International Journal of Forecasting 35 (2019) 45–66

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Google data in bridge equation models for German GDP

Thomas B. Götz[a,*], Thomas A. Knetsch[b]

[a] Deutsche Bundesbank, Macroeconomic Analysis and Projection Division, Germany
[b] Deutsche Bundesbank, General Economic Statistics Division, Germany

ARTICLE INFO

Keywords:
Big data
Bridge equation models
Forecasting
Principal components analysis
Partial least squares
LASSO
Boosting

ABSTRACT

Interest in the use of "big data" when it comes to forecasting macroeconomic time series such as private consumption or unemployment has increased; however, applications to the forecasting of GDP remain rather rare. This paper incorporates Google search data into a bridge equation model, a version of which usually belongs to the suite of forecasting models at central banks. We show how such big data information can be integrated, with an emphasis on the appeal of the underlying model in this respect. As the decision as to which Google search terms should be added to which equation is crucial –- both for the forecasting performance itself and for the economic consistency of the implied relationships –- we compare different (ad-hoc, factor and shrinkage) approaches in terms of their pseudo real time out-of-sample forecast performances for GDP, various GDP components and monthly activity indicators. We find that sizeable gains can indeed be obtained by using Google search data, where the best-performing Google variable selection approach varies according to the target variable. Thus, assigning the selection methods flexibly to the targets leads to the most robust outcomes overall in all layers of the system.

1. Introduction

The internet has become a normal part of many people's life, at least in developed countries of the world. Gone are the days when people merely used it to send emails; now the web is used for buying products, booking hotels, banking, dating, research, reading the news, investing, social interactions and countless other things. The recent advent of the "sharing" culture, be it car or room sharing, only intensified the impact of the world wide web. In Germany, for example, nearly 85% of people above the age of 10 used the internet in 2015 (Destatis, 2015), though the rate diminished with age: of people below the age of 45, nearly everyone used the internet, while among people aged 45–64 years and over 65 years still about 90% and almost 50% respectively browsed the web in 2015 (+2% and +4% on the previous year, respectively).

Given that the internet is used so widely in our personal and professional lives, the question arises as to whether we are able to generate knowledge regarding macroeconomic activity from internet data. Luckily, advances in computer technology now enable researchers at companies or institutions not only to generate vast amounts of data, but also to process the non-standard, rather unstructured data that emerge from business and social activities on the internet and other platforms. We refer to such data here as "big data".[1] This paper investigates whether such big data — or, more specifically, data derived from them — lead to improvements in forecast accuracy as far as macroeconomic quantities are concerned. In light of the omnipresence of

* Corresponding author.
E-mail addresses: thomas.goetz@bundesbank.de (T.B. Götz), thomas.knetsch@bundesbank.de (T.A. Knetsch).

[1] It is common to characterize big data using the so-called "4 Vs definition" (volume, velocity, variety, veracity). Varian (2014) provides an interesting overview of tools for manipulating and analyzing big data in general. Furthermore, Nymand-Andersen (2016) provides insights regarding the use of big data in central banks for policy purposes. See also Einav and Levin (2013) and Diebold (2012), who provide complementary views on the use of big data in econom(etr)ics.

https://doi.org/10.1016/j.ijforecast.2018.08.001
0169-2070/

the web, we focus on the gross domestic product (GDP hereafter), and analyze the extent to which internet data can help to predict macroeconomic activity. More specifically, this paper employs Google Search data, a proxy for internet usage behavior, for Germany.[2]

While early work on the use of internet data for forecasting purposes was in the field of epidemiology (see e.g. Ginsberg et al., 2009, or Johnson et al., 2004), in recent years more and more work has been devoted to improving the forecast accuracy of various macroeconomic variables. Seminal contributions were made by McLaren and Shanbhogue (2011), who examined the use of internet data for the labour and housing markets, Choi and Varian (2012), who forecast automobile sales, unemployment claims, travel destination planning and consumer confidence, and Koop and Onorante (2013), who introduced Google probabilities as model-switching determinants within a dynamic model selection approach. Focusing on specific applications, Goel, Hofmann, Lahaie, Pennock, and Watts (2010), Toth and Hajdu (2012) and Vosen and Schmidt (2011, 2012) dealt with forecasts of consumption, Askitas and Zimmermann (2009), D'Amuri and Marcucci (2017) and Tuhkuri (2016) applied Google search data to unemployment forecasts, Guzman (2011) and Seabold and Coppola (2015) analyzed the case of inflation (expectations), Humphrey (2010) considered existing home sales, Kulkarni, Haynes, Stough, and Paelinck (2009) looked at housing prices, Pan, Wu, and Song (2012) investigated forecasts of hotel room demand, and Artola, Pinto, and de Pedraza García (2015) investigated forecasts of tourism inflows.

All of these studies have focused on a specific macroeconomic indicator, usually sampled at the monthly frequency; so far, no one has investigated the potential for Google search data to achieve forecast accuracy improvements for economic activity as a whole, i.e., GDP growth.[3] This paper intends to fill this gap in the literature by incorporating Google search data into a bridge equation model (BEM hereafter), one of the workhorse models that is used for short-term GDP forecasting in many central banks (see e.g. ECB, 2008, Bell, Co, Stone, & Wallis, 2014, or Bundesbank, 2013). Indeed, the model's simplicity, transparency and structure clearly lend themselves to the investigation of whether Google data improve forecasts of GDP, and if they do, through which channels. Furthermore, as the choice of the Google search terms to enter a given model often turns out to be crucial for the forecast performance, we investigate different (Google) variable selection approaches in terms of their out-of-sample forecast performances: principal component analysis (PCA hereafter), partial least squares (PLS hereafter), the least absolute shrinkage and selection operator (LASSO hereafter), boosting, and a couple of subjective (ad hoc) methods. In addition to the forecasting power of the resulting Google-augmented models (vis-à-vis the benchmark model without Google terms), we also pay attention to the specific Google search terms that are actually chosen by the best-performing variable selection methods over time.

It turns out that a crucial question for the forecast performance is whether Google data are included next to or instead of survey indicators. While the forecast gains are hardly detectable in the former case (forecast improvements are found for PLS and stem mainly from better predictions for the manufacturing sector), the latter yields many instances in which a Google-augmented model outperforms the benchmark. This finding points towards Google data potentially being an alternative to survey variables. On a more disaggregate level, the outcomes depend quite heavily on which variable selection method is employed, showing that just choosing a single approach across the entire system is suboptimal. As a consequence, assigning the selection procedures carefully to the underlying GDP components leads to forecast improvements for almost all horizons, and the most robust results overall. Thus, if incorporated into the model properly, Google search data can lead to more precise GDP growth forecasts.

The remainder of the paper is structured as follows. Section 2 introduces the BEM used for our analysis and illustrates how it is augmented by internet search data. The Google data themselves are described in Section 3. Section 4 presents the (statistical) methods for determining which Google search terms enter which equation of the BEM. The setup and outcomes of the forecast exercise are discussed extensively in Section 5. Section 6 provides concluding remarks.

2. The bridge equation models

Bridge equation models were introduced by Klein and Sojo (1989) as a regression-based system for GDP growth forecasting, where the different GDP components of the National Accounts are modelled individually. The equations for the individual GDP components are then augmented with short-term indicators that are tailored to the specific equations in question. Thus, intuitively speaking, the information contained in various short-term indicators gets transferred, or "bridged", to the coherent structure implied by the National Accounts (Wohlrabe, 2008). There exist many applications of BEMs in the literature, among which are those of Angelini, CambaMendez, Giannone, Reichlin, and Rünstler (2011), Baffigi, Golinelli, and Parigi (2004), Camacho, Perez-Quiros, and Poncela (2013), Foroni and Marcellino (2014) and Schumacher (2014), the last of which studies compares MI(xed) DA(ta) S(ampling) models (Ghysels, Santa-Clara, & Valkanov, 2004) and BEMs as competitive approaches to dealing with the mixed-frequency characteristic of many (macro)economic datasets.

2.1. The benchmark BEM

BEMs are characterized by dynamic linear equations, where GDP growth or some component thereof represents the (low-frequency) dependent variable. In our case, let

[2] Indeed, many online activities start off with a search engine: people search for a specific product they intend to buy, look for companies they may invest in or collect information on their next potential vacation destination. Of the available internet search engines, Google is clearly the dominant one, with a market share in Germany of about 95% (and just below 90% worldwide) in 2016 (Destatis, 2015, and Statista, 2016, respectively).

[3] Wiermanski and Wilshusen (2015) aim to predict GDP growth, but use consumer credit report data.

y_t denote the quarterly growth rate of a GDP component, e.g. Manufacturing,[4] in period t $(= 1, \ldots, T)$. In addition to low-frequency lags, the regressor set may also contain time-aggregated short-term (high-frequency) indicators. Here, let x_t^q denote the sole (for explanatory purposes) short-term, directly measurable (i.e., "hard"), monthly indicator, e.g. Industrial Production, time-averaged to a quarterly frequency (hence the superscript q). Then, the corresponding dynamic linear equation is just an autoregressive distributed lag (ADL hereafter) model:

$$y_t = \mu_y + \rho_y(L)y_{t-1} + \beta(L)x_t^q + \epsilon_t^y, \tag{1}$$

where $\rho_y(L) = \sum_{i=0}^{p-1} \rho_{y,i+1}L^i$ and $\beta(L) = \sum_{i=0}^{q} \beta_i L^i$, with L representing the usual lag operator, i.e., $L^i y_t = y_{t-i}$. Note that all series are transformed to stationarity whenever necessary.[5] Time aggregation of the underlying monthly indicator, x_t^m (note the superscript m), is undertaken using the average (or sum) of all monthly values (Silvestrini & Veredas, 2008).

Eq. (1) can be estimated using ordinary least squares (OLS hereafter), after which forecasts of GDP growth, say y_{T+h}, can be computed. However, doing so requires $x_{T+1}^q, \ldots, x_{T+h}^q$, i.e., forecasts of the time-averaged monthly indicator, which are obtained in two steps: first, we compute forecasts up to period $T+h$ using an ADL model specified at the monthly frequency:

$$x_t^m = \mu_x + \rho_x(L^{1/3})x_{t-1/3}^m + \delta_x(L^{1/3})z_t^m + \epsilon_t^x, \tag{2}$$

with $\rho_x(L^{1/3})$ and $\delta_x(L^{1/3})$ defined similarly to $\rho_y(L)$ and $\beta(L)$, but at the monthly frequency using the high-frequency lag polynomial $L^{1/3}$, for which $L^{1/3}x_t^m = x_{t-i/3}^m$. Note that fractions in the subscripts represent data points within the low-frequency period t, with $i = 0, 3, 6, \ldots$ corresponding to end-of-quarter observations.[6] Second, we time-aggregate the corresponding monthly forecasts to the quarterly frequency using one of the aforementioned aggregation schemes (depending on whether x is a flow or stock variable). The variables z are survey (i.e., "soft") indicators, which — due to their timeliness — should help predict the hard indicators, which are usually characterized by publication delays, especially in the short term. The soft indicators that are modelled using a simple AR process are:

$$z_t^m = \mu_z + \rho_z(L^{1/3})z_{t-1/3}^m + \epsilon_t^z, \tag{3}$$

where $\rho_z(L^{1/3})$ is defined straightforwardly.[7]

In summary, for a given soft indicator z, hard indicator x and GDP component y, the benchmark BEM procedure consists of the following steps:

Benchmark BEM procedure.

(a) Estimate Eq. (3); use the parameter estimates to obtain monthly forecasts of z until period $T+h$.
(b) Estimate Eq. (2); use the parameter estimates and forecasts of z to compute monthly forecasts of x until period $T+h$.
(c) Temporally aggregate x (including its forecasts) to the quarterly frequency.
(d) Estimate Eq. (1); use the parameter estimates and forecasts of x to obtain quarterly forecasts of y until period $T+h$.

As y corresponds to a specific GDP component, we repeat the procedure above for all components in question and compute a final GDP growth forecast as a weighted average according to the components' shares in the National Accounts.

