International Journal of Forecasting 36 (2020) 1163–1172

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Can Google search data help predict macroeconomic series?

Robin F. Niesert, Jochem A. Oorschot *, Christian P. Veldhuisen, Kester Brons, Rutger-Jan Lange

Department of Econometrics, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, Netherlands

ARTICLE INFO

Keywords:
Bayesian methods
Forecasting practice
Kalman filter
Macroeconomic forecasting
State space models
Nowcasting
Spike-and-slab
Hamiltonian sampler

ABSTRACT

We make use of Google search data in an attempt to predict unemployment, CPI and consumer confidence for the US, UK, Canada, Germany and Japan. Google search queries have previously proven valuable in predicting macroeconomic variables in an in-sample context. However, to the best of our knowledge, the more challenging question of whether such data have out-of-sample predictive value has not yet been answered satisfactorily. We focus on out-of-sample nowcasting, and extend the Bayesian structural time series model using the Hamiltonian sampler for variable selection. We find that the search data retain their value in an out-of-sample predictive context for unemployment, but not for CPI or consumer confidence. It is possible that online search behaviours are a relatively reliable gauge of an individual's personal situation (employment status), but less reliable when it comes to variables that are unknown to the individual (CPI) or too general to be linked to specific search terms (consumer confidence).

1. Introduction

Timely and accurate economic data are invaluable for making sensible investment and policy decisions. Unfortunately, many macroeconomic time series are released with a substantial time lag and are also subject to revisions. Previous research suggests that nowcasts (predictions of contemporaneous but unknown values) that make use of Google search data can outperform both AR(1) models and survey-based predictors. Improvements in terms of the mean absolute prediction error (MAPE) have been found for US inflation (Guzman, 2011), the UK housing market (McLaren & Shanbhogue, 2011), Swedish private consumption (Lindberg, 2011), German and Israeli unemployment (Askitas & Zimmermann, 2009; Suchoy, 2009) and US private consumption (Vosen & Schmidt, 2011). Outperformance seems to be particularly pertinent at structural breaks and extreme observations. Choi and Varian's (2012) Google search data model for US unemployment claims yielded an 11% improvement in MAPE relative to an AR(1) model, but 21% during recessions. D'Amuri and Marcucci (2017) find that Google category data has predictive power for US unemployment irrespective of whether the out-of-sample period starts before, during or after the Great Recession. Similarly, Preis, Moat, and Stanley (2013) found that a trading strategy based on the relative popularity of the search query 'debt' outperformed a buy-and-hold strategy over the period 2004–2011, but in particular during the financial crisis.

We are interested in three macroeconomic variables (unemployment, the consumer price index (CPI) and consumer confidence) for five countries (US, UK, Canada, Germany and Japan). We follow Scott and Varian (2014a, 2014b) in using online search data obtained from 'Google Trends' and 'Google Correlate' as exogenous variables. Google Trends is a service that produces a single time series indicating the level of search activity in a specific country for any category predefined by Google, such as 'unemployment appeals'. Google Correlate, on the other hand, produces up to 100 time series that are highly correlated with any (user-defined) series of interest (for details, see Stephens-Davidowitz and Varian (2014)). Scott

* Corresponding author.
E-mail address: oorschot@ese.eur.nl (J.A. Oorschot).

https://doi.org/10.1016/j.ijforecast.2018.12.006
0169-2070/

and Varian (2014a, 2014b) developed the Bayesian structural time series (BSTS) model for the purpose of handling the many regressors obtained from both data sets. Estimating their model using the entire sample, they produced monthly 'nowcasts' of the macroeconomic variables and found that the resulting 'in-sample predictions' outperformed both an AR(1) benchmark and a structural time series (STS) model in terms of MAPE.[1]

Naturally, caution is always required in extrapolating the findings of such in-sample analyses to out-of-sample contexts. Several studies have focused on the out-of-sample performances of Google search data, although they have typically been limited to hand-selected series from Google Trends, ignoring Google Correlate. For example, Choi and Varian (2012) show that the categories 'trucks & SUVs' and 'automotive insurance' help to predict motor vehicle sales, while D'Amuri and Marcucci (2017) show that the 'jobs' category helps to forecast US unemployment. Similarly, Naccarato, Falorsi, Loriga, and Pierini (2018) use the frequency of the search term 'job offers' to forecast Italian youth unemployment, and Yu, Zhao, Tang, and Yang (2019) use the search terms 'oil consumption', 'oil inventory' and 'oil price' to predict (changes in) oil consumption. It could be argued that all of these out-of-sample studies use somewhat simpler (autoregressive) models than Scott and Varian's (2014a, 2014b) BSTS model.

The question remains as to whether Scott and Varian's (2014a, 2014b) BSTS model using both Google Trends and Correlate data can be employed for making effective out-of-sample forecasts. This is no easy task: Scott and Varian (2014b, p. 21) themselves note that one disadvantage of using Google Correlate is that the strongest (in-sample) predictors are often 'spurious regressors' that lack a 'plausible economic justification' (which may explain why the out-of-sample studies cited above chose to exclude Google Correlate). To the best of our knowledge, the current paper is the first to systematically use Google Correlate to produce out-of-sample nowcasts. The selection of variables is particularly challenging given the large numbers of potentially relevant time series obtained from Google Trends and Correlate. For this purpose, Scott and Varian (2014a, 2014b) integrate into the BSTS model a spike-and-slab regression with the stochastic search variable selection (SSVS) sampler (George & McCulloch, 1997). However, the SSVS sampler may suffer when either the number of predictors or the multicollinearity among them is high; see e.g. Heaton and Scott (2010). We deviate from the work of Scott and Varian (2014a, 2014b) by using not only the SSVS but also the Hamiltonian sampler, which was introduced by Pakman and Paninski (2013) and may be beneficial when using Google search data.

