TI 2019-021/III
Tinbergen Institute Discussion Paper

Can Google Search Data Help Predict Macroeconomic Series?

Robin Niesert
Jochem Oorschot[1]
Chris Veldhuisen
Kester Brons
Rutger-Jan Lange[2]

[1] PhD ESE Econometric Institute, Erasmus University, TI
[2] Assistant Professor ESE, Econometric Institute, Erasmus University

Can Google Search Data Help Predict Macroeconomic Series?

Robin Niesert, Jochem Oorschot[1],
Chris Veldhuisen, Kester Brons, Rutger-Jan Lange

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

Abstract

We use Google search data with the aim of predicting 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. To our knowledge, the more challenging question of whether such data have out-of-sample predictive value has not yet been satisfactorily answered. 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 and consumer confidence. It may be that online search behaviour is 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).

Keywords: Bayesian methods, forecasting practice, Kalman filter, macroeconomic forecasting, state space models, nowcasting, spike-and-slab, Hamiltonian sampler

[1]Corresponding author (oorschot@ese.eur.nl).

Accepted for publication by the International Journal of Forecasting, December 2018

1. Introduction

Timely and accurate economic data is invaluable in making sensible investment and policy decisions. Unfortunately, many macroeconomic time series are released with a substantial time lag and 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 mean absolute prediction error (MAPE) have been found for US inflation (Guzman, 2011), the UK housing market (McLaren and Shanbhogue, 2011), Swedish private consumption (Lindberg, 2011), German and Israeli unemployment (Askitas and Zimmermann, 2009; Suchoy, 2009) and US private consumption (Vosen and 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 is predictive of US unemployment irrespective of whether the out-of-sample period starts before, during or after the Great Recession. Similarly, Preis et al. (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, consumer price index (CPI) and consumer confidence) for five countries (US, UK, Canada, Germany and Japan). We follow Scott and Varian (2014a,b) 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 specific search term, 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 and Varian (2014a,b) 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 produce monthly 'nowcasts' of the macroeconomic variables and found that the resulting 'in-sample predictions' outperformed an AR(1) benchmark as well as a structural time series (STS) model in terms of MAPE.[2]

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 performance of Google search data, although they are typically limited to hand-selected series from Google Trends, while ignoring Google Correlate. For example, Choi and Varian (2012) show that the categories 'trucks & SUVs' and 'automotive insurance' help predict motor vehicle sales, while D'Amuri and Marcucci (2017) show that the 'jobs' category helps forecast US unemployment. Similarly, Naccarato et al. (2018) use the frequency of the search term 'job offers' to forecast Italian youth unemployment, and Yu et al. (2018) use the search terms 'oil consumption', 'oil inventory' and 'oil price' to predict (changes in) oil consumption. Arguably, all 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 to make effective out-of-sample forecasts. This is no easy task: Scott and Varian (2014b) (p. 21) themselves note that a disadvantage of using Google Correlate is that the strongest (in-sample) predictors are often 'spurious regressors' lacking 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 in making out-of-sample nowcasts. Given the high number of potentially relevant time series obtained from Google Trends and Correlate, the selection of variables

[2] 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 and out-of-sample period.

is particularly challenging. For this purpose, Scott and Varian (2014a,b) integrate into the BSTS model a spike-and-slab regression with the stochastic search variable selection (SSVS) sampler (George and McCulloch, 1997). However, the SSVS sampler may suffer when the number of predictors or the multicollinearity among them is high; see e.g. Heaton and Scott (2010). We deviate from Scott and Varian (2014a,b) 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 at a monthly frequency of the BSTS model 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, however, 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 we suspect is caused by 'spurious regressors'. Notwithstanding these results for consumer confidence and CPI, we are able to generalise Scott & Varian's (2014a,b) 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 an in- and out-of-sample setting, 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, consumer confidence) for five countries (US, UK, Canada, Germany, Japan) from February 2004 to December 2016 at a monthly frequency (155 observations) from Bloomberg. These series and countries were selected to facilitate comparison 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 lag, measured in weeks, in the release dates of the series under investigation. The unemployment series shows signs of a trend and seasonal component (Figure 1), which are absent for consumer confidence and CPI (Figures 2 and 3). For unemployment we take the natural logarithm and account for the trend and seasonality, while for consumer confidence and CPI we model only the level. All data transformations are listed in Table A.4 in Appendix A.