This paper considers an adapted submodel of the full BEM that is run routinely for short-term forecasting at the Deutsche Bundesbank (see Bundesbank, 2013 for details).[8] As we are interested in the potential benefits of Google data for GDP growth forecasting, this paper considers only an example BEM: a disaggregated BEM that covers the production side of the German National Accounts. More specifically, y in Eq. (1) corresponds to one of the 15 GDP components listed in the left column of Table 1. As far as the choice of hard monthly indicators x is concerned, two criteria are considered: first, the indicator must be economically sound; second, it must have a statistically significant impact on the target variable in question. Based on these considerations and past experience, we chose the indicators listed in the second column of Table 1. Depending on the GDP component in question, the soft indicator z is taken to be either the ifo index that assesses the current business situation in trade and industry (ifo ind hereafter) or the purchasing managers index in services (pmi serv hereafter), as given in the third column of Table 1.[9]

2.2. Augmented BEM versions

Now that the benchmark BEM has been introduced, let us discuss how we incorporate time series derived from Google search data. Given the general structure of the BEM, i.e., starting by using soft, timely indicators to forecast hard, monthly indicators with publication delays, we propose to treat the (also timely) Google data similarly to these soft

4 We avoid the inflationary use of inverted commas by using capital letters whenever we refer to specific series. This applies to GDP components, monthly indicators ("hard" or "soft", see below) and Google data.

5 In the case of cointegration between the two series, as tested for using the Engle–Granger two-step procedure, the (log-)levels of the series are added to Eq. (1), yielding an error-correction format. It is known widely that failing to account for a long-run relationship in this case would imply losses in the short-term forecast accuracy (see e.g. Clements & Hendry, 1998).

6 Consequently, $x_{t-1/3}^m$ represents the value of x in the second month of quarter t, $x_{t-2/3}^m$ that in the first month, and $x_{t-3/3}^m \equiv x_{t-1}^m$ that in the third month of the previous quarter.

7 The orders of the low- and high-frequency lag polynomials are usually determined via standard information criteria. Furthermore, the forecasts of x and z could be based on any other model just as well. However, ADL or AR models are often applied so as to keep the BEM simple and transparent.

8 The data set was downloaded from the internal database of the Deutsche Bundesbank and is generally not publicly available; however, if one wants to replicate these results, the vintage of data used in this paper can be provided upon request.

9 Note that some GDP components do not have a hard indicator assigned to them (e.g., Housing). In these cases, step (b) of the procedure is skipped and x is replaced by z in Eq. (1), as well as in steps (c) and (d) above.

Table 1
The disaggregated production-side bridge equation model.

GDP component (y)	Hard indicator (x)	Soft 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[a]	ifo ind
Hotel Industry	Sales Hotel Industry	ifo ind
Net taxes	Value-added Tax (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

Note: Although not displayed, Eq. (2) is often augmented by variables that capture the effects of bridge and vacation days, as well as weather conditions, which proves useful for some of the x-variables. These variables are either pre-determined (bridge and vacation days) or extrapolated using historical means (assuming for example a "normal" winter).
[a] The indicator Toll is only available from 2007 onward, so if we look back too far within our forecast evaluation (see Section 5) we use Industrial Production instead, to ensure that the estimation is reliable.

indicators. Indeed, most papers in the literature focus on a specific macroeconomic indicator, such as consumption (Vosen & Schmidt, 2012), that intuition suggests might be explained by specific Google search terms. After all, users are more likely to search for "jobs", "used car" or "last-minute holiday offers" than "GDP", for example. Thus, it seems natural to augment the regression models of the x-indicators rather than the equations of the corresponding quarterly GDP component. Hence, in the first augmented BEM version, Eq. (2) gets altered to:

$$x_t^m = \mu_x + \rho_x(L^{1/3})x_{t-1/3}^m + \delta_x(L^{1/3})z_t^m + \gamma_x(L^{1/3})g_t^m + \epsilon_{a,t}^x, \tag{2a}$$

where g^m represents some monthly Google time series; details on the latter are provided in the next section.

Modelling the hard indicators x in this way implies that Google data enter alongside the soft indicators already included in the benchmark model; or, alternatively, are added instead of the soft indicators. Consequently, the second augmented BEM version contains the following equation for x:

$$x_t^m = \mu_x + \rho_x(L^{1/3})x_{t-1/3}^m + \gamma_x(L^{1/3})g_t^m + \epsilon_{b,t}^x. \tag{2b}$$

Before we can adapt the BEM procedure from before, we need to specify a model for the Google series. In the spirit of handling g the same way as z, we use an AR model to extrapolate the Google series:[10]

$$g_t^m = \mu_g + \rho_g(L^{1/3})g_{t-1/3}^m + \epsilon_t^g. \tag{4}$$

Now, for some given Google series g and a given soft indicator z, hard indicator x and GDP component y, the augmented BEM procedure reads as follows.

Augmented BEM procedure.

(a) Estimate Eqs. (3) and (4); use the parameter estimates to obtain monthly forecasts of z and g until period $T + h$.
(b) Estimate Eq. (2a) or (2b); use the parameter estimates and forecasts of g (and z if necessary) to compute monthly forecasts of x until period $T + h$.
(c) Temporally aggregate x (including its forecasts) to the quarterly frequency.
(d) Estimate Eq. (1); use the parameter estimates and forecasts of x to obtain quarterly foreca sts of y until period $T + h$.

Again, we subsequently obtain a final GDP growth forecast by computing a weighted average of all GDP component forecasts for which the above procedure is repeated.

Due to the fact that we once add Google series to the equation for x on top of soft indicators and once use Google series instead of soft indicators, we deal with two augmented BEM versions. Thus, Table 2 summarizes the three models we that consider and defines labels for the remainder of the paper.

Note that we only allow Google series to enter the equations for the hard indicators x. In cases where a GDP component does not have an x-indicator assigned to it (see Table 1), the corresponding soft indicator remains the sole regressor (apart from lags of y) that affects its forecasts, no matter which augmented model version we consider.[11]

[10] The lag order is also chosen via information criteria. We considered alternative models, e.g. vector autoregressions on different groups of Google series, but the results were less promising. Shimshoni, Efron, and Matias (2009) analyzed the predictability of Google data and found that accounting for seasonality proves very beneficial. In future, instead of seasonally adjusting the Google series a priori, we could consider specifications that model a seasonal component explicitly.

[11] Again, step (b) of the procedure is skipped and x is replaced by z in Eq. (1), as well as in steps (c) and (d). As a consequence, the forecasts of the GDP components without hard indicators are unaffected by any of the Google augmentations we consider. This means that we could disregard those equations altogether in our analysis, but we chose to leave them in the routine for the sake of covering the entire span of the National

Table 2
The three models considered in this paper.

Label	Model description
BM	Benchmark model based on Eq. (2): z appears in the equation for x
AM-gz	Augmented model based on Eq. (2a): g and z appear in the equation for x
AM-g	Augmented model based on Eq. (2b): g appears in the equation for x

Table 3
The categories of Google data used.

ECB Google data: categories		
Autos & Vehicles	Beauty & Fitness	Business & Industrial
Computers & Electronics	Finance	Food & Drink
Health	Home & Garden	Internet & Telecom
Jobs & Education	Law & Government	News
Real Estate	Sensitive Subjects	Shopping
Sports	Travel	

3. Google data

The Google search data that we employ in this paper stem from a data set that Google provides to the European Central Bank (ECB hereafter) on a weekly basis.[12] The data are available as of 2004 and appear without any publication delay. The data set comprises query searches of keyword categories; i.e., it measures the total number of searches for a particular category relative to all search queries. Hence, only relative changes in search volumes can be assessed, not absolute search volumes (Koivupalo, 2014). This is crucial, because a search term may have a higher relative search volume in a certain time period while having a lower absolute number of searches. The data do not get revised, but are based on random samples from all Google search queries during a day. Although the weekly data constitute an average over the corresponding seven consecutive days, the data change slightly whenever an updated data set is considered. The data are normalized to start at one (so that other figures indicate deviations from the starting value), and are greater than or equal to zero (where the latter represents query numbers that fall short of Google's privacy filter).[13]

We average the Google data over time so as to match them with the monthly frequency of our indicators in the BEM, where we assign each week to the month into which most of its days fall. As the Google data are not seasonally adjusted (in contrast to the macroeconomic variables), we apply the ARIMA-X12 approach (instead of relying on year-on-year growth rates). As far as the order of integration is concerned, we use the bootstrap sequential quantile test (BSQT) of Smeekes (2015) to check for unit roots in a time series panel,[14] in which many series may be dependent on each another. We use ten equally-spaced quantiles for our panel of a total of 200 Google search series (see below). The BSQT actually returns zero rejections of a unit root (i.e., a zero proportion of $I(0)$ series), such that we compute first differences of all Google search variables in our dataset.

The data cover 14 different countries of the European Union, but we focus on the ones for Germany. Furthermore, the various search terms are allocated into 26 categories (see Table 3) and, for a finer distinction, 269 subcategories (Table B.1 in Appendix B). However, we disregarded a priori nine categories that we deemed unfitting (Arts & Entertainment, Books & Literature, Games, Hobbies & Leisure, Online Communities, People & Society, Pets & Animals, Reference and Science), as well as several subcategories from Sensitive Subjects that were particularly prone to outliers or zero-values. In the end, our Google data set consisted of 200 series.

The timeliness of the Google data means that we may expect forecast accuracy gains from including them, particularly for short forecast horizons; say, for now- and one-quarter-ahead forecasting. For backcasts, though, the Google series may not add much information, as most of the relevant indicators will have been published over the reference period.

Fig. B.1 (in Appendix B) provides a glimpse of their forecasting potential by displaying two of the macroeconomic indicators whose equations are augmented with Google series, namely Real Retail Sales (incl. cars) and Sales Hotel Industry, together with two intuitively "fitting" Google time series, namely Autos & Vehicles and Hotels & Accommodations, respectively. In both cases, the development of the Google series is very similar to that of the respective macroeconomic indicator; in fact, it looks as if the former leads the latter by one month. For the troughs in particular, the Google data appear to be a promising leading indicator.

Accounts. However, we also tried to assess whether adding Google series g to these equations would improve the forecast accuracy. Unfortunately, we discovered that the corresponding GDP components are very sensitive to the addition of Google variables, which is why we opted for the augmentations introduced here. For details of these analyses, we refer the reader to the working paper version of the manuscript (Götz & Knetsch, 2017).

12 For confidentiality reasons, the dataset is not available publicly. The provision of the latest-available vintage — for the entire dataset, actually — may be possible upon request.

13 In contrast to the ECB data, data obtained from the Google Trends application *Insights for Search* start with a value of zero, have the maximum query share over a specific time period normalized to 100, and cover more countries as well as deeper levels of categories (for more information on the publicly-available Google data and their informative contents, see Bontempi, Golinelli, & Squadrani, 2016). Crucially, though, the random samples on which the data are based are much smaller, and thus, the ECB data, which we judge to be sufficiently granular, should be more accurate.

14 We use the openly-available GAUSS code from Smeekes (2015).

Note also, though, that the time series show some disconnection in the early years of the sample period, which may have to do with Google Trends data being available only from 2004 onward.

4. Google variable selection

Now that we have introduced the model and the data, it is time to discuss how we choose which Google (sub)categories enter which equation of the BEM. Using the notation in Section 2, the aim of this section is determine which series $g \in G$ are chosen to enter which x-equation, where G represents the entire set of 200 Google series described in the previous section. Naturally, including all of the candidate series is neither practical nor feasible, given the number of different (sub)categories. Also, even if there is a statistically significant relationship between a macroeconomic indicator and a Google search term, the relationship may turn out to be unjustifiable from an economic perspective; something we may label a "spurious relationship". Finally, a Google search term that might be expected intuitively to help in forecasting a given indicator might in fact not have a beneficial effect, due to either the low popularity of the search term or adverse search behaviors.[15] Consequently, the issue of which Google variable to select may be a subtle one, with a potentially large impact on the forecast performances of the augmented models. The remainder of this section describes three classes of Google variable selection methods that we considered.

4.1. Ad-hoc methods

We employ two ad-hoc approaches, a subjective selection and one based on Google Correlate. For the former, we choose the Google search data, once on a category- and once on a subcategory level, by hand, i.e., based on "common sense". The corresponding assignment of query terms is presented in Table B.2. For the latter, we use the Google Trends tool Google Correlate, which allows a user to search for queries that follow a pattern similar to a specific target series. In our case, we choose each hard indicator x in turn and let Google Correlate determine the search queries with the time series that possess the largest correlation coefficients. By shifting the target series by several time periods, we can inspect the relevance of lagged search terms as well. Subsequently, we manually (i.e., based on "common sense" again) filter out the search terms that suggest "spurious relationships" and then manually (i.e., based on "common sense" again) look for those "non-spurious" (sub)categories in our data set that correspond as closely as possible to the queries obtained using Google Correlate. The resulting selection is summarized in Table B.3.