We compare nowcasts of the BSTS model at a monthly frequency against those of the STS benchmark, which does not make use of Google search data, and find that the BSTS model usually outperforms the benchmark in in-sample settings. In an out-of-sample context, though, the BSTS model based on Google Trends data fails to ourperform the benchmark for consumer confidence and CPI. Moreover, adding Google Correlate data does not improve the performance, a finding that we suspect is caused by 'spurious regressors'. Notwithstanding these results for consumer confidence and CPI, we are able to generalise Scott and Varian's (2014a, 2014b) in-sample findings to an out-of-sample context for unemployment, for which the problem of spurious regressors appears minimal. In sum, it seems that online search behaviour is a relatively reliable gauge of an individual's personal situation (employment status), but is less reliable when it comes to variables that are unknown to the individual (CPI) or too general to be linked to specific search terms (consumer confidence).

Section 2 describes the data, while Section 3 describes the BSTS model and the Hamiltonian sampler. Section 4 presents the results for both in- and out-of-sample settings, followed by a brief exploration of alternative transformations and selection approaches. Finally, we interpret the findings in a broader context.

2. Data

2.1. Macroeconomic series

We obtain three macroeconomic series (unemployment, CPI and consumer confidence) for five countries (US, UK, Canada, Germany and Japan) from February 2004 to December 2016 at a monthly frequency (155 observations) from Bloomberg. These series and countries were selected to facilitate comparisons with Scott and Varian's (2014b) earlier findings. While Bloomberg does not report release dates for these series, we obtained approximate release dates from the reports of the national statistics agencies of the five countries investigated here. Based on this information, Table 1 shows the approximate time lags, measured in weeks, in the release dates of the series under investigation. The unemployment series shows signs of a trend and a seasonal component (Fig. 1), which are absent for consumer confidence and the CPI (Figs. 2 and 3). For unemployment, we take the natural logarithm and account for the trend and seasonality, while for consumer confidence and the CPI we model only the level. All data transformations are listed in Table A.4 in Appendix A.

2.2. Google trends

Google Trends is a public service that has been available since January 2004, providing time series of worldwide search activity for (i) specific (user-defined) search terms and (ii) predefined search categories. Queries in any category are assigned by Google to a particular country based on the IP address of the user.[2] For further details on the construction of the Google Trends data, see Stephens-Davidowitz and Varian (2014). For each macroeconomic

[1] We thank an anonymous referee for alerting us to the fact that the BSTS software has since been updated to allow the user to split the full sample into an in-sample period and an out-of-sample period.

[2] If the IP address of the user is unavailable, the domain of the search engine is used instead; e.g. queries from google.de are assigned to Germany.

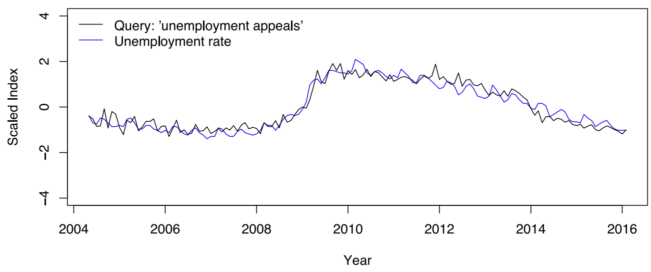

Fig. 1. Unemployment and the Google search term 'unemployment appeals' (US).

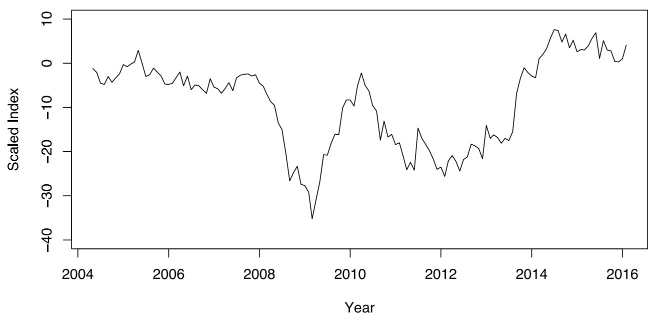

Fig. 2. Consumer confidence (UK).

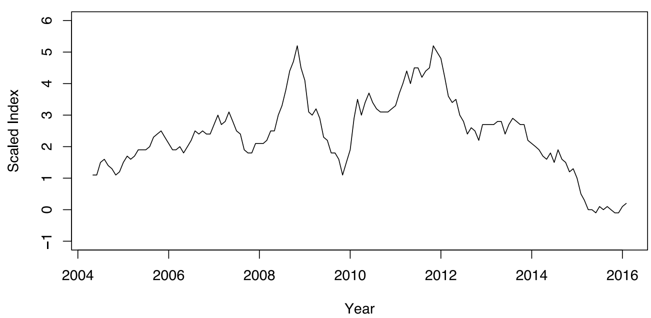

Fig. 3. The consumer price index (UK).

series in each country, we select approximately 60 distinct potentially-relevant Google categories (i.e., 3 × 60 categories per country). Each category consists of 155 monthly observations from February 2004 to December 2016. To illustrate, the categories selected for unemployment include 'unemployment appeals' and 'job listings'. The Google category data associated with unemployment often contain both trends and seasonal patterns, as is illustrated in Fig. 4 for the category 'job listings'. We 'whiten' the Google Trends data as per Scott and Varian (2014a) to ensure that the regression component does not interfere with the structural components of the BSTS model. We take first differences to remove the time-varying trend, deseasonalise to remove any time-constant seasonality, and demean the remainder. We select potentially-relevant Google categories once, based on their descriptions from Google, and eliminate any forward-looking bias by using only data that are available at the time of our nowcasts.

