, we consider an adapted submodel of the full BEM routinely run for short-term forecasting at the Deutsche Bundesbank (see Bundesbank, 2013 for details) as an example. In particular, it is a disaggregated BEM covering the production side of the German NAs and is summarized in Table 1 below.[3] Whenever an x-indicator is absent in a row, step (2) is dropped and a time-aggregated version of z enters step (3) instead of x_t^q. For detailed information on data features, i.e., transformations, publication delays and so forth, we refer to Table 6 of the full working paper version (WPV).

Table 1: The disaggregated production-side Bridge Equation Model

GDP Component (y)	Monthly Indicators (x)	Survey Indicator (z)
Mining	Production Mining	ifo ind
Manufacturing	Industrial Production	ifo ind
Energy & Water Supply	Energy Production	ifo ind
Construction	Production in Construction	ifo ind
Trade (incl. cars)	Real Retail Sales (incl. cars)	ifo ind
Traffic	Toll (Industrial Production)	ifo ind
Hotel Industry	Sales Hotel Industry	ifo ind
Net taxes	VAT	ifo ind
Agriculture & Forestry		ifo ind
Information & Communication		ifo ind
Housing		ifo ind
Financial Services		pmi serv
Corporate Services		pmi serv
Public Services, Health & Education		pmi serv
Other Services		pmi serv

The Google search data we employ in this paper stem from a data set that is provided to the European Central Bank by Google, but are very similar to the ones derived from the Google Trends application *Insights for Search*: they are available as of 2004, appear on a weekly basis without publication lag, measure relative changes in search volumes, are based on random samples from all queries during a day, do not get revised and are published in a seasonally unadjusted fashion. We end up with 17 categories (e.g., Autos & Vehicles) and 183 subcategories (e.g., Vehicle Brands).[4] Prior to incorporating the Google data into our BEM, we

[3] "ifo ind" corresponds to the ifo index assessing the current business situation in trade and industry, whereas "pmi serv" represents the purchasing managers index in services.

[4] We a-priori disregarded nine categories and some specific subcategories we deemed unfitting. Details as well as a full list of (sub)categories used are presented in Tables 2 and 3 of the WPV.

time-average them to the monthly frequency and apply seasonal adjustment using the ARIMA-X12-approach. The bootstrap sequential quantile test[5] of Smeekes (2015) returns zero rejections of a unit root such that we compute first differences of all Google search variables in our dataset.

As far as the augmentation of our BEM with Google indicators is concerned, we propose to treat the Google data similarly to survey indicators. To be more precise, rather than attempting to predict aggregate GDP directly using Google search data, we intend to let the effect run indirectly through the monthly x-indicators and thereby the GDP components. Indeed, users are more likely to search for "jobs", "used car" or "last-minute holiday offers" than, e.g., "GDP". Furthermore, we delete the survey indicators from those equations we augment with Google data, i.e., equations (2). Hence, the augmented BEM is obtained by amending step (2) above to (2*) and adding step (1*); steps (1) and (3) remain unchanged, whereby surveys are only used in the equations for those GDP components not containing an x_t^q in their regressor set:

$$\begin{aligned} (1^*) \quad g_t^m &= \mu_g + \rho_g(L^{1/3})g_{t-1/3}^m + u_t^g, \\ (2^*) \quad x_t^m &= \mu_x + \rho_x(L^{1/3})x_{t-1/3}^m + \gamma_x(L^{1/3})g_t^m + u_t^x, \end{aligned}$$

where g is generically employed for Google variables. This implies that (i) we do not augment equations (3) in case an x-indicator is absent (e.g., Housing) and (ii) we add Google indicators g *instead* of survey indicators z in the remaining equations.[6] Schematically, the proposed augmentation of the BEM in Table 1 can be represented by renaming the third column into "Survey & Google Indicators (z or g)" and replacing "ifo ind" by g in the first eight rows.

3. Google Variable Selection

In this section we discuss how we choose which Google (sub)categories enter which equation of the BEM for it is neither practical nor feasible to include all of the candidate series. Note that this issue may be a subtle one, as a statistically meaningful Google (sub)category may be unjustifiable economically (something we may label a "spurious relationship"), or an intuitively "fitting" indicator might, in fact, not have a beneficial effect due to either low popularity (e.g., industry-related ones) or adverse search behavior (e.g., Vehicle Brands in light of the recent emission scandal affecting many car brands). We consider and evaluate various alternative out-of-sample procedures.[7]

Subjectively: We choose the Google search data, once on a category- and once on a subcategory level, by hand, i.e., based on "common sense". All indicators enter with one lag.