4.2. Factor methods

It has become quite common in the literature to address the dimensionality problem that one is confronted with when forecasting economic time series in a data-rich environment by imposing a factor structure on the regressors. Summarizing the information that is present in the usually vast number of predictors using factor techniques allows a user to balance the trade-off between exploiting as much information as possible and minimizing the number of parameters to be estimated.[16] We use two methods for extracting common factors: PCA (see, e.g., Forni, Hallin, Lippi, & Reichlin, 2005, or Stock & Watson, 2002; or, for an application to forecasting with Google series, Vosen & Schmidt, 2012) and PLS (originally introduced by Wold, 1985; see also Groen & Kapetanios, 2016, or Cubadda & Guardabascio, 2012). Technical details and information on the numbers of factors and lags are provided in Appendix A.

We attempt to a priori avoid the factors loading on nonintuitive Google search terms, i.e., "spurious factor loadings", by restricting the set of eligible Google variables for a given x-series. More precisely, for a given monthly indicator or GDP component, we allow the factors to load only on subcategories that correspond to the categories that we considered for the subjective approach outlined above (the categories in the middle column of Table B.2).[17] Note that all of the other data-based selection procedures discussed below are based on the same pre-selection.[18]

Two versions of each factor method are considered. First, we compute the factor loadings unrestrictedly over all subcategories that "survive" the aforementioned pre-selection. Second, we group the eligible subcategories and subsequently draw category-specific factors. The reason for considering the latter is economic interpretability: it may be more intuitive to have, e.g., separate Autos-&-Vehicles-factors and Business-&-Industrial-factors for Industrial Production, than various factors loading on a mixture of all the corresponding subcategories. In addition, the resulting category-specific factors can be seen as data-driven alternatives to the Google categories, which, in contrast, are obtained using a data-driven criterion.

4.3. Shrinkage methods

Relying on its documented combination of statistical accuracy for prediction and variable selection, and computational feasibility (see e.g. Bühlmann & van de Geer, 2011, or Gasso, Rakotomamonjy, & Canu, 2009), we consider the least absolute shrinkage and selection operator

[15] For example, in recent times, Vehicle Brands may have been looked up by either potential car buyers or people investigating the emission scandal affecting many car brands.

[16] Using the notation in Section 2, we obviously select not $g \in G$, but linear combinations of all $g \in G$, say f^g.

[17] Such pre-selection is known to be quite beneficial for forecasting (Girardi, Golinelli, & Pappalardo, 2017), and is therefore rather common in the literature: for example, Vosen and Schmidt (2011, 2012) select 56 and 41 consumption-relevant categories respectively, before extracting factors from them.

[18] A data-driven alternative to the ad-hoc pre-selection of Google categories is the so-called *group lasso* (Hastie, Tibshirani, & Wainwright, 2015), which performs shrinkage on a group of variables instead of on each variable individually, and thus, lends itself to hierarchical or grouped data structures. One can even go one step further in the form of the *sparse group lasso* and apply shrinkage simultaneously to the members of the selected groups. However, as nothing prevents the group lasso from selecting nonintuitive, i.e., spurious, categories, which would make the selection difficult to interpret economically, we stick to the manual pre-selection described here.

Table 4
The selection methods we consider.

Label	Selection method
Subj-Cat	Subjectively chosen categories (ad hoc method, see Table B.2)
Subj-Subcat	Subjectively chosen subcategories (ad hoc method, see Table B.2)
Google-Corr	(Sub)categories based on Google Correlate (ad hoc method, see Table B.3)
PCA-Cat	PCA with category-specific factors (factor method, see Appendix A)
PCA	PCA with unrestricted factors (factor method, see Appendix A)
PLS-Cat	PLS with category-specific factors (factor method, see Appendix A)
PLS	PLS with unrestricted factors (factor method, see Appendix A)
LASSO	Usual version of LASSO (shrinkage method, see Appendix A)
AdaLASSO	Adapted version of LASSO (shrinkage method, see Appendix A)
Boosting	Boosting selection procedure (shrinkage method, see Appendix A)

(LASSO) proposed by Tibshirani (1996). In situations where the number of predictors far exceeds the time dimension, as is the case with our Google series here, the usual OLS estimator fails (its variance tends to infinity) and it overfits the model greatly. The idea of LASSO is to regularize the complexity of the model by adding a penalty term, thereby assigning non-zero coefficients to only a few variables and shrinking the remaining ones to zero. Hence, LASSO performs estimation and variable selection at the same time. We compute two versions of the LASSO: the usual one and an adapted one, to address potential inconsistencies (see e.g. Zou, 2006, or Konzen & Ziegelmann, 2016).

Finally, we also propose a variant of the boosting approach, which originates from the machine learning community (see Schapire, 1990, or Freund, 1995). Recently, the forecast performance of boosting has been analyzed in several forecast studies (among others Lehmann & Wohlrabe, 2016, or Buchen & Wohlrabe, 2011, 2014), marking it a competitive alternative to existing approaches in data-rich settings. It is an iterative procedure that starts off with a simple model which is then improved, i.e., "boosted", sequentially by adding the series with most explanatory power at each step. As before, technical details as well as information on key inputs of the shrinkage methods are provided in Appendix A, where we also explain how the boosting procedure is adapted to form a variable selection device.[19]

4.4. Overview

Table 4 summarizes all of the selection approaches that we consider and assigns labels to them for future reference.

Finally, note that we construct PLS factors for and apply shrinkage methods to the Google variables only, excluding from these manipulations lags of the respective target variable or the soft indicators.

5. Forecast exercise

5.1. Setup

Being equipped with all ingredients, it is now time to perform a forecast exercise in order to assess whether including Google data either next to (model AM-*gz*) or instead of (model AM-*g*) soft indicators in the equations for the hard indicators improves the prediction accuracy for GDP growth, its various components and the underlying hard indicators themselves. We conduct the analysis in pseudo-real time; i.e, we mimic the regular routine of a forecaster while abstracting from eventual data revisions.[20] Table 5 provides an overview of various data characteristics: availability, characteristics, transformations and publication delay.

Having downloaded the data at the end of December 2016, we consider an increasing sequence of estimation samples, starting with 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 a sufficient number of observations. Furthermore, as was outlined in the introduction, the impact of the internet on our daily lives has increased over time. As a larger, and more representative, share of the population starts to use the internet regularly every year, it may be that the ability of Google search data to improve macroeconomic forecasts has grown more visible in the recent past. Also, technically, given that Google search data are only available as of 2004 and Google is known for tinkering regularly with its search algorithms, forecasts that correspond to longer evaluation periods could suffer from higher levels of estimation uncertainty and more breaks in the data.[21]

We synchronize the timing of our forecasts with the bi-weekly publication frequency of our hard and soft indicators. As a consequence, we compute forecasts after the first and third weeks of each month, and label them "early" and "late", respectively. Note that we stick to the publication calendar of the indicators, making sure that we

19 We also considered ridge and elastic net regressions (the latter being a linear combination of LASSO and ridge), but they both performed worse than LASSO, which is why we do not list or discuss them here.

20 The reason why we do not consider a real-time data set is that the data vintages for some of the series do not reach far enough back into the past. Although we assume the results to be rather robust to the presence of revisions, it would be an interesting sensitivity analysis to perform in the future.

21 In light of these arguments, we also investigated the robustness of our findings with respect to (i) a longer evaluation period and (ii) estimation on a rolling window. It turned out that the outcomes either do not differ much qualitatively from those presented here, or they do so in the anticipated way: (i) soft indicators tend to have a greater relevance in the more distant past, and (ii) the increased reliability associated with larger estimation samples seems to outweigh the effects of the eventual structural changes. For the sake of brevity, we abstain from showing or discussing the results here; the interested reader is referred to Sections 5.5 and 5.6 of the working paper version (Götz & Knetsch, 2017).

Table 5
Data features.

Variable	Start	Characteristics & transformations	Publ. delay
15 GDP components	1991:Q1	in chained prices of previous years, 2010 = 100, sca, log-diff.	6
Production Mining	1991:M1	2010 = 100, sca, log-diff.	5
Industrial Production	1991:M1	2010 = 100, sca, log-diff.	5
Energy Production	1991:M1	2010 = 100, sca, log-diff.	5
Production in Construction	1991:M1	2010 = 100, sca, log-diff.	5
Real Retail Sales (+ cars)	1994:M1	in constant prices, 2010 = 100, sca, logs	9
Toll	2007:M1	in kilometres, sca, log-diff.	7
Sales Hotel Industry	1994:M1	in current prices, 2010 = 100, sca, log-diff.	7
VAT	1991:M1	in millions DM/EUR, sa, log-diff.	9
ifo ind	1991:M1	2005 = 100, sca, logs	−1
pmi serv	1997:M6	in percentages, sa, logs	−1
All Google series	2004:W1	in query shares, 2003 = 1, sa	0

Notes. sca: seasonally and calender-adjusted; sa: seasonally adjusted; Q: quarter; M: month; W: week. Assuming four weeks per month, the publication delay is defined as the number of weeks between the moment at which a variable gets published and the end of the reference period; i.e., Industrial Production in June 2016 gets published five weeks after the end of June, that is, one week into August. Note that a monotonic transformation is applied to the survey indicators so as to ensure positivity of the entries.

never use data that would have not been available at that time. As our focus is on short-term GDP forecasting, i.e., up to the next quarter, we determine the forecast period by the publication dates of GDP. More precisely, given the publication delay of German GDP (see the final column of Table 5), new "rounds" of forecasts always start late in M2, M5, M8 and M11. Let us go through one of these "rounds", the one starting late 2014:M2 say, to illustrate the inherent forecast horizons h as far as y, x and z (or g) are concerned, and whether we deal with now-, fore- or backcasts (NC, FC or BC hereafter). Similarly to the publication delay in Table 5, we define the forecast horizon as the number of weeks between the moment when we make the forecast and the end of the reference period. For y we have:

- Late 2014:M2 $\Rightarrow \hat{y}_{2014:Q1}$ (NC; $h = 5$), $\hat{y}_{2014:Q2}$ (FC; $h = 17$)
- Early 2014:M3 $\Rightarrow \hat{y}_{2014:Q1}$ (NC; $h = 3$), $\hat{y}_{2014:Q2}$ (FC; $h = 15$)
- Late 2014:M3 $\Rightarrow \hat{y}_{2014:Q1}$ (NC; $h = 1$), $\hat{y}_{2014:Q2}$ (FC; $h = 13$)
- Early 2014:M4 $\Rightarrow \hat{y}_{2014:Q1}$ (BC; $h = -1$), $\hat{y}_{2014:Q2}$ (NC; $h = 11$)
- Late 2014:M4 $\Rightarrow \hat{y}_{2014:Q1}$ (BC; $h = -3$), $\hat{y}_{2014:Q2}$ (NC; $h = 9$)
- Early 2014:M5 $\Rightarrow \hat{y}_{2014:Q1}$ (BC; $h = -5$), $\hat{y}_{2014:Q2}$ (NC; $h = 7$)
- Late 2014:M5 $\Rightarrow y_{2014:Q1}$ was published; the next "round" starts...

The situation for x, z and g is complicated by the ragged edge feature of the dataset. As was explained in Section 2, we need to forecast any figures of x, z and g that are not available over the forecast period. Hence, (generically) for x we have:

- Late 2014:M2 $\Rightarrow \hat{x}_{2014:M1}, \ldots, \hat{x}_{2014:M6}$ (1 BC, 1 NC, 4 FC; $h = -3, 1, 5, 9, 13, 17$)
- Early 2014:M3 $\Rightarrow \hat{x}_{2014:M1}, \ldots, \hat{x}_{2014:M6}$ (2 BC, 1 NC, 3 FC; $h = -5, -1, 3, 7, 11, 15$)
- Late 2014:M3 $\Rightarrow \hat{x}_{2014:M1}, \ldots, \hat{x}_{2014:M6}$ (2 BC, 1 NC, 3 FC; $h = -7, -3, 1, 5, 9, 13$)
- and so forth...

Obviously, figures that already became available do not need to be forecast; for example, ifo ind and pmi serv get published late in the respective quarter, implying that we never compute backcasts for these series ($\hat{z}_t = z_t$ then).[22]

Quite importantly, note that the outcomes for the monthly and quarterly series at a specific forecast horizon are not comparable directly. Due to temporal aggregation of the monthly series in step (c) of the BEM procedures, monthly forecasts with a specific horizon enter several quarterly forecasts: for example, monthly ($h = 9$) forecasts enter the equations for quarterly ($h = 9$), ($h = 13$) and ($h = 17$) forecasts. Likewise, quarterly forecasts with a specific horizon depend on several monthly forecasts: for example, quarterly ($h = 9$) forecasts are obtained using ($h = 1$), ($h = 5$) and ($h = 9$) forecasts.

The figures in the following section represent relative root mean squared forecast errors (RMSFEs) of an augmented BEM (AM-*gz* or AM-*g*) compared to the benchmark system (BM). Thus, values larger than one favour the model without any internet data, whereas values smaller than one indicate that the respective augmentation by Google series improves the forecast accuracy.