2.3. Google correlate

Like Google Trends, Google Correlate provides time series of Google search terms dating back to January 2004. Unlike Trends, though, Correlate returns multiple time series that are highly correlated with any (user-defined) series of interest. Naturally, we obtain time series that are correlated strongly (either positively or negatively) with our macroeconomic series. For example, Fig. 1 illustrates that the frequency of the search term 'unemployment appeals' tracks the macroeconomic US unemployment series closely. We select at most 50 positively and 50 negatively correlated queries for each macroeconomic series for each country and remove time series that are constant for more than 12 consecutive observations. Again, we 'whiten' the data and take the log of time series which we suspect to contain multiplicative noise; all transformations are listed in Table A.5 in Appendix A. We ensure that our nowcasts are genuinely out-of-sample by feeding only the historic

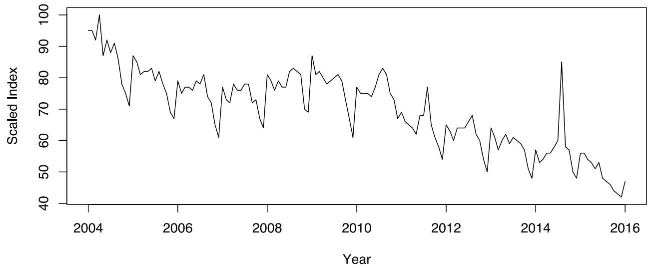

Fig. 4. Google category 'job listings' (US).

Table 1
Sources and approximate release lags of the macroeconomic series.

		Release lag (weeks)	Source
UN	US	≤1	Bureau of Labor Statistics
	GE	8	German Federal Statistical Office
	CA	1	Statistics Canada
	JA	4	Statistics Bureau, Ministry of Internal Affairs and Communications
	UK	6	UK Office for National Statistics
CPI	US	2	Bureau of Labor Statistics
	GE	2	German Federal Statistical Office
	CA	3	Statistics Canada
	JA	4	Statistics Bureau, Ministry of Internal Affairs and Communications
	UK	2	UK Office for National Statistics
CC	US	2	University of Michigan Consumer Sentiment Index
	GE	∗	ICON Wirtschafts- und Finanzmarktforschung
	CA	∗	∗
	JA	≤1	Economic and Social Research Institute Japan
	UK	4	European Commission

Notes: UN = unemployment, CPI = consumer price index, CC = consumer confidence, US = United States, GE = Germany, CA = Canada, JA = Japan, UK = United Kingdom, ∗ = release dates not found.

part of the macroeconomic series to Google Correlate. We amend our list of search terms annually, in January, after which the values of the selected series are updated monthly; that is, our out-of-sample nowcasts for 2015 are based on Google search terms that proved informative over the period from February 2004 to December 2014.

3. The BSTS model

3.1. Model formulation

The BSTS model (Scott and Varian, 2014a, 2014b) decomposes a time series y_t as the sum of its structural and regression components as follows:

$$
\begin{aligned}
y_t &= \mu_t + \tau_t + \boldsymbol{\beta}'\boldsymbol{x}_t + \varepsilon_t, && \varepsilon_t \sim N(0, \sigma_\varepsilon^2),\\
\mu_t &= \mu_{t-1} + \delta_{t-1} + u_t, && u_t \sim N(0, \sigma_u^2),\\
\delta_t &= \delta_{t-1} + \upsilon_t, && \upsilon_t \sim N(0, \sigma_\upsilon^2),\\
\tau_t &= -\sum_{s=1}^{S-1} \tau_{t-s} + w_t, && w_t \sim N(0, \sigma_w^2).
\end{aligned}
\tag{1}
$$

This allows for the presence of a trend with latent level μ_t, slope δ_t, and $S = 12$ monthly seasonal components $\{\tau_t, \tau_{t-1}, \ldots, \tau_{t-S+1}\}$. Together, these structural components form the state vector

$$\boldsymbol{\alpha}_t = (\mu_t, \delta_t, \{\tau_t, \tau_{t-1}, \ldots, \tau_{t-S+1}\})'$$

of the (implicit) state space model (see Appendix B). Furthermore, the triple $(\mu_t, \delta_t, \tau_t)'$ is subject to state innovations $\boldsymbol{\eta}_t = (u_t, \upsilon_t, w_t)'$, which are assumed to be independent such that their covariance matrix $\boldsymbol{Q}$ is diagonal. The $k \times 1$ regression component $\boldsymbol{x}_t$ that contains Google search data affects the (scalar) dependent variable y_t through the parameter vector $\boldsymbol{\beta}$. Finally, y_t is exposed to random observation noise ε_t that is independent of the state innovations. Henceforth, we suppress the subscripts t for denoting the entire time series, e.g. $\boldsymbol{y} := (y_1, y_2, \ldots, y_n)'$.

As our benchmark model, we take the model in Eq. (1) under the restriction $\boldsymbol{\beta} = \mathbf{0}$, such that no Google search data are used — the 'structural time series' (STS) model. Our benchmark is more sophisticated than the AR(1) benchmark, which is used often in the literature. An interesting extension would be to allow the variance of the error u_t to vary over time; see e.g. Stock and Watson (2007) or Clark (2011). However, we do not pursue this approach here in order to maintain comparability with Scott and Varian (2014a, 2014b).

3.2. Sampling

We estimate Eq. (1) by sampling from its full posterior $p(\boldsymbol{\alpha}, \boldsymbol{Q}, \boldsymbol{\beta}, \sigma_\varepsilon^2|\boldsymbol{y})$ using a Gibbs sampler. Specifically, the BSTS algorithm (Scott & Varian, 2014b) iterates over the following three steps:

1. Sample the states $\boldsymbol{\alpha}$ from $p(\boldsymbol{\alpha}|\boldsymbol{y}, \boldsymbol{Q}, \boldsymbol{\beta}, \sigma_\varepsilon^2)$ using Durbin and Koopman's (2002) state simulation smoother.
2. Sample the state variances $\boldsymbol{Q}$ from $p(\boldsymbol{Q}|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\varepsilon^2)$ as per Scott and Varian (2014a, p. 132).
3. (a) Select variables by drawing samples of the auxiliary variable $\boldsymbol{\gamma}$ using the SSVS or Hamiltonian sampler, and
   (b) Sample $\boldsymbol{\beta}$ and σ_ε^2 from $p(\boldsymbol{\beta}, \sigma_\varepsilon^2|\boldsymbol{y}, \boldsymbol{\alpha}, \boldsymbol{Q}, \boldsymbol{\gamma})$.