Google Correlate: We use this Google Trends tool to find search terms that have the largest correlation with the respective x-indicator. Then, we manually filter out terms suggesting "spurious relationships" and look for those (sub)categories in our data set corresponding as "closely" as possible to the remaining ones.

PCA: Based on a pre-selection of eligible Google variables for each x-series, we draw unrestricted, i.e., over all viable subcategories, and category-specific PCA-factors (the latter labeled PCA-Cat). The number of factors is determined by a scree test, the lag length by the Schwartz information criterion (SIC).

PLS: Contrary to PCA, PLS takes the relationship of the subcategories with the corresponding x-indicator into account by drawing the factors so as to maximize their correlation with the target (conditional on previously drawn factors). Similar to PCA, we draw unrestricted and category-specific PLS-factors (the latter labeled PLS-Cat). The number of factors and lag lengths are jointly determined by the SIC.

LASSO: In the presence of a large amount of regressors, an alternative to factor methods are shrinkage procedures, LASSO being one of the most popular ones. We use the SIC adapted to the LASSO (see Zou et

[5]This testing procedure accounts for the possibility that many series in the time series panel under investigation may be dependent on one another.

[6]Both restrictions are relaxed in the full WPV, but are found to worsen or at least not significantly improve forecast accuracy vis-à-vis the benchmark BEM. Furthermore, a data-driven, simultaneous selection of Google and survey indicators is included in the WPV (Section 5.7).

[7]The precise assignments of query terms for the initial two (ad-hoc) approaches as well as the pre-selection underlying the remaining methods are presented in Tables 4 and 5 of the WPV.

al., 2007) to determine the penalty parameter λ. To guarantee a large degree of shrinkage, we only consider λ-values leading to at most six non-zero coefficients in the model. As lagged Google observations are contained in the set of regressors, we perform variable and lag selection at the same time.

Adaptive LASSO: To address the potential inconsistency of the usual LASSO estimator, we also consider an adapted Lasso (AdaLASSO) version, where we weight the penalty term (by absolute OLS estimates) in order to penalize irrelevant variables to a higher degree than relevant ones (see, e.g., Smeekes and Wijler, 2016). All other settings are equivalent to the regular LASSO case.

Boosting: Another successful machine learning tool is Boosting, an iterative procedure starting off with a simple model, that is sequentially "boosted" by adding the series with most explanatory power at each step. We set the shrinkage parameter and stopping criterion to 0.1 and 250, respectively. In order for our variant to to function as a variable selection approach, we only select those Google (sub)categories that get chosen 20% of the times.[8]

4. Forecast Exercise – Setup and Results

To assess whether replacing survey indicators by Google search variables, selected via the approaches presented in the previous section, in the equations of the "hard" monthly indicators improves forecast accuracy, we conduct a forecast exercise in pseudo-real time, i.e, we mimic the regular routine of a forecaster while abstracting from eventual data revisions. We consider an increasing sequence of estimation samples starting from 1991:M1 - 2013:M6 and ending with 1991:M1 - 2016:M12. Although the evaluation period might appear rather short compared to the length of our sample, it is long enough for our forecast accuracy measures to be based on sufficient observations. Furthermore, both, the impact of the internet on our daily lives as well as the use of search engines such as Google, have increased over time, suggesting that the ability of Google search data to improve macroeconomic forecasts is more visible in the recent past.[9]

We follow standard practice at the Deutsche Bundesbank of synchronizing the timing of our forecasts with the publication of "hard" and "soft" indicators, implying two forecast dates per month: after the first and after the third week (Bundesbank, 2013). We stick to the publication calendar of the indicators, making sure that we never use data that would have not been available at the time. Depending on the moment at which we compute forecasts, we either obtain one nowcast and one forecast (e.g., Q1 and Q2 if "today" is March) or one backcast and one nowcast (e.g., Q1 and Q2 if "today" is April) for GDP (or a component thereof). This is due to the fact that GDP for quarter t is published with a delay of about six weeks, which then implies forecast horizons $h = -5, -3, \ldots, 17$, the latter being defined as the amount of weeks between "today" and the end of the reference period. As we need to forecast the monthly indicators until the end of the forecast period, we have $h = -7, -5, \ldots, 17$ in this case, where the publication delays of the series determine which horizons actually apply. Note that the outcomes for the monthly and quarterly series and a specific forecast horizon are not directly comparable.[10] Table 2 summarizes the GDP growth results for the comparison between the benchmark BEM in steps (1)-(3) and the augmented version described at the end of section 2. Figures represent relative root mean squared forecast errors (RMSFEs) of the latter compared to the former.