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.

2.2. Google Trends

Google Trends is a public service available from 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.[3] For more details on the construction of the Google Trends data, see Stephens-Davidowitz and Varian (2014). For each macroeconomic 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, categories selected for unemployment include 'unemployment appeals' and 'job listings'. Google category data associated with unemployment often contains both trends and seasonal patterns, as illustrated in Figure 4 for the category 'job listings'. We 'whiten' the Google Trends data as in 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 description by Google, and eliminate any forward-looking bias by using only data 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, however, Correlate returns multiple time series that are highly correlated with any (user-defined) series of interest. Naturally, we obtain time series that are strongly (positively or negatively) correlated with our macroeconomic series. For example, Figure 1 illustrates that the frequency of the search term 'unemployment appeals' closely

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

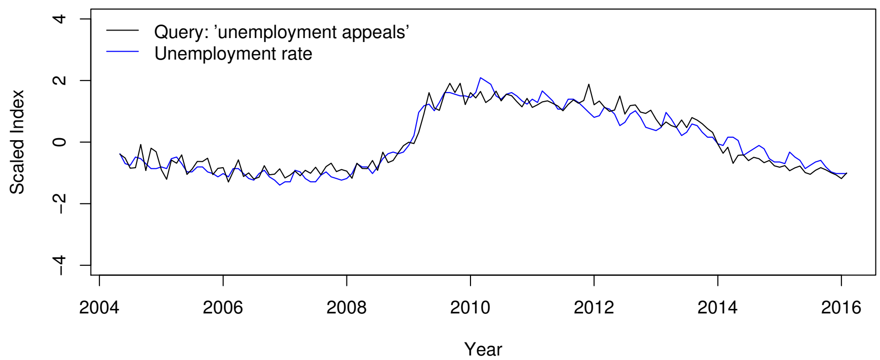

Figure 1: Unemployment and Google search term 'unemployment appeals' (US)

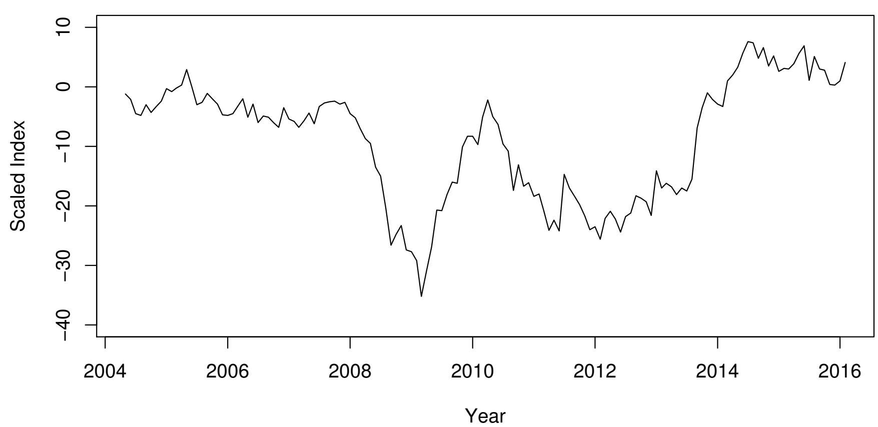

Figure 2: Consumer confidence (UK)

tracks the macroeconomic US unemployment series. We select at most 50 positively and 50 negatively correlated queries for each macroeconomic series per 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. To make genuine out-of-sample nowcasts, we feed only