5.2. AM-gz vs. BM: Google next to survey indicators

We start off by putting the model AM-*gz* to the test; i.e., we add Google data to the soft indicators that already appear in the equations for the hard indicators. In other words, we investigate whether internet data provide valuable information beyond that contained in the survey indicators we consider. Table B.4 in Appendix B contains the outcomes with respect to GDP growth, i.e., the weighted average of our 15 GDP components according to their share in the National Accounts, for the various Google variable selection approaches summarized in Section 4.4. As the outcomes for AdaLASSO turn out to be almost identical to those for the usual LASSO (i.e., inconsistencies do not appear to be an issue for the latter), we do not show them here to save on space.[23]

[22] Given the publication delays of the variables under consideration, we actually never obtain forecast horizons smaller than −7.

[23] However, the complete set of results can be inspected in the working paper version (Götz & Knetsch, 2017).

Focusing first on the ad hoc variable selection methods in the first three columns of Table B.4, we ascertain that, by and large, the forecast accuracy of the augmented models is the same as that of the benchmark model. Indeed, the relative RMSFEs hover around one, with a few instances where AM-*gz* outperforms the benchmark (e.g. short-horizon nowcasts for Subj-Cat) and a few instances where BM dominates (e.g. $-3 \leq h \leq 3$ for Subj-Subcat). All in all, the internet data selected either subjectively or by means of Google Correlate do not seem to add new information on top of the soft indicators that already appear in the x equations.

Turning to the factor methods, the outcomes for the PCA variants are mostly similar: only near-term nowcasts ($h = 1,3$) seem to gain somewhat from the augmentation. For the PLS versions, though, particularly the unrestricted one, the situation looks different: the Google-augmented model delivers more precise GDP growth now- and forecasts for $h \geq 5$, with improvements of up to 14%. Backcasts also seem to benefit, but near-term nowcasts apparently present a challenge. Somewhat disappointingly, none of the shrinkage methods are able to generate competitive back- or nowcasts; in fact, the addition of Google series to the model often harms the forecast performance. Only the forecasts appear to be slightly better off from the model augmentation.

All in all, adding Google data to survey indicators in the equations for the hard indicators hardly leads to any forecast accuracy improvements, with the only exception being PLS, and even then the relative RMSFEs rarely fall below 0.9. It seems that the Google series do not add enough new, unique information for the augmented model to result in more precise GDP growth forecasts. However, let us dig deeper into the results for the well-performing selection method PLS, and inspect the forecasting results for the GDP components individually as well as for the underlying hard, monthly indicators. Table B.5 in Appendix B contains the respective set of outcomes.

Focusing first on the outcomes for the GDP components, it emerges that the good results for GDP growth stem mainly from improved nowcasts and forecasts of Manufacturing, the biggest GDP component by far according to its weight in the National Accounts. There are further instances of forecast improvements, but they mostly refer to forecasts: Hotel Industry, Mining and Trade (incl. cars). However, there are also many instances of forecast accuracy losses; for example, Net Taxes, as well as (almost always) Traffic and Construction. Relating the figures for the monthly indicators to their corresponding GDP components sometimes gives the puzzling picture of a better (or worse) forecast performance for the former not being accompanied by an equally good (or bad) performance for the latter. Retail Sales and Trade, as well as Energy Production and Energy & Water (for nowcasts and especially backcasts) are two such cases. However, one should keep in mind that the outcomes for the monthly and quarterly series and a specific forecast horizon cannot be compared directly. Furthermore, temporal aggregation of the x-variables may impact the quarterly forecasts of the GDP components in a more or less fortunate way. It could, however, also point towards a disentanglement of a GDP component and its assigned monthly indicator.

5.3. AM-g vs. BM: Google instead of survey indicators

Rather than investigating whether Google search data can provide additional information when used on top of survey indicators, let us analyze the situation in which we include them instead of survey variables. Using the notation of Section 2, we now compare model AM-*g*, where the equation for a hard indicator x is specified as in Eq. (2b), with the benchmark model BM. By excluding soft indicators z from the regressor set at this stage, we check whether the Google indicators carry unique information that is obscured when g is added on top of z, perhaps due to correlation effects. Put somewhat differently, we assess whether the Google and survey indicators in question are substitutes rather than complements in the equations for x. Table B.6 contains the outcomes with respect to GDP growth.

A most noteworthy finding was that the outcomes improved relative to Table B.4 in 80% of instances, with the vast majority of relative RMSFEs being clearly below one. Large gains (about 20% on average but up to 46%) are possible for forecasting and medium- to long-horizon nowcasts (i.e., $h \geq 5$), while backcasts also seem to benefit from using Google instead of survey variables, albeit to a much lesser degree. However, near-term nowcasts continue to present a challenge for Google search data. Indeed, the sometimes-large relative RMSFEs point towards the Google data missing some crucial information that is contained in the survey variables. Overall, though, the results point towards a great potential of Google search data to improve short-term GDP growth forecasts.

We determine whether the forecast accuracy improvements that are observed widely over Table B.6 are statistically significant by conducting a Diebold and Mariano (1995) test (DM test hereafter) for each pairwise comparison; as model AM-*g* is not nested in model BM, the use of the DM test is valid in this setting.[24] Note though the results have to be treated with care, as the number of forecast errors at our disposal for a specific horizon is rather small, and both size distortions and losses of power are known to be quite common in such situations (see e.g. Harvey, Leybourne, & Newbold, 1997, or Clark, 1999).[25] That being said, though, whenever the forecast performances of models AM-*g* and BM differ significantly from each other, the corresponding figure in Table B.6 is highlighted in bold.[26]

24 See Clark and McCracken (2001) and McCracken (2007) for tests of equal forecast accuracy (the latter including derivations of asymptotic distributions) between two nested models.

25 The reason for the small samples is that the forecast "rounds" (see Section 5.1), which each contain each of the 12 horizons under consideration exactly once, are run at a quarterly frequency. We tried alleviating this problem by considering the modification of Harvey et al. (1997), but the results did not differ much. We account for two further issues that may occur in small samples (see Clark, 1999), namely (i) serial correlation patterns differing from their theoretical ones, i.e., an h-step-ahead error should follow an MA($h-1$), and (ii) potentially not positive semi-definite variance estimates, by relying on a HAC estimator using the Bartlett kernel and a corresponding bandwidth selection (Newey & West, 1994).

26 Instead of differentiating between various levels of significance (using numbers of stars as labels), which may impair the readability of the table, bold figures summarize all DM tests that yielding p-values smaller than 0.1.

The outcomes of the DM tests mostly confirm what might have been expected from the figures themselves: the two competing models perform almost always equally well when backcasting, short-horizon nowcasts favor the benchmark model (if any), and long-term nowcasts and forecasts ($h \geq 5$) are often significantly more accurate when using the augmented model. The latter two conclusions are spread somewhat unevenly over the different variable selection methods: BM is a dominant strategy (in the sense that it is never beaten) if Google variables are selected using Subj-Subcat or PCA-Cat. In contrast, AM-g is the dominant choice for Subj-Cat, Google-Corr, PCA, LASSO and both PLS variants, where the PLS variants yield the most convincing scenarios (the augmented model outperforms the benchmark significantly for all $h \geq 5$). For boosting, the dominant model depends on the forecast horizon. On the whole, the tests for equal forecast accuracy underline the conclusions drawn from the RMSFEs themselves.

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 altogether in practice, even given their apparently superior performance as far as near-term nowcasts are concerned. The validity and usefulness of indicators derived from surveys that are designed specifically for various sectors of a macroeconomy is well established and documented for various model specifications (beyond the example BEM we employ here), time periods, applications, and so forth. In addition, survey indicators are available for longer periods of time and are very transparent as to how they are obtained, guaranteeing a certain level of representativity and reliability. However, the outcomes presented here at least point towards the potential for internet search data in general, and Google indicators in particular, to contain information that is not embedded in survey variables and that could prove beneficial for macroeconomic forecasting. This potential can be expected to increase in future, as, on the one hand, the increasing length of Google time series makes estimation more reliable, and, on the other, even more people use the internet, making the data themselves more informative.

Let us now inspect the more disaggregate set of results again, namely the GDP component and hard indicator forecasts. As the outcomes for PLS lead to the most robust outcomes overall, we only show the full set of results for this variable selection method; Table B.7 contains the outcomes in the same spirit as Table B.5, with statistically significant DM tests again being highlighted in bold. As before, improved nowcasts and forecasts of Manufacturing and (to a lesser degree) Mining and Hotel Industry appear to be driving the good results for GDP growth. For these three GDP components, one can often also link the gains in forecast performance to the corresponding monthly indicators, while this is less the case for others (e.g., Toll and Traffic). In general, it seems as if the "best" variable selection method depends on the monthly indicator (and GDP component) in question: though inferior to the PLS variants as far as GDP growth in Table B.6 is concerned, Subj-Cat in fact yields quite robust outcomes for Production Mining and Retail Sales, while LASSO performs better than its competitors for Energy Production, Toll and Sales Hotel Industry. For Construction and Net Taxes, however, refraining from any Google augmentation seems to be the optimal choice.[27] Consequently, the ultimate forecast performance is quite sensitive to the way in which the Google data are selected, implying that careful and competent monitoring by the researcher is unavoidable.

5.4. Augmented model with targeted selection

In view of the fact that several different variable selection methods, including not augmenting the respective hard indicator equation at all, deliver the best results at the disaggregate level, the question arises as to the performance of an augmented model that uses the optimal selection method for each GDP component (and respective hard indicator) instead of the using same method for all of them. We attempt to get an idea of the forecast performance of such an approach with respect to GDP growth by assigning the Google variable selection methods to the eight hard indicators based on their disaggregate performances in the models BM, AM-gz and AM-g. Hence, for each hard indicator we determine whether Google indicators are included next to or instead of soft indicators, if they are added at all.[28] Table 6 gives an overview of this "targeted selection" of regressors.[29]

Table 7 contains the corresponding outcomes for GDP growth, with bold figures representing rejections of the respective DM tests. It turns out that forecast accuracy gains occur over almost the entire range of forecast horizons; the only cases in which the augmented model results in slightly larger RMSFEs are $h = 1$ and $h = -5$, with the latter, as the furthest backcast, presenting a particular challenge, given that almost all of the indicators have been published over the reference period. Regarding the PLS variants in the model AM-g, the augmented model based on the selection above dominates the benchmark statistically for $h \geq 5$ (and $h = -1$), while never being truly inferior to it. Although the gains tend to be a bit smaller for some instances compared to the "pure" AM-gz or AM-g scenarios, they are more robust in the sense that no cases of severe forecast deterioration are present. Indeed, a relative RMSFE of 1.25 (as for PLS, $h = 1$ under AM-g) is hard to

27 Detailed results for all methods are available upon request, and are provided in part in the aforementioned working paper version.

28 While soft and Google indicators seem to explain the same variation in the dependent variable based on the outcomes described in Section 5.2, the results from Section 5.3 point towards them not containing the exact same information, with differences in informational content influencing forecasts on a more disaggregate level. We shed further light on this issue by conducting a further robustness analysis, where we let the data decide whether to include Google or soft indicators or both. On balance, though, the Google series showed a larger degree of commonality with the respective hard indicator than either soft variables or the two combined. For the details, we refer again to the working paper version (Götz & Knetsch, 2017).

29 The forecast performance with respect to Manufacturing is somewhat more robust under model AM-gz than under model AM-g (which delivers larger forecast accuracy gains for some horizons, but large losses for others).

Table 6
The targeted selection of regressors.

Hard indicator (x)	Selection method	Type of x-equation
Production Mining	Subj-Cat	AM-g
Industrial Production	PLS	AM-gz
Energy Production	LASSO	AM-g
Production in Construction	./.	BM
Real Retail Sales (incl. cars)	Subj-Cat	AM-g
Toll	LASSO	AM-g
Sales Hotel Industry	LASSO	AM-g
Value-added Tax (VAT)	./.	BM

Table 7
GDP growth: AM based on targeted selection vs. BM.