While the first two steps are standard, a more detailed description of the last step, a spike-and-slab regression using the two different samplers, is warranted before we move on to a description of our out-of-sample nowcasting procedure. We sample from the conditional posterior of $\boldsymbol{\beta}$ and σ_{ε}^{2} by using the SSVS algorithm with the conjugate spike-and-slab prior setup that was popularised by George and McCulloch (1997) and given in the context of the BSTS model by Eqs. (4)–(6) of Scott and Varian (2014b). The prior setup imposes a normal hierarchical mixture prior on the regression coefficients $\boldsymbol{\beta}$ by introducing a binary parameter vector $\boldsymbol{\gamma}$ that determines which regressors are included in the model. Conditional on $\boldsymbol{\gamma}$, the posterior distribution of $\boldsymbol{\beta}$ and σ_{ε}^{2} is the well-known posterior of an ordinary linear regression model with conjugate priors (see Eq. (7) of Scott and Varian (2014b)).

Alternative prior specifications, which are not explored here, include Carvalho, Polson, and Scott's (2009) horseshoe prior and Ročková and George's (2018) spike-and-slab lasso. We follow Scott and Varian (2014a, 2014b) in using the conjugate priors described above, as these are computationally tractable in combination with the sampler used.

Samples of the conditional posterior of $\boldsymbol{\gamma}$ (given by Eq. (8) of Scott and Varian (2014b)) are constructed by means of an (embedded) Gibbs sampling routine that draws sequentially from the conditional Bernoulli distribution of γ_i given $\boldsymbol{\gamma}_{-i}$. (Here, γ_i denotes the ith element of $\boldsymbol{\gamma}$, while $\boldsymbol{\gamma}_{-i}$ is the vector $\boldsymbol{\gamma}$ that excludes the ith element.) However, as Heaton and Scott (2010) point out, traditional Markov chain Monte Carlo (MCMC) variable selection methods, which are used for large sets of regressors, frequently miss regressor combinations with a high posterior probability. We use the Hamiltonian Monte Carlo (HMC) method, which is often more efficient at exploring the parameter space than traditional MCMC methods (Neal, 2011).

We sample from the posterior of $\boldsymbol{\gamma}$ using HMC by means of Pakman and Paninski's (2014) exact Hamiltonian sampler for binary variables. To that end, we augment the parameter space with a continuous random vector $\boldsymbol{z}$ of the same dimension as $\boldsymbol{\gamma}$. The auxiliary variable $\boldsymbol{z}$ is related to $\boldsymbol{\gamma}$ by means of

$$\gamma_i = \begin{cases} 0 & \text{if } z_i < 0, \\ 1 & \text{if } z_i \geq 0, \qquad \forall\, i = 1, 2, \ldots, k, \end{cases} \tag{2}$$

which we modified slightly from the work of Pakman and Paninski (2013) to match a binary variable defined on $\{0, 1\}$. The joint distribution of $\boldsymbol{z}$ and $\boldsymbol{\gamma}$ is then given by

$$p(\boldsymbol{\gamma}, \boldsymbol{z}) = p(\boldsymbol{\gamma})p(\boldsymbol{z}|\boldsymbol{\gamma}). \tag{3}$$

For $p(\boldsymbol{z}|\boldsymbol{\gamma})$, we adopt the truncated Gaussian distribution, following Pakman and Paninski (2014). The choice of $p(\boldsymbol{z}|\boldsymbol{\gamma})$ in combination with the posterior of $\boldsymbol{\gamma}$ leads to the following potential energy function:

$$\begin{aligned} U(\boldsymbol{z}) &= -\log p(\boldsymbol{z}|\boldsymbol{\gamma}) - \log p(\boldsymbol{\gamma}|\dot{\boldsymbol{y}}) \\ &\propto -\frac{\boldsymbol{z}'\boldsymbol{z}}{2} - \frac{1}{2}\log|\boldsymbol{\Omega}_{\gamma}^{-1}| + \frac{1}{2}\log|\boldsymbol{V}_{\gamma}^{-1}| \\ &\quad + \frac{\nu_{\epsilon} + n}{2}\log(ss_{\epsilon} + SS_{\gamma}) \\ &\quad - \boldsymbol{\iota}'\boldsymbol{\gamma}\log\varrho - (k - \boldsymbol{\iota}'\boldsymbol{\gamma})\log(1-\varrho), \end{aligned} \tag{4}$$

where the vector $\boldsymbol{\iota}$ consists of ones and is of appropriate length.

3.3. Out-of-sample nowcasts

We produce in-sample nowcasts of a macroeconomic variable y_{t+1} by estimating the model using the entire dataset, as is standard in the literature. On the other hand, when making out-of-sample nowcasts of y_{t+1} we must consider the (posterior) predictive distribution of y_{t+1} conditional on the information set $\mathcal{I}_{t+1}$, which contains the predictors up to (and including) time $t+1$, while the macroeconomic series are only included up to (and including) time t. To illustrate, on 1 February we may use US Google search data, where we include data from January, in order to produce a nowcast of the January US CPI, while the 'actual' CPI numbers are not released by the Bureau of Labor Statistics until two weeks later (mid-February). We obtain nowcasts (point predictions) by taking the mean of the posterior predictive distribution $p(y_{t+1}|\mathcal{I}_{t+1})$ and evaluate them using the root mean squared error (RMSE) criterion. We also report the mean absolute prediction error (MAPE) so as to facilitate comparisons with the previous literature.