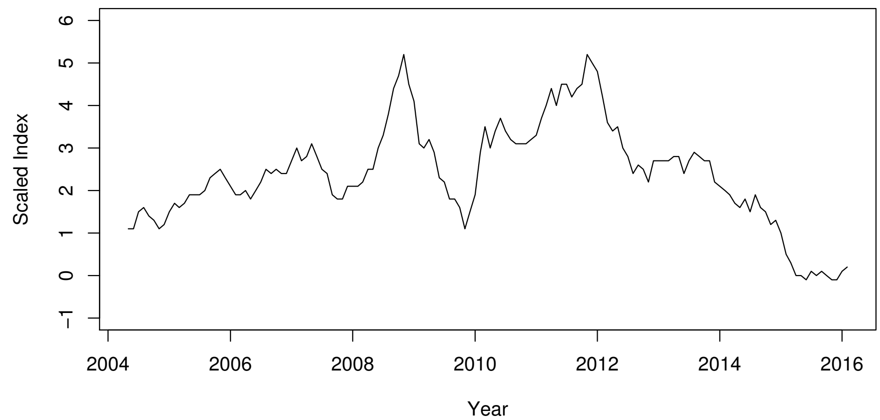

Figure 3: Consumer price index (UK)

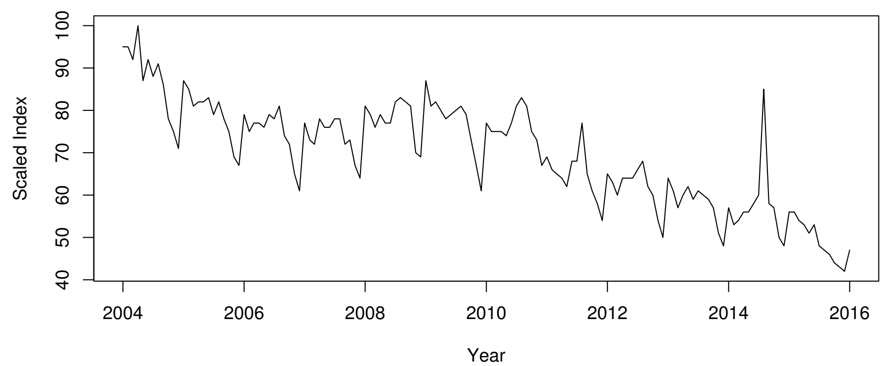

Figure 4: Google category 'job listings' (US)

the historic 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 in the period from February 2004 to December 2014.

3. The BSTS model

3.1. Model formulation

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

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

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

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

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 $\boldsymbol{x}_t$ containing 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 to denote the entire time series, e.g. $\boldsymbol{y} := (y_1, y_2, ..., y_n)'$.

As our benchmark model, we take model (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 often used 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). To maintain comparability with Scott and Varian (2014a,b), however, we do not pursue this approach here.

3.2. Sampling

To estimate model (1), we sample 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 and 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 in 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, spike-and-slab regression using the two different samplers, is warranted before we move onto a description of our out-of-sample nowcasting procedure. To sample from the conditional posterior of $\boldsymbol{\beta}$ and σ_ε^2, we use the SSVS algorithm with the conjugate spike-and-slab prior setup, popularised by George and McCulloch (1997) and given in the context of the BSTS model by equations (4)-(6) in 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 equation (7) in Scott and Varian (2014b)).

Alternative prior specifications, which are not explored here, include Carvalho et al.'s (2009) horseshoe prior and Ročková and George's (2016) spike-and-slab lasso. We follow Scott and Varian (2014a,b) 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 equation (8) in Scott and Varian (2014b)) are constructed by means of an (embedded) Gibbs sampling routine that sequentially draws from the conditional Bernoulli distribution of γ_i given $\boldsymbol{\gamma}_{-i}$. (Here, γ_i denotes the i-th element of $\boldsymbol{\gamma}$, while $\boldsymbol{\gamma}_{-i}$ is the vector $\boldsymbol{\gamma}$ excluding the i-th 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 than traditional MCMC methods at exploring the parameter space (Neal, 2011).

To sample from the posterior of $\boldsymbol{\gamma}$ using HMC, we use 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

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

which we modified slightly from 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:

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

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

3.3. Out-of-sample nowcasts

To make in-sample nowcasts of a macroeconomic variable y_{t+1}, the model is estimated using the entire dataset, as is standard in the literature. To make out-of-sample nowcasts of y_{t+1}, on the other hand, 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 US CPI in January, while '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 these using the root mean squared error (RMSE) criterion. We also report the mean absolute prediction error (MAPE) to facilitate comparison with previous literature.