Forecast horizon											
Forecasts			Nowcasts						Backcasts		
17	15	13	11	9	7	5	3	1	−1	−3	−5
0.91	**0.90**	**0.93**	**0.95**	**0.85**	**0.8**	**0.85**	0.95	1.00	**0.94**	0.95	1.02

sell, even if the difference in the underlying forecast errors is not statistically significant. With forecast improvements of up to 20%, though, the results are remarkable enough.[30]

5.5. The Google series chosen

Finally, let us take a closer look at the Google series that actually get selected by the data-driven approaches. As an illustration, we consider PLS, as representing the factor-based methods, and LASSO, one of the shrinkage approaches. Starting with PLS, we consider Industrial Production and Sales Hotel Industry as examples, since PLS performs quite well for these two x indicators (see Table B.7). We focus here on the composition of the first PLS-factors, to save space; after all, it is the one that has the largest correlations with the eligible Google subcategories and the target variable in question.[31] Figs. B.2 and B.3 (in Appendix B) show the loadings of the first PLS-factors over time. Note that the loadings are scaled such that they add up to 100%, which enables us to better compare the relative importance of one search term over the other ones in that category.

Starting with the factor for Industrial Production, the largest weights — in absolute terms — are recorded for Vehicle Brands, Vehicle Codes & Driving Laws, Business Education, Vehicle Shopping, Boats & Watercraft and Chemicals Industry. Apparently, search terms related to the production or sale of vehicles show a lot of co-movement with Industrial Production. Somewhat surprisingly, the two search terms with the largest weights in the Business-&-Industrial category are Business Education and Chemicals Industry rather than Manufacturing, but internet users might simply be more inclined to look for such related terms. Another interesting observation is that the majority of subcategories load negatively on the first PLS factor. Turning to the factor for Sales Hotel Industry, the five subcategories with the largest average (absolute) weights are Travel Guides & Travelogues, Luggage & Travel Accessories (though with a negative loading), Car Rental & Taxi Services, Tourist Destinations and Carpooling & Vehicle Sharing. All of these search terms appear logical, especially in light of the many online services that center around such issues (e.g., *Tripadvisor*). Note that the Travel subcategories mostly load positively on the factor, whereas the Food-&-Drinks and Sensitive-Subjects subcategories receive negative loadings most of the time.[32]

Let us now turn to the Google variables chosen by LASSO and focus as an example on Industrial Production, Energy Production and VAT, all three of which actually achieve improvements in forecasting performances using this approach.[33] Fig. B.4 shows the results and should be read as follows: whenever a color appears as a vertical bar, the corresponding subcategory is selected and enters the monthly indicator equation "fully"; i.e., if two colors share the vertical space, both Google search terms are selected with weights of one. All in all, the outcomes appear quite intuitive and mirror some of the results above on PLS selection. Vehicle Brands and Boats & Watercraft seem most important for Industrial Production, the "spot-on" subcategory Energy & Utilities proves useful for Energy Production, and VAT is dominated by Banking.

30 Note that, by construction, the forecast performance with respect to the underlying GDP components and hard indicators is better (or as good as) under any of the models considered thus far, because it is exactly what we based the targeted assignment on.

31 It turns out that for Industrial Production, three factors are chosen over almost the entire evaluation period; the only exception is May 2014, when only two factors enter the model. For Sales Hotel Industry we obtain two factors most of the time; a third factor is added in 2013:M7–2013:M11, 2014:M3–2014:M4, 2014:M9–2014:M10 and 2015:M1. The results are available upon request.

32 Note however that, when discussing the sign of the factor loadings, one should keep in mind that we would normally also inspect the respective PLS factor coefficient in the equations for Industrial Production and Sales Hotel Industry. We stick to an illustrative discussion here.

33 As was stated earlier, we refer to the working paper version for details. We do not show the graphs that correspond to Production Mining and Sales Hotel Industry, further indicators for which larger forecast accuracy gains were detected, because they are not very illustrative: Production Mining is influenced predominantly by the Automotive Industry, and for Sales Hotel Industry the subcategory Travel Guides & Travelogues is the only one selected over the entire evaluation period.

6. Conclusion

This paper has analyzed whether internet data contain information that is useful for predictions of economic activity. In particular, we incorporated Google search data into a bridge equation model for the German macroeconomy, to assess whether they can improve the GDP growth forecasts. Treating the Google variables similarly to survey indicators, they affect the GDP growth forecasts through underlying hard indicators, which have an effect on the GDP components in an intermediate step. We addressed the crucial issue of which Google search terms to choose by considering variable selection approaches from three classes: ad hoc, factor and shrinkage methods. Using an out-of-sample, pseudo real time forecast exercise, the performances of two augmented model versions — one including Google data next to soft (survey) indicators and one including Google data instead of them — were compared to that of a status-quo model that did not include internet data at all.

It turned out that hardly any forecast improvements were found when Google and survey data appeared alongside each other in the equations for the hard indicators. Only when compressing the information embedded in the Google series using partial least squares were forecast accuracy gains detected. Inspecting the outcomes on a more disaggregate level revealed that these gains stemmed mainly from improved forecasts for the manufacturing sector. However, large forecast accuracy gains were detected when the survey variables were replaced with Google indicators, especially for forecasts and long-horizon nowcasts, which provides some evidence that Google data may potentially form a alternative to survey variables. The outcomes for the GDP components and underlying hard indicators depended somewhat on the variable selection method used. As a consequence, assigning selection methods individually to the hard indicators led to forecast improvements for almost all horizons, and to the most robust results overall.

In future work it might be worthwhile to consider an even tighter Google variable pre-selection, or the use of a data-driven approach like the (sparse) group lasso, to further improve the forecast performance. Considering specific, tailored Google search terms instead of categorized versions might be another promising alternative. However, it is also possible that Google search terms are not the right data source for all of the indicators in question. Many internet platforms or software applications have emerged that target specific markets or groups, and it could be that data derived from *Amazon* or *Autoscout24* would provide a better fit for predicting Real Retail Sales (incl. cars), and similarly data from *Tripadvisor* or *HRS* for the Hotel Industry and *Immobilienscout24* for Production in Construction. Finally, one should also keep in mind that the bridge equation model that we have considered in this paper is merely an example. An interesting future analysis would be to incorporate Google data into alternative model specifications, e.g. dynamic factor models. All in all, though, we feel confident in concluding that, although there are still many open issues and possible pitfalls with the use of internet search data, they surely show enough of a potential to improve macroeconomic forecasts.

Acknowledgments

We particularly thank Stephan Smeekes for an extensive review and many insightful remarks on an earlier version of the paper. Furthermore, we thank Etiënne Wijler, Tatjana Dahlhaus, Malte Knüppel, Daniela Rahn, Sandra Paterlini, Klemens Hauzenberger, Alain Hecq, participants at the 2016 CFE in Seville, the Workshop on Advances in Quantitative Economics II in Maastricht and members of the Big Data Project Group of the Deutsche Bundesbank for valuable comments and discussions. The views expressed in this paper are solely ours and should not be interpreted as reflecting the views of the Deutsche Bundesbank.

Appendix A. Details on Google variable selection methods

This section contains technical details on the data-driven Google variable selection approaches, as well as information on various input choices for the methods under consideration.

A.1. Principal component analysis

Intuitively speaking, PCA extracts factors in such a way as to maximize the variance accounted for within the group of predictors. Technically, each additional factor (i.e., each linear combination of the respective regressors) maximizes the remaining variance within the set of regressors (conditional on being orthogonal to the previous factors). Detailed explanations and formulae can be found in numerous textbooks on the related subject matters.

As far as the number of factors that we use in our analysis is concerned, Vosen and Schmidt (2011) used the Kaiser–Guttman criterion, which leads to a comparatively large number of factors. Given the length of their sample period, a re-adjustment of the number of factors was necessary to avoid overfitting of the model. No such issues emerged when using the scree test of Vosen and Schmidt (2012), so we opted for this criterion. The lag length of the resulting factors is determined via the Schwartz information criterion (SIC hereafter).

A.2. Partial least squares

Unlike PCA, PLS takes the relationship between the regressors and the target variable into account when extracting the factors. Technically speaking, each additional PLS factor is defined as the linear combination that best explains the target variable, conditional on being orthogonal to the previous factors. The weights, w, of the next PLS factor are equal to the covariances between the predictors and a new target variable, which is obtained by removing the linear effects of all previously computed PLS factors:

$$w_{i+1} = \Sigma_{xy} - \Sigma_{xx}\Omega_i(\Omega_i'\Sigma_{xx}\Omega_i)^{-1}\Omega_i'\Sigma_{xy}, \quad i = 1, \ldots, K-1, \tag{5}$$

with $w_1 = \Sigma_{xy}$, where Σ_{xy} is the sample equivalent of the covariance between the predictors and the target (Σ_{xx} follows straightforwardly), $\Omega_i = (w_1, \ldots, w_i)$, and K

is the number of predictors. For details on the formulae, we refer to Groen and Kapetanios (2016) or Cubadda and Guardabascio (2012). PLS can be interpreted as a middle ground between PCA and canonical correlations analysis (CCA hereafter), where the target variable is usually a vector rather than a single time series (Götz, Hecq, & Smeekes, 2016).[34]

Regarding the number of PLS factors and the respective lag lengths, we experimented with several fixed numbers, information criteria and a cross-validation approach, which takes the last two years of available observations as a validation sample and the remaining ones as a training sample. It turns out that the SIC leads to the most stable and reliable outcomes.

A.3. LASSO

The usual LASSO regression is of the form

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \frac{1}{T}\sum_{t=1}^{T}(y_t - X_t\beta)^2 + \lambda\sum_{j=1}^{K}|\beta_j|, \tag{6}$$

where β is a K-dimensional vector, with K being the number of predictors. $\lambda \geq 0$ represents the penalty parameter, where $\lambda = 0$ corresponds to the OLS estimator and $\lambda \to \infty$ leads to shrinking all parameters to zero.

We determine the value of λ using the SIC adapted to the LASSO, i.e., where the degrees of freedom are adjusted based on the framework of Stein's unbiased risk estimation (Zou, Hastie, & Tibshirani, 2007). Another commonly-used alternative is time series cross-validation (see e.g. Smeekes & Wijler, 2016).[35] In an extensive Monte Carlo study of several shrinkage (and factor) methods, Smeekes and Wijler (2016) show that the SIC seems to have an edge over time series cross-validation (and over the Akaike information criterion, for that matter). We ensure a large degree of shrinkage among the pre-specified set of candidate Google series by considering only λ-values that lead to at most six non-zero coefficients in the model. Note that lagged Google observations are contained in the set of regressors; i.e., we perform variable and lag selection at the same time.

A.4. Adaptive LASSO

In the adapted LASSO version, the parameters in the penalty term are weighted in order to penalize irrelevant variables to a higher degree than relevant ones:

$$\hat{\beta}(\lambda) = \arg\min_{\beta} \frac{1}{T}\sum_{t=1}^{T}(y_t - X_t\beta)^2 + \lambda\sum_{j=1}^{K}\frac{|\beta_j|}{w_j}, \tag{7}$$

where the weights w are determined using cleverly chosen initial estimators, for which the absolute values of OLS or ridge coefficients are common choices (Smeekes & Wijler, 2016). We follow the former approach in order to obtain an initial estimator, namely $w_j = |\hat{\beta}_j^{OLS}|$. In the second stage, Eq. (7) is estimated, where we use a reformulation (see e.g. Bühlmann & van de Geer, 2011) of the adaptive Lasso into its ordinary counterpart from Eq. (6): set $\tilde{X}^{(j)} = w_j X^{(j)}$ and $\tilde{\beta}_j = \frac{\beta_j}{w_j}$, where $X^{(j)}$ denotes the jth column of the $T \times K$ data matrix X; then, estimating β in Eq. (7) boils down to estimating $\tilde{\beta}$ using a conventional Lasso regression. Finally, an estimator for β is obtained by back-transformation, i.e., $\beta^* = w_j\tilde{\beta}^*$, where stars indicate the estimators to be the solutions of the respective optimization problems underlying Eqs. (6) and (7).

Again, we allow at most six non-zero coefficients and employ SIC for determining the tuning parameter λ.

A.5. Boosting

The boosting procedure can be summarized in the following steps:

1. Initialize $\hat{f}_{t,0} = \bar{y}$ for each t. Set $m = 0$.
2. Increase m by one. For each t compute $u_t = y_t - \hat{f}_{t,m-1}$.
3. For each potential regressor k, regress u_t on $g_{t,k}$ and compute the sum of squared residuals, $SSR_k = \sum_{t=1}^{T}(u_t - g_{t,k}\hat{\theta}_k)^2$. $\hat{\theta}_k$ is just the corresponding regression coefficient. Hence, we implicitly opt for an L_2-loss function and OLS as a base learner.
4. Choose g_{t,k_m^*} s.t. $SSR_{k^*} = \arg\min_k SSR_k$ and set $\hat{f}_{t,m} = g_{t,k_m^*}\hat{\theta}_{k_m^*}$.
5. Update $\hat{f}_{t,m} = \hat{f}_{t,m-1} + \nu\hat{f}_{t,m}$ for each t, where $0 < \nu < 1$.
6. Repeat steps 2 to 5 until $m = M$.