4. Results

This section compares the BSTS and STS models in order to test whether Scott and Varian's (2014a, 2014b) in-sample results persist in an out-of-sample context for three macroeconomic series and five countries between March 2004 and December 2016 (154 monthly observations).[3] Like Scott and Varian (2014a, 2014b), we focus on nowcasts at a monthly frequency. For the out-of-sample analysis, we use an initial estimation window from March 2004 to August 2012 (104 observations, roughly two-thirds of the data) to produce predictions for the remaining period using an expanding window. We present results based on (i) exclusively category (Trends) data and (ii) both category and Correlate data. Furthermore, for each of these we use both the SSVS and the Hamiltonian sampler, leading to four different BSTS models. The STS model nowcasts are used as the benchmark. We report two performance measures — the root mean square error (RMSE) and the mean absolute prediction error (MAPE) — for all five models, five countries and three macroeconomic series, leading to $2 \times 5 \times 5 \times 3 = 150$ numbers. We report these numbers separately for the in-sample (Table 2) and out-of-sample (Table 3) settings.

We facilitate across-country comparisons by ranking all models separately for each country. This allows us to calculate an average (across-country) rank for each model, where rank 1 denotes the best predictions.

[3] The number of nowcasts is one fewer than the number of observations, as we use first differences to make the nowcasts.

Table 2
In-sample nowcasts at a monthly frequency for the unemployment rate, CPI and consumer confidence for all countries relative to the benchmark model (STS).

		MAPE					RMSE				
			Category		Category & Correlate			Category		Category & Correlate	
		STS	SSVS	HAM	SSVS	HAM	STS	SSVS	HAM	SSVS	HAM
	US	2.729	−0.009	−0.029	−0.030	−0.109	3.711	−0.007	−0.024	−0.059	−0.140
	GE	1.145	−0.023	−0.024	−0.019	−0.015	1.807	−0.019	−0.019	−0.017	−0.014
UN	CA	2.510	−0.013	+0.002	−0.011	−0.005	3.468	−0.020	−0.016	−0.010	−0.013
$\times 10^{-2}$	JA	3.266	−0.006	−0.026	−0.056	−0.006	4.166	−0.034	−0.046	−0.083	−0.040
	UK	1.139	−0.040	−0.019	−0.038	−0.056	1.455	−0.065	−0.044	−0.060	−0.083
Average rank		4.6	2.4	3	2	2.4	5	2.4	2.4	2.6	2.4
	US	3.663	+0.009	+0.012	−0.006	−0.022	5.153	−0.018	−0.006	−0.023	−0.051
	GE	2.319	−0.118	−0.083	−0.052	−0.114	3.052	−0.160	−0.106	−0.057	−0.147
CPI	CA	3.354	−0.039	−0.021	∗	∗	4.266	−0.108	−0.066	∗	∗
$\times 10^{-1}$	JA	2.272	+0.022	+0.007	+0.008	+0.005	3.275	−0.017	−0.026	−0.014	−0.015
	UK	2.289	+0.022	+0.018	∗	∗	2.986	−0.002	−0.008	∗	∗
Average rank		3	3.3	3.7	3.3	1.7	5	2	2.3	3.3	2
	US	3.295	−0.134	−0.098	−0.087	−0.028	4.2431	−0.1983	−0.1718	−0.1370	−0.0375
	GE	1.851	+0.029	+0.018	+0.027	+0.020	2.4854	+0.0040	−0.0138	−0.0052	−0.0113
CC	CA	4.352	−0.145	−0.167	−0.290	−0.416	5.6590	−0.2054	−0.2385	−0.4132	−0.5489
	JA	1.202	−0.019	−0.008	−0.024	−0.029	1.5390	0.0329	−0.0112	−0.0343	−0.0472
	UK	2.725	−0.019	−0.012	−0.001	−0.002	2.0453	−0.0068	−0.0014	+0.0018	+0.0025
Average rank		4.2	2.8	2.6	3	2.4	4.4	2.8	2.4	2.6	2.6

Notes: UN = unemployment, CPI = consumer price index, CC = consumer confidence, US = United States, GE = Germany, CA = Canada, JA = Japan, UK = United Kingdom, STS = structural state space (benchmark) model, SSVS = stochastic search variable selection sampler, HAM = Hamiltonian sampler, MAPE = mean absolute prediction error, RMSE = root mean square error. The table shows the absolute difference in MAPE and RMSE between the Bayesian structural time series (BSTS) model with two different samplers, SSVS and Hamiltonian, and the benchmark model. All models use data from March 2004 to December 2016 (154 monthly observations). Two different sets of regressors are tested for each sampler: a set with only category data and a set with both category and Correlate data. The Correlate series of CPI for CA and UK could not be obtained (indicated by ∗). For each country and performance criterion, the models are ranked from 1 to 5, where 1 is the best performing model. The average rank for each model is given; for CPI, this is calculated excluding CA and UK.

Table 3
Out-of-sample nowcasts at a monthly frequency for the unemployment rate, CPI and consumer confidence of all countries relative to the benchmark model (STS).