In fact, large gains are possible for forecasting and long-horizon nowcasts (i.e., $h > 3$); for back- and especially nowcasts, though, it is much harder to beat the benchmark BEM. Looking at the best-performing methods, PLS and LASSO appear to give the largest and most robust improvements overall.[11] Note that forecast accuracy gains can be tremendous in some instances, e.g., PLS achieve more than 45% for $h = 7$.

Even though suitably chosen Google search data seem to constitute a valid alternative to survey variables, we do not claim that they should replace survey variables in practice altogether. The validity and usefulness of indicators derived from surveys, that are specifically designed for various sectors of a macroeconomy, is well

[8]If no Google regressor surpasses the δM-barrier, we select the one being chosen most of the times.

[9]We investigate the robustness of our findings with respect to a longer evaluation period and a rolling window estimation as well, see Sections 5.5 and 5.6 in the WPV.

[10]Indeed, due to temporal aggregation, monthly forecasts with a specific horizon enter several quarterly forecasts; likewise, quarterly forecasts with a specific horizon depend on several monthly forecasts.

[11]The most robust ad-hoc selection method, subjectively chosen categories, is almost always outperformed by either PLS or LASSO (or both); the only exceptions are $h = 1$ and 3.

Table 2: Relative RMSFEs augmented vs. benchmark BEM for GDP growth forecasts

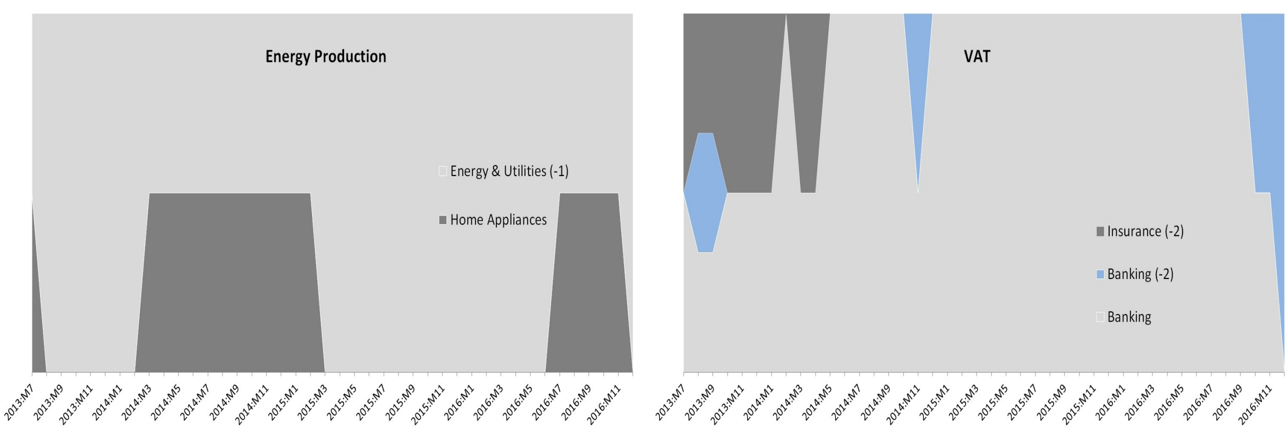
Method	Subj.-Cat.	Subj.-Subcat.	Google Corr.	PCA-Cat	PCA	PLS-Cat	PLS	LASSO	AdaLASSO	Boosting
Forecast Horizon										
17	0.82	0.91	0.83	0.84	0.78	0.72	0.71	0.74	0.74	0.78
15	0.83	0.91	0.84	0.86	0.81	0.73	0.74	0.74	0.74	0.77
13	0.90	0.98	0.91	1.02	0.91	0.85	0.89	0.78	0.78	0.85
11	0.85	0.93	0.86	0.96	0.88	0.84	0.88	0.81	0.81	0.85
9	0.83	0.89	0.79	0.93	0.83	0.69	0.67	0.84	0.84	0.86
7	0.73	0.78	0.71	0.80	0.73	0.57	0.54	0.83	0.83	0.80
5	0.79	0.89	0.77	0.89	0.77	0.68	0.65	0.87	0.87	0.89
3	0.99	1.24	1.09	1.18	1.01	1.12	1.13	1.02	1.02	1.12
1	1.04	1.27	1.12	1.22	1.07	1.21	1.25	1.06	1.06	1.15
-1	0.96	1.10	0.96	0.99	0.93	0.94	0.92	0.97	0.97	0.97
-3	0.95	1.11	0.96	0.96	0.92	0.94	0.91	0.98	0.98	0.98
-5	1.00	0.97	1.00	0.99	1.00	1.01	1.00	0.99	0.99	0.97