In general, two inputs are key to the functioning of boosting: the step size or shrinkage parameter (usually labeled ν) and the stopping criterion (usually denoted by M). For the former, we take the commonly-used value of 0.1 (see e.g. Bühlmann & Hothorn, 2007), and for the latter we choose 250. However, we make boosting more directly comparable to the other approaches by adapting its methodology slightly so that it functions solely as a variable selection approach: after saving the chosen Google regressors g_{t,k_m^*} at each step m, we select the series that were chosen at least δM times, where $\delta \in [0, 1]$ determines the severity with which we select the candidates. We consider a grid of δ values, i.e., $\delta = 0.05, 0.1, 0.15, 0.2, 0.25, 0.5$, and find $\delta = 0.2$ to give the most satisfactory outcomes. We also include an additional rule, which often improves the outcomes when it applies: whenever no Google regressor surpasses the δM-barrier, we select the one that is chosen most often.

Appendix B. Tables & figures

See Tables B.1–B.7 and Figs. B.1–B.4.

[34] In CCA, linear combinations on both sides of the equation are determined in such a way as to maximize the covariance between them (again conditional on them being orthogonal to the previous factors). In such systems (often vector autoregressive models), PCA, PLS and CCA are often used to unravel an underlying reduced rank structure in the model (see e.g. Cubadda, Hecq, & Palm, 2009).

[35] Note that the usual k-fold cross-validation (see e.g. Bühlmann & van de Geer, 2011) is not valid in a time series setting.

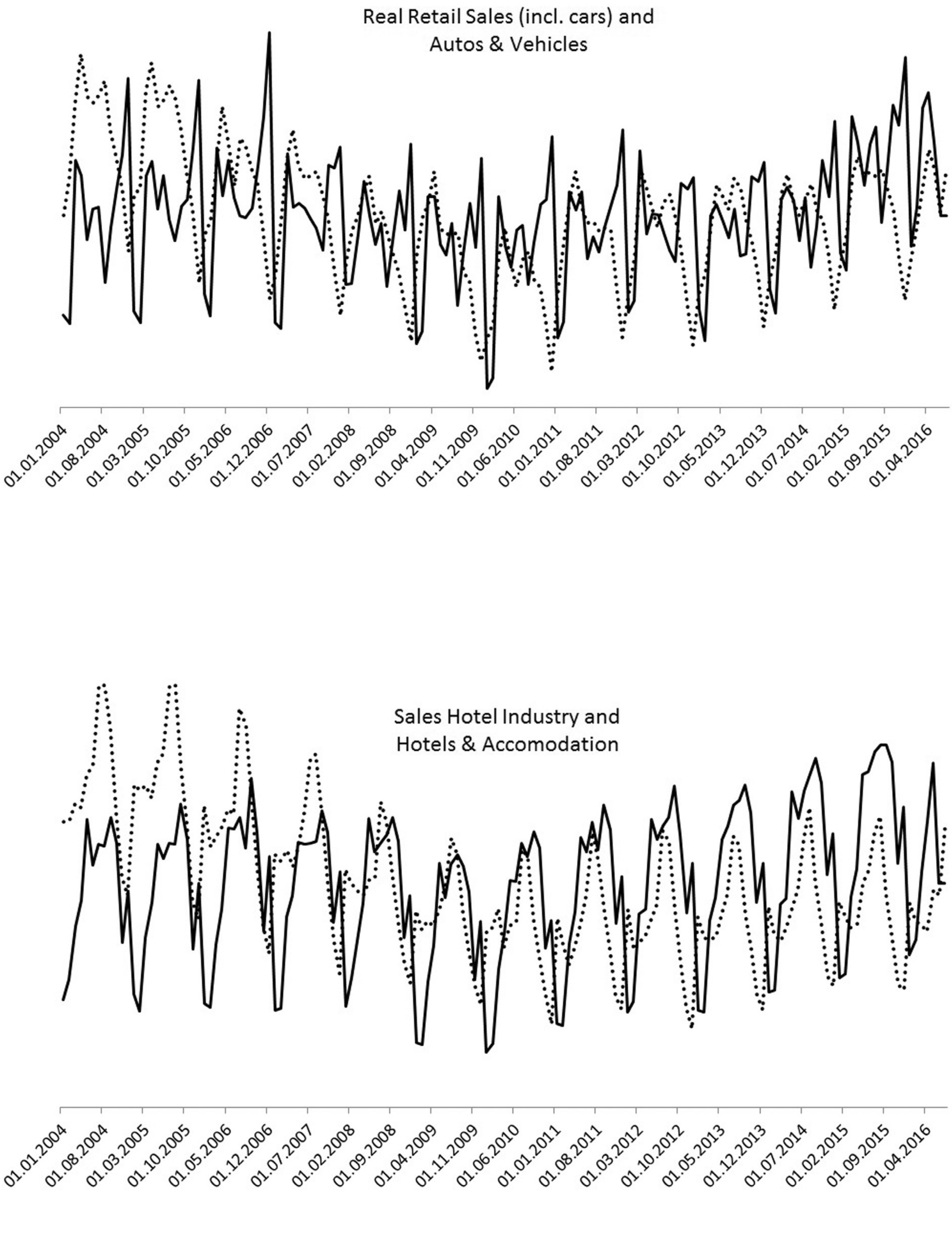

Fig. B.1. The potential of Google series for prediction. Note: The macroeconomic indicators appear as solid lines, while the Google time series are depicted as dotted lines. The series shown here are seasonally unadjusted, standardized and represented at a monthly frequency; i.e., the weekly Google observations are aggregated temporally, as explained in Section 3.

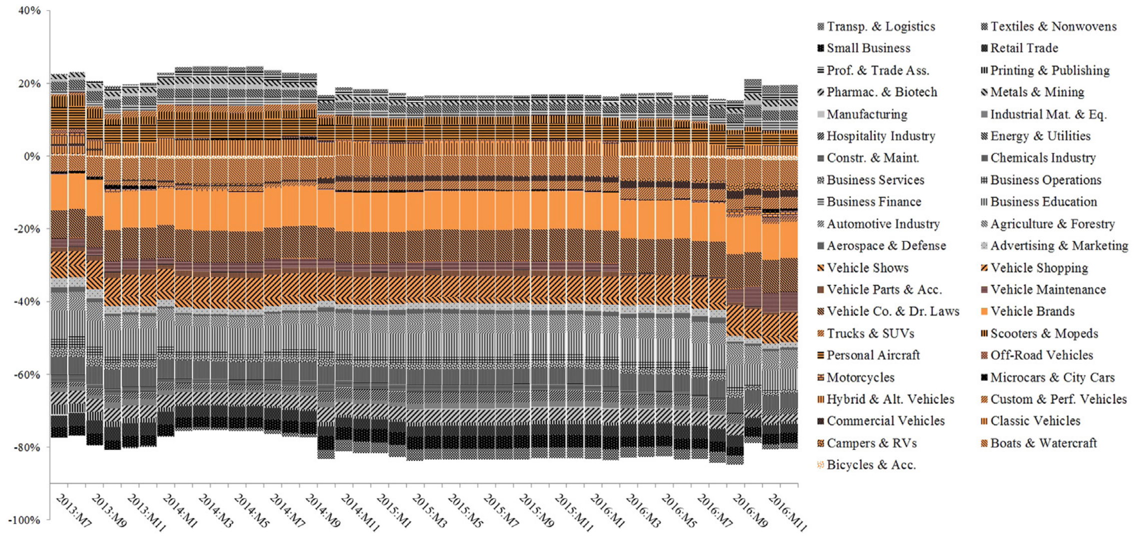

Fig. B.2. First PLS-factor loadings, Industrial Production: AM-g. Note: The graph is based on the Google-variable-augmented AM-g (see Section 2.2) where, apart from autoregressive lags, only Google indicators appear in the equation for x (see Eq. (2b)). The loadings are scaled such that they add up to 100%. The orange-shaded loadings correspond to the subcategories of Autos & Vehicles, the grey-shaded ones to those of Business & Industrial. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

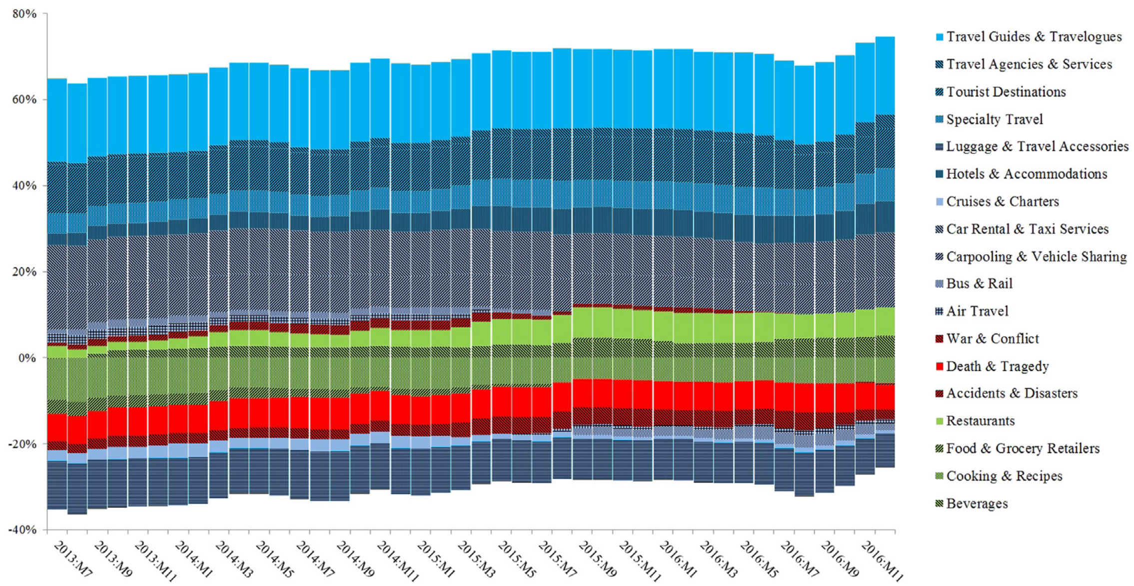

Fig. B.3. First PLS-factor loadings, Sales Hotel Industry: AM-g. Note: For the underlying model, see Fig. B.2. The green-shaded loadings correspond to the subcategories of Food & Drink, the red-shaded ones to those of Sensitive Subjects, and the blue-shaded ones to those of Travel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

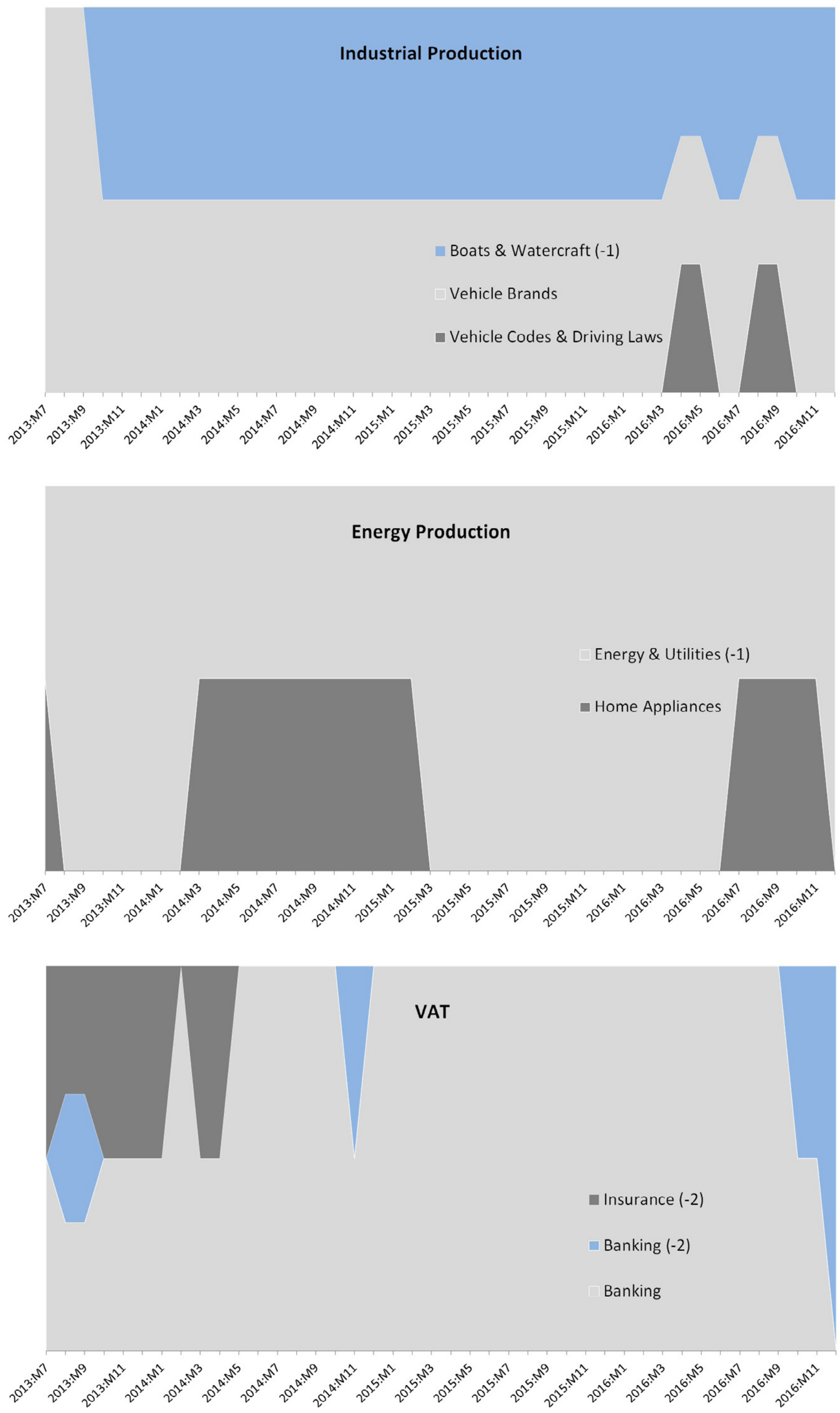

Fig. B.4. Google variable selection by LASSO: AM-*g*.