		MAPE					RMSE				
			Category		Category & Correlate			Category		Category & Correlate	
		STS	SSVS	HAM	SSVS	HAM	STS	SSVS	HAM	SSVS	HAM
	US	2.563	−0.072	−0.037	−0.088	−0.094	3.267	−0.165	−0.138	−0.218	−0.208
	GE	0.877	+0.064	+0.051	+0.050	+0.051	1.032	+0.087	+0.073	+0.072	+0.072
UN	CA	1.870	−0.034	−0.062	−0.047	−0.026	2.349	−0.002	−0.013	+0.011	+0.021
$\times 10^{-2}$	JA	2.910	−0.018	−0.046	+0.004	+0.041	3.749	+0.090	+0.033	+0.065	+0.068
	UK	1.260	−0.067	−0.078	−0.048	−0.057	1.607	−0.185	−0.195	−0.159	−0.173
Average rank		3.8	3	2.2	2.8	3	3	3.4	2.4	2.8	3.4
	US	2.538	+0.081	+0.045	+0.058	+0.017	3.150	+0.046	+0.040	+0.041	+0.016
	GE	2.086	−0.045	−0.143	−0.121	−0.054	2.626	+0.010	−0.099	−0.022	−0.002
CPI	CA	2.558	+0.022	−0.010	∗	∗	3.129	+0.005	−0.050	∗	∗
$\times 10^{-1}$	JA	2.442	+0.051	+0.055	+0.025	+0.070	3.973	+0.001	+0.014	+0.006	+0.056
	UK	1.750	+0.055	+0.040	∗	∗	2.257	+0.045	+0.019	∗	∗
Average rank		2.3	4	2.7	2.7	3.3	2	4	2.7	3	3.3
	US	2.837	−0.023	+0.017	+0.033	−0.075	3.571	−0.058	−0.017	−0.045	−0.073
	GE	1.039	+0.042	+0.036	+0.032	+0.030	1.494	+0.011	+0.008	+0.002	+0.009
CC	CA	3.120	+0.064	+0.099	+0.128	+0.085	3.800	+0.104	+0.137	+0.247	+0.209
	JA	1.106	+0.005	+0.015	+0.021	+0.022	1.418	+0.004	−0.002	+0.005	+0.016
	UK	2.128	+0.022	+0.033	+0.009	+0.018	2.918	+0.017	+0.015	−0.015	+0.017
Average rank		1.4	3	4	3.8	2.8	2.2	3.4	2.8	3	3.6

Note: The initial estimation window covers the period from March 2004 to August 2012 (104 observations). Out-of-sample nowcasts are made for the remaining period from September 2012 to December 2016 (50 observations) using an expanding window. See also the notes to Table 2.

We use the same default prior settings as Scott and Varian (2014b) across all series and models, which implies $\kappa = 1$, $w = 0.5$, $\nu = 0.01$, $R_e^2 = 0.5$ and the expected model size $m = 5$. For the Hamiltonian sampler we use a static travel time of $T = 2\frac{1}{2}\pi$. We draw 3000 samples from the posterior distribution and use a burn-in period of 1000 draws for all series and models, which proved sufficient for stable predictions.[4]

4 Increasing the number of samples to 20,000 for selected periods reduced the variance of the posterior mean predictions, but did not

4.1. In-sample estimates

In an in-sample context, we find that the BSTS models generally produce more accurate estimates than the STS benchmark for all macroeconomic series under investigation and all countries, irrespective of the performance measure used (Table 2). The relative improvement over the benchmark is in the range of 1%–5% for both performance measures.[5] The BSTS model using both category and Correlate data does not consistently improve on the BSTS model without Correlate data, irrespective of the sampler used. For the data investigated here, the Hamiltonian sampler does not appear to outperform the SSVS sampler.

4.2. Out-of-sample nowcasts

In an out-of-sample context, the BSTS models generally produce more accurate predictions than the STS benchmark for the unemployment series, but not for the consumer confidence and CPI series (Table 3). This finding seems to hold for most countries and both performance measures.

For the unemployment series, the use of Google category data leads to gains for four out of five countries (with Germany being the exception), while using both category and Correlate data leads to gains for three out of five countries (with Germany and Japan being the exceptions). The improvements are in the range of 1%–5%, which are relatively modest gains; however, recall that our in-sample results were in the same range. In this light, the fact that Google search data yields roughly the same improvements in both the in- and out-of-sample contexts testifies to its robust value in predicting unemployment.

For consumer confidence and CPI, on the other hand, we find that the use of Google category data does not improve our out-of-sample nowcasts systematically. For consumer confidence in particular, the nowcast errors are larger than those of the benchmark. Likewise, we find that the use of Google Correlate data does not improve our nowcasts of consumer confidence and CPI in an out-of-sample context. Instead, these correlations often break down after the estimation period on which they are based, rendering them useless for out-of-sample nowcasts. Indeed, the results may be worse than those obtained using category data alone. Thus, the strength of Google Correlate, i.e., the ability to return many potentially relevant series, is also its weakness, since it may also identify many search queries that are highly correlated with a given time series even in the absence of any underlying (predictive) relationship. We investigate the number of spurious correlations by focusing on the US and simply counting the number of correlated series for which the out-of-sample correlation is less than half the in-sample correlation. For both consumer confidence and CPI, the majority of the 89 series retrieved can be classified as spurious (48 and 77, respectively), which explains why the BSTS models with Correlate data do not outperform those without. For unemployment, on the other hand, we find only one spurious correlation.[6]

The best-performing version of the BSTS model for US unemployment uses both category and Correlate data. Fig. 5 depicts the cumulative squared prediction errors (sum of squared errors, SSE) over time for both the benchmark model and the BSTS model, again using both samplers. Prediction errors accumulate slowly but consistently in all models during the initial estimation window from March 2004 to August 2012, but more quickly for the benchmark model. Thus, the added value of using Google search data is spread out over time, and all nowcasts are improved somewhat. However, some improve more than others, since we see an upward shift in the SSE of the benchmark model during the 2008 financial crisis relative to both BSTS models. This echoes Choi and Varian's (2012) finding that Google search data can be especially valuable in predicting turning points, such as financial crises. Following our initial estimation window, the end of which is indicated by the dotted line, the SSE of the benchmark model continues to diverge from those of the two BSTS models (perhaps even at a slightly faster rate), confirming our view that Google category and Correlate data have robust out-of-sample predictive value for unemployment.