4. Results

This section compares the BSTS and STS models 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).[4] Like Scott and Varian (2014a,b), 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. Further, 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 – root mean square error (RMSE) and 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.

To facilitate across-country comparisons, we rank 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.

We use the same default prior settings as in 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 $3,000$ samples from the posterior distribution and use a burn-in of $1,000$ draws for all series and models, which proved sufficient for stable predictions.[5]

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

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

[5] Increasing the number of samples to $20,000$ for selected periods reduced the variance of the posterior mean predictions, but did not noticeably improve our predictions or change the relative performance of the models.

used (Table 2). The relative improvement over the benchmark is in the range of $1-5\%$ for both performance measures.[6] The BSTS model using both category and Correlate data does not consistently improve over 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, using Google category data leads to gains for four out of five countries (Germany being the exception), while using both category and Correlate data leads to gains for three out of five countries (Germany and Japan being the exceptions). Improvements are in the range of $1-5\%$ percent – relatively modest gains, but recall that our in-sample results were in the same range. In this light, the fact that Google search data yields roughly the same improvement in both 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 using Google category data does not systematically improve our out-of-sample nowcasts. For consumer confidence in particular, the nowcast errors are larger than those of the benchmark. We find that using 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. The strength of Google Correlate, i.e. the ability to return many potentially

[6] Scott and Varian (2014a) report a relative improvement of roughly 14 percent 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.

relevant series, is thus also its weakness, since it can also identify many search queries that are highly correlated with a given time series even in the absence of any underlying (predictive) relationship. To investigate the number of spurious correlations, we focus on the US and simply count the number of correlated series for which the out-of-sample correlation is less than half the in-sample correlation. For consumer confidence and CPI, the majority of the 89 retrieved series 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[7]

The best performing version of the BSTS model for US unemployment uses both category and Correlate data. Figure 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. The added value of using Google search data is thus spread out over time; all nowcasts are somewhat improved. However, some improve more than others, since during the 2008 financial crisis we see an upward shift in the SSE of the benchmark model 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. After our initial estimation window, the end of which is indicated by the dotted line, the SSE of the benchmark model continues to diverge from that of both 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.

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

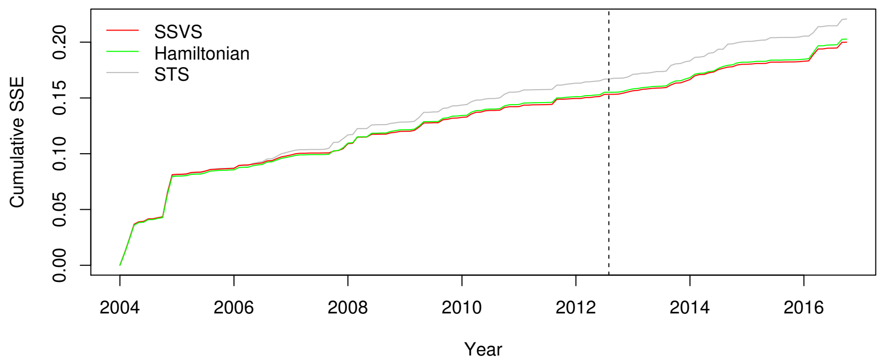

Figure 5: US unemployment cumulative SSE for the BSTS model with SSVS and Hamiltonian sampler and the STS model. The BSTS models use Google Correlate and category data.

4.3. Sensitivity analysis

In this section we zoom in on the US macroeconomic series and consider 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 Google category data alone. The BSTS model is designed to handle a large number of predictors, but at the heart of its effectiveness is still a bias-variance trade-off. It may be argued that including 50 to 75 (monthly) categories is not necessarily optimal with respect to this trade-off. Therefore, we explore whether the use of 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, (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 monthly category data: improvements in the prediction errors range between $1-3\%$ for both the Hamiltonian and the SSVS sampler. 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 arguably simplifies the variable-selection problem. For (ii)-(v), we also find that the general result of section 4.2 holds: Google search data help 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 consistently improved compared to those of the STS model. The selected categories and corresponding out-of-sample results are available on request.