Table B.1
ECB Google data: subcategories.

Autos & Vehicles		
Bicycles & Accessories	Boats & Watercraft	Campers & RVs
Classic Vehicles	Commercial Vehicles	Custom & Performance Vehicles
Hybrid & Alternative Vehicles	Microcars & City Cars	Motorcycles
Off-Road Vehicles	Personal Aircraft	Scooters & Mopeds
Trucks and SUVs	Vehicle Brands	Vehicle Codes & Driving Laws
Vehicle Maintenance	Vehicle Parts & Accessories	Vehicle Shopping
Vehicle Shows		
Beauty & Fitness		
Beauty Pageants	Body Art	Cosmetic Procedures
Cosmetology & Beauty Professionals	Face & Body Care	Fashion & Style
Fitness	Hair Care	Spas & Beauty Services
Weight Loss		
Business & Industrial		
Advertising & Marketing	Aerospace & Defence	Agriculture & Forestry
Automotive Industry	Business Education	Business Finance
Business Operations	Business Services	Chemicals Industry
Construction & Maintenance	Energy & Utilities	Hospitality Industry
Industrial Materials & Equipment	Manufacturing	Metals & Mining
Pharmaceuticals & Biotech	Printing & Publishing	Professional & Trade Associations
Retail Trade	Small Business	Textiles & Nonwovens
Transportation & Logistics		
Computers & Electronics		
CAD & RAM	Computer Hardware	Computer Security
Consumer Electronics	Electronics & Electrical	Enterprise Technology
Networking	Programming	Software
Finance		
Accounting & Auditing	Banking	Credit & Lending
Financial Planning & Management	Grants, Scholarships & Financial Aid	Insurance
Investing		
Food & Drink		
Beverages	Cooking & Recipes	Food & Grocery Retailers
Restaurants		
Health		
Ageing & Geriatrics	Alternative & Natural Medicine	Health Conditions
Health Education & Medical Training	Health Foundations & Medical Research	Medical Devices & Equipment
Medical Facilities & Services	Medical Literature	Men's Health
Mental Health	Nursing	Nutrition
Oral & Dental Care	Pediatrics	Pharmacy
Public Health	Reproductive Health	Substance Abuse
Vision Care	Women's Health	
Home & Garden		
Bed & Bath	Domestic Services	Gardening & Landscaping
Home Appliances	Home Furnishing	Home Improvement
Homemaking & Interior Decor	Home Storage & Shelving	HVAC & Climate Control
Kitchen & Dining	Laundry	Nursery & Playroom
Pest Control	Swimming Pools & Spas	Yard & Patio
Internet & Telecom		
Communications Equipment	Email & Messaging	Mobile & Wireless
Search Engines	Service Providers	Teleconferencing
Web Apps & Online Tools	Web Portals	Web Services
Jobs & Education		
Education	Jobs	
Law & Government		
Government	Legal	Military
Public Safety	Social Services	

(*continued on next page*)

Table B.1 (*continued*)

News		
Broadcast & Network News	Business News	Gossip & Tabloid News
Health News	Journalism & News Industry	Local News
Newspapers	Politics	Sports News
Technology News	Weather	World News
Real Estate		
Apartments & Residential Rentals	Commercial & Investment Real Estate	Property Development
Property Inspections & Appraisals	Property Management	Real Estate Agencies
Real Estate Listings	Timeshares & Vacation Properties	
Sensitive Subjects		
Accidents & Disasters	Death & Tragedy	War & Conflict
Shopping		
Antiques & Collectibles	Apparel	Auctions
Classifieds	Consumer Resources	Entertainment Media
Gifts & Special Event Items	Luxury Goods	Mass Merchants & Department Stores
Photo & Video Services	Shopping Portals & Search Engines	Swap Meets & Outdoor Markets
Tobacco Products	Toys	Wholesalers & Liquidators
Sports		
College Sports	Combat Sports	Extreme Sports
Fantasy Sports	Individual Sports	Motor Sports
Sporting Goods	Sports Coaching & Training	Team Sports
Water Sports	Winter Sports	World Sports Competitions
Travel		
Air Travel	Bus & Rail	Carpooling & Vehicle Sharing
Car Rental & Taxi Services	Cruises & Charters	Hotels & Accomodations
Luggage & Travel Accessories	Specialty Travel	Tourist Destinations
Travel Agencies & Services	Travel Guides & Travelogues	

Table B.2
Google variable selection: subjectively.

Monthly indicator/GDP component	Category-level	Subcategory-level
Prod. Mining	Business & Industrial	Agriculture & Forestry, Metals & Mining
Ind. Prod.	Autos & Vehicles, Business & Industrial	Classic Vehicles, Automotive Industry, Chemicals Industry, Industrial Materials & Equipment, Manufacturing
Energy Prod.	Business & Industrial, Home & Garden	Energy & Utilities, HVAC & Climate Control
Production Constr.	Business & Industrial, Home & Garden, Real Estate	Construction & Maintenance, Gardening & Landscaping, Property Development
Retail Sales	Autos & Vehicles, Sensitive Subjects, Shopping	Classic Vehicles, Vehicle Shopping, War & Conflict, Shopping Portals & Search Engines
Toll	Autos & Vehicles, Business & Industrial	Commercial Vehicles, Trucks & SUVs, Automotive Industry, Transportation & Logistics
Hotel Ind.	Food & Drink, Sensitive Subjects, Travel	Restaurants, War & Conflict, Hotels & Accommodations
VAT	Finance, Law & Government, News	Financial Planning & Management, Legal, Business News
Agric. & Fores.	Business & Industrial	Agriculture & Forestry
Info. & Comm.	Computer & Electronics, Internet & Telecom	Consumer Electronics, Communications Equipment, Email & Messaging
Housing	Home & Garden, Real Estate	Home Furnishing, Real Estate Agencies
Financial Services	Finance	Credit & Lending, Financial Planning & Management
Corporate Services	Business & Industrial, Finance, News	Business Services, Banking, Business News
Public Services, Health & Educ.	Finance, Health, Jobs & Education	Grants, Scholarships & Financial Aid, Medical Facilities & Services, Social Services, Education
Other Services	Business & Industrial, Internet & Telecom, Travel	Spas & Beauty Services, Domestic Services, Service Providers, Car Rental & Taxi Services, Travel Agencies & Services

Note: The first eight rows correspond to monthly indicators (x in Table 1), while the remaining seven capture those GDP components that do not get augmented with an x-indicator such that the Google series enter directly in a time-aggregated fashion. All Google indicators enter with one (monthly) lag in the case of a monthly indicator equation and with no (quarterly) lags in the case of a quarterly GDP component equation.

Table B.3
Google variable selection: Google Correlate.

Macro indicator/GDP component	Subcategories
Prod. Mining	Business Operations (+1 lag), Industrial Materials & Equipment
Ind. Prod.	Metals & Mining (+1 lag), Apartments & Residential Rentals (+1 lag)
Energy Prod.	Spas & Beauty Services (+2 lags), HVAC & Climate Control, Winter Sports (+2 lags)
Production Constr.	Construction & Maintenance, Manufacturing, Home Improvement
Retail Sales	Toys
Toll	Construction & Maintenance, Transportation & Logistics, Public Safety
Hotel Ind.	Gifts & Special Events Items
VAT	Restaurants, Car Rental & Taxi Services (+2 lags)
Agric. & Fores.	Chemicals Industry
Info. & Comm.	Software, Web Apps & Online Tools
Housing	Real Estate Listings
Financial Services	Business News
Corporate Services	Financial Planning & Management, Service Providers
Public Services, Health & Educ.	Web Services
Other Services	Education, Travel Agencies & Services

Note: Lags are indicated in brackets whenever present. For the rest, see Table B.2.

Table B.4
GDP growth: AM-*gz* vs. BM.

Method		Subj-Cat	Subj-Subcat	Google-Corr	PCA-Cat	PCA	PLS-Cat	PLS	LASSO	Boosting
Forecast horizon	17	1.00	1.01	1.01	0.99	0.99	0.99	0.92	0.97	0.96
	15	1.00	1.01	1.01	1.00	0.99	0.99	0.92	0.98	0.97
	13	0.98	0.99	1.01	1.01	1.00	1.00	0.95	0.98	0.98
	11	0.98	0.99	1.00	1.02	1.02	1.01	0.98	1.01	1.00
	9	0.99	1.00	0.99	1.03	1.03	0.94	0.92	1.03	1.01
	7	1.00	0.96	1.00	1.00	1.03	0.86	0.86	1.02	1.00
	5	1.00	1.02	1.00	1.03	1.02	0.94	0.91	1.06	1.04
	3	0.92	1.15	1.04	1.07	0.95	1.03	1.05	1.07	1.07
	1	0.92	1.10	1.03	1.04	0.94	1.04	1.09	1.07	1.05
	−1	1.03	1.09	0.99	1.01	1.00	1.03	0.96	1.05	1.03
	−3	1.02	1.10	1.00	0.98	0.98	1.02	0.95	1.06	1.04
	−5	1.01	0.97	1.00	1.00	1.00	1.01	1.00	0.99	0.96

Note: The target variable is quarterly GDP growth, the weighted average of the GDP components y listed in the first column of Table 1. The figures represent RMSFEs of the Google-variable-augmented AM-*gz* (see Section 2.2), where both Google and soft indicators appear in the equation for x (see Eq. (2a)), relative to the benchmark model without any Google series (see Section 2.1, Eqs. (1)–(3)). The various Google variable selection methods are described in Section 4. The forecast horizons −5, −3 and −1 correspond to backcasts, 1 to 11 correspond to nowcasts, and 13, 15 and 17 correspond to forecasts.

Table B.5
GDP components and monthly indicators: AM-*gz* (PLS) vs. BM.

		PLS							
		Mining	Manufacturing	Energy & Water	Construction	Trade	Traffic	Hotel Industry	Net Taxes
Forecast horizon	17	0.94	0.94	1.00	1.09	0.98	1.05	0.99	1.00
	13	0.97	0.96	0.98	1.08	0.99	1.01	0.96	1.01
	9	1.02	0.88	1.01	0.98	1.17	1.02	0.99	1.01
	5	1.05	0.87	1.08	1.11	1.16	1.00	1.00	1.02
	1	1.10	0.97	1.02	1.12	1.10	0.99	1.09	1.06
	−3	1.06	1.00	0.99	1.07	1.11	1.00	1.09	1.03
		PLS							
		Production Mining	Industrial Production	Energy Production	Production Construc-tion	Retail Sales	Toll	Sales Hotel Industry	VAT
Forecast horizon	17	1.03	0.87	0.94	1.21	0.99	0.95	1.00	0.96
	15	0.96	0.88	0.96	1.26	0.98	0.96	0.98	0.95
	13	0.97	0.86	1.00	1.14	0.99	0.94	0.96	0.92
	11	0.98	0.90	1.02	1.02	0.98	0.92	0.97	0.95
	9	1.00	0.91	1.04	1.00	0.99	0.98	1.00	0.97
	7	0.98	0.88	1.01	1.07	0.98	0.98	0.99	0.98

(*continued on next page*)

Table B.5 (*continued*)

	PLS							
	Production Mining	Industrial Production	Energy Production	Production Construc-tion	Retail Sales	Toll	Sales Hotel Industry	VAT
5	0.99	0.86	1.07	1.05	1.00	1.00	0.92	1.02
3	0.98	0.93	1.07	1.08	0.93	1.01	0.89	1.04
1	1.09	0.99	1.13	1.02	1.01	1.25	0.89	1.21
−1	1.13	1.01	1.23	1.04	1.03	1.25	0.89	1.18
−3	1.13	1.01	1.23	1.04	1.03		0.98	1.18
−5					1.10		0.98	1.11
−7					1.10			1.11

Note: The target variables in the top panel are the eight quarterly GDP components y that have hard, monthly indicators x assigned to them, while those in the bottom panel are the corresponding hard indicators x themselves (see the first and second columns in Table 1). For the monthly indicators, the forecast horizons −7 to −1 correspond to backcasts, 1 and 3 correspond to nowcasts and 5 to 17 correspond to forecasts. For the underlying models, see Table B.4.