4.3. Sensitivity analysis

This section zooms in on the US macroeconomic series and considers how our out-of-sample results change if we use other transformations, selection approaches and data frequencies. As the results in the previous section suggest that Google Correlate data is of limited use in our application, we focus on the Google category data alone. The BSTS model is designed to handle a large number of predictors, but there is still a bias–variance trade-off at the heart of its effectiveness. It could be argued that including 50 to 75 (monthly) categories is not necessarily optimal with respect to this trade-off. Therefore, we explore whether either using fewer categories or using category data at a weekly frequency affects our results. Specifically, we (i) use category data at a weekly frequency and apply the usual transformations; (ii) log difference the category series but do not remove the structural components; (iii) difference the category series but do not remove the structural components; (iv) difference the category series and remove the structural components; and (v) select only 10 to 20 categories for each of the three macroeconomic series and apply the transformations as usual.

For the unemployment series, we find that the prediction errors of the BSTS model with weekly category data are lower than those of the model with monthly category data: the improvements in the prediction errors are in the range 1%–3% for both the Hamiltonian

noticeably improve our predictions or change the relative performances of the models.

[5] Scott and Varian (2014a) report a relative improvement of roughly 14% for the BSTS model over an AR(1) model for the US consumer confidence series. Our findings relative to an AR(1) model (not reported) are in line with this result.

[6] US unemployment is correlated highly with the search term ‘spider solitaire’ in the in-sample period but not in the out-of-sample period. While one may be tempted to speculate that playing computer games leads to unemployment, this correlation is spurious.

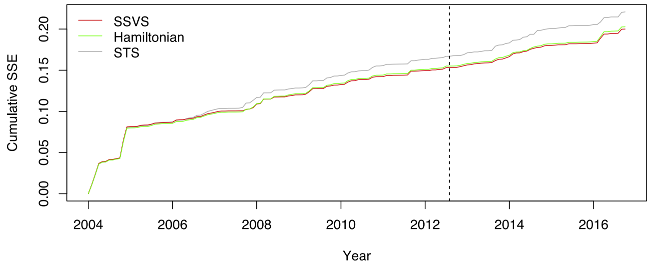

Fig. 5. US unemployment cumulative SSE values for the BSTS models with the SSVS and Hamiltonian samplers and for the STS model. The BSTS models use Google Correlate and category data.

and SSVS samplers. However, caution is needed in interpreting this as evidence that aggregating Google search queries leads to information loss, as fewer categories were available for weekly data, which could be argued to simplify the variable-selection problem. For (ii)–(v), we also find that the general result of Section 4.2 holds: Google search data help to nowcast unemployment but not CPI and consumer confidence. The MAPEs and RMSEs of the BSTS models are lower than those of the STS model for the unemployment series, whereas the results for the consumer confidence and CPI series are not improved consistently compared to those of the STS model. The selected categories and corresponding out-of-sample results are available on request.

5. Discussion and conclusion

We found that the BSTS model outperforms the STS model in an in-sample setting for all three macroeconomic series, confirming the in-sample results reported by Scott and Varian (2014b). The out-of-sample outperformance persisted only for unemployment: for four out of five countries when using category data, and for three of out five countries when using both category and Correlate data. In other words, we have been able to generalise Scott and Varian's (2014a, 2014b) in-sample findings for unemployment to an out-of-sample context, but not those for consumer confidence and CPI. In addition, we have demonstrated the viability of using the Hamiltonian sampler for the BSTS model, although it appeared to have little added value over the SSVS sampler for this particular application.

We conclude from these findings that Google search data appear to be most helpful when the series under investigation is related directly to an individual's personal situation and is linked closely with specific search behaviours (such as employment status), but is less reliable when it comes to macroeconomic measures that are unknown to the individual (such as CPI) or too general to be linked to specific search terms (such as consumer confidence). For example, many unemployed people may have known in advance that they were at risk of becoming unemployed, and this knowledge would have generated specific and predictable online search behaviours. Conversely, few individuals can estimate monthly CPI figures precisely, and even if they could, the impact of this on their search behaviour is likely to be either minimal or subject to high individual variation. Similarly, although in principle consumer confidence is determined by a sum over households, each of which can be assumed to know whether or not confidence is warranted based on its own circumstances, this knowledge does not appear to be sufficient to generate specific and predictable search behaviours. Our finding that improvements over the baseline model are confined to predictions of macroeconomic series that have particularly close relationships with user search behaviour echoes work in the field of consumer action; for example, Goel, Hofman, Lahaie, Pennock, and Watts (2010) find that search data are predictive of specific consumer actions that occur in the near future, such as going to the cinema.

The weak links between search behaviour and both CPI and consumer confidence are likely to be among the main causes of the many spurious queries obtained by Google Correlate. The monthly frequency of the macroeconomic series yields only a limited number of observations (155 observations starting in February 2004); thus, search queries that are genuinely related to our macroeconomic series may be swamped by the many spurious correlations. Possibly for the same reason, both our in-sample and out-of-sample predictions of unemployment improved when we used weekly rather than monthly data, even though (or perhaps because) fewer categories were available. For CPI and consumer confidence, these spurious correlations cannot be filtered out effectively, and researchers who wish to predict such variables may be better off hand-picking Google search terms.

Finally, our results are generally consistent across countries. A notable exception is Germany, for which unemployment nowcasts were not improved by Google search data. Although we have no immediate explanation for this exception, we note that unemployment dropped steadily in Germany over the years following the financial crisis, unlike in the other countries investigated. Additional research into more macroeconomic series in different regions could further test the robustness of our results.