established and documented for various model specifications (beyond the example BEM we employ here), time periods, applications and so forth. On top of that, survey indicators are available for a longer period of time and are very transparent as to how they are obtained, guaranteeing a certain level of representativity and reliability. But the outcomes presented thus far at least point towards the potential of internet search data to contain information that is not embedded in survey variables.

Investigating the GDP component and monthly indicator forecasts[12] unveils that forecast improvements of Manufacturing, the by far biggest GDP component according to its weight in the NAs, seem to be mainly driving the good results for GDP growth. Quite often, though, there seem to exist mismatches between the forecasts of the x-indicators and the respective GDP components, whereby it should be recalled that the monthly and quarterly outcomes are not directly comparable.

To close this section, let us have a quick look at the Google series, that actually get selected by one of the most promising approaches, LASSO.[13] As examples, we focus on Energy Production and VAT, both recording improvements in forecasting performance using this approach. Figure 1 shows the results, whereby a color appearing as a vertical bar implies the corresponding subcategory to be selected. All in all, the outcomes appear quite intuitive; the "spot-on" subcategory Energy & Utilities proves useful for Energy Production and VAT is dominated by Banking

Figure 1: Google Variable Selection by LASSO



[12]Not shown here to save on space. We refer the reader to Table 12 in the WPV.

[13]Examples for PLS are discussed in Section 5.8 of the WPV.

5. Conclusions

In this paper we analyzed whether (data derived from) "big data" carry useful information for predictions of economic activity. In particular, we incorporated Google search data into a BEM for Germany to assess whether they can improve the GDP growth forecast performance. To address the crucial issue of which Google search terms to choose, we considered several variable selection approaches: ad-hoc, factor-based or of the shrinkage-type. We found that large forecast accuracy gains are possible, especially for fore- and late nowcasts, when replacing survey by Google variables in equations of the underlying "hard", monthly indicators. Hence, some evidence for Google data to be potential alternatives to survey variables was detected. This result, however, only partly extended to the underlying GDP components and monthly indicators.

One should keep in mind that the BEM we considered in this paper is merely an example. An interesting future analysis would be to incorporate Google data into alternative model specifications, e.g., a dynamic factor model. Other promising avenues would be an even tighter, or data-driven, Google variable pre-selection, the use of specific, tailored Google search terms instead of categorized versions and alternative representatives of internet data (e.g., *Tripadvisor* for the Hotel Industry). All in all, though, we feel confident in concluding that, although there are still many open issues and pitfalls with using internet search data, they surely show potential to improve macroeconomic forecasts.

References

Bell, V., Co, L. W., Stone, S. & Wallis, G. (2014). Nowcasting UK GDP growth. *Bank of England Quarterly Bulletin*, 54(1): 58-68.

Bundesbank (2013). Forecasting models in short-term business cycle analysis - a workshop report. *Deutsche Bundesbank Monthly Report*, September 2013, 69-83.

Destatis (2015). Nearly every second person over the age of 65 uses the internet. *Press release from 16 December, 2015*, Retrieved on 5 August, 2016.

ECB (2008). Short-term forecasts of economic activity in the euro area. *ECB Monthly Bulletin*, April 2008, 69-74.

Klein, L. R. & Sojo, E. (1989). Combinations of high and low frequency data in macroeconometric models. In *Advanced Studies in Theoretical and Applied Econometrics*, vol. 17, ch. 1, 3-16. Springer Netherlands.

Varian, H. R. (2014). Big data: new tricks for econometrics. *Journal of Economic Perspectives*, 28(2): 3-28.

Silvestrini & Veredas (2008). Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys*, 22(3): 458-497.

Smeekes, S. (2015). Bootstrap sequential tests to determine the order of integration of individual units in a time series panel. *Journal of Time Series Analysis*, 36(3): 398-415.

Smeekes, S. & Wijler, E. (2016). Macroeconomic forecasting using penalized regression methods. Research Memorandum 039, Maastricht University, GSBE.

Zou, H., Hastie, T. & Tibshirani, R. (2007). On the degrees of freedom of the lasso. *The Annals of Statistics*, 35(5): 2173-2192.