Table B.6
GDP growth: AM-*g* vs. BM.

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

Note: The figures represent RMSFEs of the Google-variable-augmented AM-*g* (see Section 2.2) where, apart from autoregressive lags, only Google indicators appear in the equation for x (see Eq. (2b)), relative to the benchmark model without any Google series (see Section 2.1, Eqs. (1)–(3)). Bold figures indicate rejections of the DM test for equal predictive accuracy. For the rest see Table B.4.

Table B.7
GDP components and monthly indicators: AM-*g* (PLS) vs. BM.

		PLS							
		Mining	Manufacturing	Energy & Water	Construction	Trade	Traffic	Hotel Industry	Net Taxes
Forecast horizon	17	**0.87**	**0.90**	1.01	**1.33**	0.89	1.08	0.94	**1.00**
	13	**0.92**	1.14	**0.99**	**1.29**	**0.93**	**1.02**	**0.90**	**1.02**
	9	**0.98**	**0.84**	**1.02**	**1.16**	**1.15**	1.01	1.03	1.02
	5	1.04	**0.68**	**1.07**	**1.26**	**1.15**	**1.01**	1.01	**1.05**
	1	1.05	1.22	**1.02**	1.07	1.09	0.99	1.06	**1.08**
	−3	1.05	1.12	1.00	1.02	**1.10**	1.00	**1.08**	1.06
		PLS							
		Production Mining	Industrial Production	Energy Production	Production Construc-tion	Retail Sales	Toll	Sales Hotel Industry	VAT
Forecast horizon	17	**0.95**	**0.81**	0.91	**1.40**	**0.97**	**0.91**	1.02	0.91
	15	**0.90**	**0.87**	0.93	**1.44**	**0.95**	**0.91**	1.01	0.92
	13	**0.91**	**0.80**	0.96	**1.32**	0.98	**0.90**	0.99	0.87
	11	**0.93**	**0.88**	0.98	1.10	0.97	**0.87**	1.01	**0.92**
	9	**0.95**	**0.80**	1.00	1.09	0.99	0.94	0.98	0.92
	7	**0.94**	**0.78**	0.98	1.09	0.99	0.94	0.97	**0.96**
	5	**0.95**	**0.72**	1.02	1.11	1.01	0.97	**0.83**	**0.99**
	3	**0.94**	**0.93**	1.03	1.10	0.95	**0.98**	**0.80**	**1.04**
	1	1.06	**0.94**	**1.11**	1.04	1.04	**1.15**	**0.85**	**1.22**
	−1	1.07	**0.99**	**1.18**	1.05	1.07	1.16	**0.86**	**1.19**
	−3	1.06	**0.99**	**1.18**	1.05	1.06		**1.01**	**1.19**
	−5					1.12		1.01	**1.12**
	−7					1.12			**1.12**

Note: For the underlying models and the meaning of bold values, see Table B.6. For the setup of the rest, see Table B.5.

References

Angelini, E., CambaMendez, G., Giannone, D., Reichlin, L., & Rünstler, G. (2011). Short-term forecasts of euro area GDP growth. *The Econometrics Journal*, *14*(1), C25–C44.

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

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

Baffigi, A., Golinelli, R., & Parigi, G. (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting*, *20*(3), 447–460.

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

Bontempi, M. E., Golinelli, R., & Squadrani, M. (2016). A new index of uncertainty based on internet searches: A friend or foe of other indicators? Working paper 1062. Universita' di Bologna, Dipartimento Scienze Economiche.

Buchen, T., & Wohlrabe, K. (2011). Forecasting with many predictors: Is boosting a viable alternative? *Economics Letters*, *113*(1), 16–18.

Buchen, T., & Wohlrabe, K. (2014). Assessing the macroeconomic forecasting performance of boosting: Evidence for the United States, the euro area and Germany. *Journal of Forecasting*, *33*(4), 231–242.

Bühlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, *22*(4), 477–505.

Bühlmann, P., & van de Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications* (1st Ed.). Springer Publishing Company.

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

Camacho, M., Perez-Quiros, G., & Poncela, P. (2013). Short-term forecasting for empirical economists: A survey of the recently proposed algorithms. *Foundations and Trends in Econometrics*, *6*(2), 101–161.

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

Clark, T. E. (1999). Finite-sample properties of tests for equal forecast accuracy. *Journal of Forecasting*, *18*(7), 489–504.

Clark, T. E., & McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics*, *105*(1), 85–110.

Clements, M., & Hendry, D. (1998). *Forecasting economic time series*. Cambridge University Press.

Cubadda, G., & Guardabascio, B. (2012). A medium-N approach to macroeconomic forecasting. *Economic Modelling*, *29*(4), 1099–1105.

Cubadda, G., Hecq, A., & Palm, F. C. (2009). Studying co-movements in large multivariate data prior to multivariate modelling. *Journal of Econometrics*, *148*(1), 25–35.

D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, *33*(4), 801–816.

Destatis, (2015). IT-Nutzung. https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2015/12/PD15_466_63931pdf.pdf?__blob=publicationFile. (Retrieved 05 August 2016).

Diebold, F. (2012). A personal perspective on the origin(s) and development of 'big data': the phenomenon, the term, and the discipline. Working paper 13-003. PIER.

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

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

Einav, L., & Levin, J. (2013). The data revolution and economic analysis. Discussion papers 12-017. Stanford Institute for Economic Policy Research.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*, 830–840.

Foroni, C., & Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting*, *30*(3), 554–568.

Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and Computation*, *121*(2), 256–285.

Gasso, G., Rakotomamonjy, A., & Canu, S. (2009). Recovering sparse signals with a certain family of nonconvex penalties and DC programming. *Transactions on Signal Processing*, *57*(12), 4686–4698.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. CIRANO Working papers 2004s-20. CIRANO.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinksi, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014.

Girardi, A., Golinelli, R., & Pappalardo, C. (2017). The role of indicator selection in nowcasting Euro-area GDP in pseudo real time. *Empirical Economics*, *53*(1), 79–99.

Goel, S., Hofmann, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences of the United States of America*, *1077*(41), 17486–17490.

Götz, T. B., Hecq, A., & Smeekes, S. (2016). Testing for Granger causality in large mixed-frequency VARs. *Journal of Econometrics*, *193*(2), 418–432.

Götz, T. B., & Knetsch, T. (2017). Google data in bridge equation models for German GDP. Bundesbank discussion paper 18/2017. Deutsche Bundesbank.

Groen, J. J., & Kapetanios, G. (2016). Revisiting useful approaches to data-rich macroeconomic forecasting. *Computational Statistics & Data Analysis*, *100*, 221–239.

Guzman, G. (2011). Internet search behavior as an economic forecasting tool: The case of inflation expectations. *The Journal of Economic and Social Measurement*, *36*(3), 119–167.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.

Hastie, T. R., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC Press.

Humphrey, B. D. (2010). *Forecasting existing home sales using Google seach engine queries* (Honors thesis), Duke University.

Johnson, H. A., Wagner, M. M., Hogan, W. R., Chapman, W., Olszewski, R. T., Dowling, J., et al. (2004). Analysis of Web access logs for surveillance and influenza. *Studies in Health Technology and Informatics*, *107*, 1202–1206.

Klein, L. R., & Sojo, E. (1989). Combinations of high and low frequency data in macroeconometric models. In L. R. Klein, & J. Marquez (Eds.), *Economics in theory and practice: an eclectic approach, vol. 17* (pp. 3–16). Springer Netherlands.

Koivupalo, H. (2014). Google data. mimeo. European Central Bank.

Konzen, E., & Ziegelmann, F. A. (2016). Lasso-type penalties for covariate selection and forecasting in time series. *Journal of Forecasting*, *35*, 592–612.

Koop, G., & Onorante, L. (2013). Macroeconomic nowcasting using Google probabilities. Working paper. European Central Bank.

Kulkarni, R., Haynes, K. E., Stough, R. R., & Paelinck, J. H. P. (2009). Forecasting housing prices with Google econometrics. Working paper no. 2009-10. George Mason University.

Lehmann, R., & Wohlrabe, K. (2016). Looking into the black box of boosting: the case of Germany. *Applied Economics Letters*, *23*(17), 1229–1233.

McCracken, M. W. (2007). Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics*, *140*(2), 719–752.

McLaren, N., & Shanbhogue, R. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin*, *51*(2), 134–140.

Newey, W. K., & West, K. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, *61*(4), 631–653.

Nymand-Andersen, P. (2016). *Big data: the hunt for timely insights and decision certainty*. Bank for International Settlements.

Pan, B., Wu, D. C., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology*, *3*(3), 196–210.

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, *5*(2), 197–227.

Schumacher, C. (2014). MIDAS and bridge equations. Discussion papers 26/2014. Deutsche Bundesbank, Research Centre.

Seabold, S., & Coppola, A. (2015). Nowcasting prices using Google trends: an application to Central America. Policy research working paper series 7398. The World Bank.

Shimshoni, Y., Efron, N., & Matias, Y. (2009). *On the predictability of search trends. Technical report*. Google, Israel Labs.
Silvestrini, A., & Veredas, D. (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, Graduate School of Business and Economics (GSBE).
Statista, (2016). Worldwide desktop market share of leading search engines. http://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/. (Retrieved 05 August 2008).
Stock, J., & Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*, 1167–1179.
Tibshirani, R. (1996). Regression analysis and selection via the Lasso. *Journal of the Royal Statistical Society. Series B.*, *58*, 267–288.
Toth, I. J., & Hajdu, M. (2012). Google as a tool for nowcasting household consumption: estimations on Hungarian data. Working paper. Institute for Economic and Enterprise Research.
Tuhkuri, J. (2016). Forecasting unemployment with Google searches. ETLA working papers 35. The Research Institute of the Finnish Economy.
Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28.
Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, *30*(6), 565–578.
Vosen, S., & Schmidt, T. (2012). A monthly consumption indicator for Germany based on Internet search query data. *Applied Economics Letters*, *19*(7), 683–687.
Wiermanski, C., & Wilshusen, S. M. (2015). Exploring the use of anonymized consumer credit information to estimate economic conditions: An application of Big Data. Discussion paper. Payment Cards Center.
Wohlrabe, K. (2008). *Forecasting with mixed-frequency time series models* (Ph.D. thesis), Ludwig-Maximilians-Universität München.
Wold, H. (1985). Partial least squares. In *Encyclopedia of statistical sciences*. John Wiley & Sons, Inc..
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429.
Zou, H., Hastie, T., & Tibshirani, R. (2007). On the "degrees of freedom" of the lasso. *The Annals of Statistics*, *35*(5), 2173–2192.

Thomas Götz has a Ph.D. from Maastricht University and is currently working at the Deutsche Bundesbank in the Macroeconomic Analysis and Projection Division. His main research interests lie in the field of theoretical and applied time series econometrics (non-stationary time series, unit root testing, co-integration, common features in multiple time series, mixed-frequency data, forecasting, big data, shrinkage methods) for macroeconomic data, particularly economic activity and inflation. He published in various journals including the *Journal of Econometrics*, the *International Journal of Forecasting and Economics Letters*.

Thomas Knetsch received his Ph.D. in Economics from the Freie Universität Berlin. He joined the Deutsche Bundesbank in 2001, worked in the Macroeconomic Analysis and Projection Division until 2016 and is currently Deputy Head of the General Economic Statistics Division. His main research interests lie in the field of empirical macroeconomics, time series econometrics, property market and price statistics as well as big data. He has published in journals including the *Oxford Bulletin of Economics and Statistics*, the *Journal of Forecasting and Empirical Economics*.