Appendix A. Data transformations

We took log differences of the categories retrieved from Google Trends; the differenced series are more meaningful to interpret, both economically and statistically, given the downward trend. Next, we removed the remaining structural components of the log-differenced

$$\boldsymbol{Z} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{T} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Box I.

series to avoid interference with the structural component of the BSTS model. Intuitively, if the structural components of a Google category series are of importance for modelling a macroeconomic series, a seasonal or trending pattern should be seen in the series itself. Since the structural components are already modelled in the BSTS model, they can be removed safely from the Google category series. These transformations effectively 'whiten' the category data. We decided not to deseasonalise or detrend the Google Correlate data, as these consist of more specific search queries with structural components that do not necessarily appear to be stable over time. The specific transformations of the Google search data are shown in Table A.4.

For the macroeconomic series, we took the log of unemployment, which is likely to have multiplicative noise, as the magnitude of the shocks depends on the level. As the transformed unemployment series for our sample still seemed to contain a trend and a seasonal component, we detrended it.

Table A.4
Transformations and structural components of macroeconomic data.

		Transformations	Structural components		
		Log	Level	Trend	Seasonal
UN	US	✓	✓	✓	✓
	GE	✓	✓	✓	✓
	CA	✓	✓	✓	✓
	JA	✓	✓	✓	✓
	UK[a]	✓	✓	✓	
CPI	US		✓		
	GE		✓		
	CA		✓		
	JA		✓		
	UK		✓		
CC	US		✓		
	GE		✓		
	CA		✓		
	JA		✓		
	UK		✓		

[a]UK unemployment data was only available in seasonally-adjusted form.

Appendix B. State space matrices

A generic linear Gaussian state space model formulation is:

$$\begin{aligned} y_t &= \boldsymbol{Z}'\boldsymbol{\alpha}_t + \boldsymbol{\beta}'\boldsymbol{x}_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{T}\boldsymbol{\alpha}_t + \boldsymbol{R}\,\boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim N(\boldsymbol{0}, \boldsymbol{Q}), \end{aligned} \tag{B.1}$$

for $t = 1, \ldots, n$. The observation equation contains a (scalar) dependent variable y_t, an $m \times 1$ latent state vector $\boldsymbol{\alpha}_t$, a $k \times 1$ regression component $\boldsymbol{x}_t$ and a random observation noise ε_t with variance σ_ε^2. The matrix $\boldsymbol{Z}$ and the vector $\boldsymbol{\beta}$, which are assumed to be of appropriate dimensions, describe how the state $\boldsymbol{\alpha_t}$ and the regression

Table A.5
Transformations of the Google search data.

		Log	Difference	Detrend	Deseasonalize	Demean
Category	UN	✓	✓	✓	✓	✓
	CPI	✓	✓	✓	✓	✓
	CC	✓	✓	✓	✓	✓
Correlate	UN	✓	✓			
	CPI		✓			
	CC		✓			

component $\boldsymbol{x}_t$, respectively, influence the observation y_t. The state transition equation contains a (square) 'transfer' matrix $\boldsymbol{T}$, a 'selector' matrix $\boldsymbol{R}$, and a state disturbance vector $\boldsymbol{\eta}_t$ with covariance matrix $\boldsymbol{Q}$. In Box I, we specify the system matrices $\boldsymbol{T}$, $\boldsymbol{R}$ and $\boldsymbol{Z}$ that are used to obtain the BSTS model.

References

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

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2009). Handling sparsity via the horseshoe. In *Proceedings of the 12th international conference on artificial intelligence and statistics* (pp. 73–80).

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

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, *29*(3), 327–341.

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

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, *89*(3), 603–616.

George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, *7*(2), 339–373.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with web search. *Proceedings of the National Academy of Sciences*, *107*(41), 17486–17490.

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

Heaton, M. J., & Scott, J. (2010). *Frontiers of statistical decision making and Bayesian analysis*. New York: Springer.

Lindberg, F. (2011). *Nowcasting Swedish retail sales with Google search query data* (Master's Thesis), Stockholm University.

McLaren, N., & Shanbhogue, R. (2011). Using Internet Search Data as Economic Indicators. Bank of England Quarterly Bulletin, No. 2011 Q2.

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, *130*, 114–122.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo, Vol. 2* (pp. 113–162).

Pakman, A., & Paninski, L. (2013). Auxiliary-variable exact hamiltonian monte carlo samplers for binary distributions. In *Advances in neural information processing systems* (pp. 2490–2498).

Pakman, A., & Paninski, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, *23*(2), 518–542.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*, 1684.

Ročková, V., & George, E. I. (2018). The spike-and-slab lasso. *Journal of the American Statistical Association*, *113*(521), 431–444.

Scott, S. L., & Varian, H. R. (2014a). Bayesian variable selection for nowcasting economic time series. In A. Goldfarb, S. M. Greenstein, & C. E. Tucker (Eds.), *Economic analysis of the digital economy* (pp. 119–135). University of Chicago Press.

Scott, S. L., & Varian, H. R. (2014b). Predicting the present with Bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, *5*(1–2), 4–23.

Stephens-Davidowitz, S., & Varian, H. (2014). A Hands-on Guide to Google Data. Technical report.

Stock, J. H., & Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*(s1), 3–33.

Suchoy, T. (2009). Query Indices and a 2008 Downturn: Israeli Data. Technical report, Bank of Israel.

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: Survey-based indicators vs. Google trends. *Journal of Forecasting*, *30*(6), 565–578.

Yu, L., Zhao, Y., Tang, L., & Yang, Z. (2019). Online big data-driven oil consumption forecasting with Google Trends. *International Journal of Forecasting*, *35*(1), 213–223.

Rutger-Jan Lange is an assistant professor in Quantitative Finance at the Econometric Institute at the Erasmus University Rotterdam. He holds a PhD in management science and operations research from the University of Cambridge (UK) and has published in econometrics, finance and theoretical physics. Robin Niesert, Jochem Oorschot, Chris Veldhuisen and Kester Brons were masters students in Econometrics and Management Science at the Erasmus University Rotterdam at the time this research was done.