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

		MAPE					RMSE				
			Cat.		Cat. & Corr.			Cat.		Cat. & Corr.	
		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
$*10^{-2}$	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	*	*
$*10^{-1}$	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 of the Bayesian structural time series (BSTS) model with two different samplers, SSVS and Hamiltonian, relative to the benchmark model. All models use data from March 2004 to December 2016 (154 monthly observations). For each sampler two different sets of regressors are tested; 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 per model is given; for CPI this is calculated excluding CA and UK.

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

		MAPE					RMSE				
			Cat.		Cat. & Corr.			Cat.		Cat. & Corr.	
		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
$*10^{-2}$	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	*	*
$*10^{-1}$	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: Same notes as table 2 apply. The initial estimation window covers 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.

5. Discussion and conclusion

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

From these findings we conclude that Google search data appear most helpful when the series under investigation directly relates to an individual's personal situation and is closely linked with specific search behaviour (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, knowledge that would have generated specific and predictable online search behaviour. Conversely, few individuals can precisely estimate monthly CPI figures and, even if they could, the impact on their search behaviour is likely to be either minimal or subject to high individual variation. Similarly, although consumer confidence is in principle determined by a sum over households, each of which can be assumed to know whether confidence is warranted (or otherwise) based on its own circumstances, this knowledge appears insufficient to generate specific and predictable search behaviour. Our finding that improvements over the baseline model are confined to predictions of macroeconomic series that have a particularly close relationship with user search behaviour echoes work in the field of consumer action; for example, Goel et al. (2010) find that search data are predictive of specific consumer actions occurring in the near future, such as going to the cinema.

The weak link between search behaviour and CPI as well as consumer confidence is likely to be one of 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). Search queries genuinely related to our macroeconomic series may thus be swamped by many spurious correlations. Possibly for the same reason, both our in-sample and out-of-sample predictions of unemployment improved when using weekly rather than monthly data, even though (or perhaps because) fewer categories were available. For CPI and consumer confidence, these spurious correlations cannot effectively be filtered out and researchers trying to predict such variables may be better off by 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 in Germany, unlike in the other countries investigated, dropped steadily over the years following the financial crisis. Further research into more macroeconomic series in different regions could further test the robustness of our results.

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: 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, 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. Springer, New York.

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, 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. Handbook of Markov Chain Monte Carlo 2, 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.

Ročková, V., George, E. I., 2016. The Spike-and-Slab Lasso. Journal of the American Statistical Association (in press).

Scott, S. L., Varian, H. R., 2014a. Bayesian Variable Selection for Nowcasting Economic Time Series. In: Economic Analysis of the Digital Economy. University of Chicago Press, pp. 119–135.

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., 2018. Online Big Data-Driven Oil Consumption Forecasting with Google Trends. International Journal of Forecasting.

AppendixA. Data Transformations

We took log differences of the categories retrieved from Google Trends; the differenced series are economically, and statistically, more meaningful to interpret given the downward trend. Thereafter we removed the remaining structural components of the log-differenced 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 safely be removed 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 whose structural components do not necessarily appear 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 likely has multiplicative noise, as the magnitude of shocks is dependent on the level. As the transformed unemployment series still seemed to contain a trend and a seasonal component for our sample, thus 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*	✓	✓	✓	
CPI	US		✓		
	GE		✓		
	CA		✓		
	JA		✓		
	UK		✓		
CC	US		✓		
	GE		✓		
	CA		✓		
	JA		✓		
	UK		✓		

*UK unemployment data was only seasonally adjusted available.

Table A.5: Transformations of Google search data

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

AppendixB. State space matrices

A generic linear Gaussian state space model formulation is:

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

for $t = 1, ..., n$. 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 vector $\boldsymbol{\beta}$, assumed to be of appropriate dimensions, describe how the state $\boldsymbol{\alpha_t}$ and the regression 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 $\boldsymbol{Q}$. Below, we specify the system matrices $\boldsymbol{T}$, $\boldsymbol{R}$ and $\boldsymbol{Z}$ that are used to obtain the BSTS model:

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

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

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