

This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Economic Analysis of the Digital Economy

Volume Author/Editor: Avi Goldfarb, Shane M. Greenstein, and Catherine E. Tucker, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-20684-X; 978-0-226-20684-4

Volume URL: <http://www.nber.org/books/gree13-1>

Conference Date: June 6–7, 2013

Publication Date: April 2015

Chapter Title: Bayesian Variable Selection for Nowcasting Economic Time Series

Chapter Author(s): Steven L. Scott, Hal R. Varian

Chapter URL: <http://www.nber.org/chapters/c12995>

Chapter pages in book: (p. 119 – 135)

Bayesian Variable Selection for Nowcasting Economic Time Series

Steven L. Scott and Hal R. Varian

4.1 Introduction

Computers are now in the middle of many economic transactions. The details of these “computer-mediated transactions” can be captured in databases and used in subsequent analyses (Varian 2010). However, such databases can contain vast amounts of data, so it is normally necessary to do some sort of data reduction.

Our motivating example for this work is Google Trends, a system that produces an index of search activity on queries entered into Google. A related system, Google Correlate, produces an index of queries that are correlated with a time series entered by a user. There are many uses for these data, but in this chapter we focus on how to use the data to make short-run forecasts of economic metrics.

Choi and Varian (2009a, 2009b, 2011, 2012) described how to use search engine data to forecast contemporaneous values of macroeconomic indicators. This type of contemporaneous forecasting, or “nowcasting,” is of particular interest to central banks, and there have been several subsequent research studies from researchers at these institutions. See, for example, Arola and Galan (2012), McLaren and Shanbhoge (2011), Hellerstein and Middeldorp (2012), Suhoy (2009), and Carrière-Swallow and Labbé (2011). Choi and Varian (2012) contains several other references to work in this area. Wu and Brynjolfsson (chapter 3, this volume) describe an application of Google Trends data to the real estate market using cross-state data.

Steven L. Scott is a statistician at Google, Inc. Hal R. Varian is the chief economist at Google, Inc. and emeritus professor at the University of California at Berkeley.

For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see <http://www.nber.org/chapters/c12995.ack>.

In these studies, the researchers selected predictors using their judgment of relevance to the particular prediction problem. For example, it seems natural that search engine queries in the “vehicle shopping” category would be good candidates for forecasting automobile sales while queries such as “file for unemployment” would be useful in forecasting initial claims for unemployment benefits.

One difficulty with using human judgment is that it does not easily scale to models where the number of possible predictors exceeds the number of observations—the so-called “fat regression” problem. For example, the Google Trends service provides data for millions of search queries and hundreds of search categories extending back to January 1, 2004. Even if we restrict ourselves to using only category data, we have several hundred possible predictors and about 100 months of data. In this chapter we describe a scalable approach to time series prediction for fat regressions of this sort.

4.2 Approaches to Variable Selection

Castle, Qin, and Reed (2009) and Castle, Fawcett, and Hendry (2010) describe and compare twenty-one techniques for variable selection for time-series forecasting. These techniques fall into four major categories:

- Significance testing (forward and backward stepwise regression, Gets)
- Information criteria (AIC, BIC)
- Principle component and factor models (e.g., Stock and Watson 2010)
- Lasso, ridge regression, and other penalized regression models (e.g., Hastie, Tibshirani, and Friedman 2009)

Our approach combines three statistical methods into an integrated system we call Bayesian Structural Time Series, or BSTS for short. The three methods are:

- a “basic structural model” for trend and seasonality, estimated using Kalman filters;
- spike and slab regression for variable selection; and
- Bayesian model averaging over the best performing models for the final forecast.

We briefly review each of these methods and describe how they fit into our framework.

4.2.1 Structural Time Series and the Kalman Filter

Harvey (1991), Durbin and Koopman (2001), Petris, Petrone, and Campagnoli (2009) and many others have advocated the use of Kalman filters for time series forecasting. The “basic structural model” decomposes the time series into four components: a level, a local trend, seasonal effects, and an error term. The model described here drops the seasonal effect for simplic-

ity and adds a regression component; it is then called a “local linear trend model with regressors.”

This model is a stochastic generalization of the classic constant-trend regression model

$$y_t = \mu + bt + \beta x_t + e_t.$$

In this classic model the level (μ) and trend (b) parameters are constant, (x_t) is a vector of contemporaneous regressors, β is a vector of regression coefficients, and e_t is an error term.

In a local linear trend model each of these structural components is stochastic. In particular, the level and slope terms each follow a random walk model.

- (1) $y_t = \mu_t + z_t + v_t \quad v_t \sim N(0, V)$
- (2) $\mu_t = \mu_{t-1} + b_{t-1} + w_{1t} \quad w_{1t} \sim N(0, W_1)$
- (3) $b_t = b_{t-1} + w_{2t} \quad w_{2t} \sim N(0, W_2)$
- (4) $z_t = \beta x_t$

The unknown parameters to be estimated in this system are the variance terms (V, W_1, W_2) and the regression coefficients, β .

If we drop the trend and regression coefficients by setting $b_t = 0$ and $\beta = 0$, the “local trend model” becomes the “local level” model. When $V = 0$, the local level model is a random walk, so the best forecast of y_{t+1} is y_t . When $W_1 = 0$, the local level model is a constant mean model, so the best forecast of y_{t+1} is the average of all previously observed values of y_t . Hence, the local level model has two popular time series models as special cases.

It is easy to add a seasonal component to the local linear trend model, in which case it is referred to as the “basic structural model.” In the appendix we describe a general structural time series model that contains these and other models in the literature as special cases.

It is also possible to allow for time-varying regression coefficients by simply including them as another set of state variables. In practice, one would want to limit this to just a few coefficients, particularly when dealing with small sample sizes.

4.2.2 Spike-and-Slab Variable Selection

The spike-and-slab approach to model selection was developed by George and McCulloch (1997) and Madigan and Raftery (1994).

Let γ denote a vector of zeros and ones that indicates whether or not a particular regressor is included in a regression. More precisely, γ is a vector the same length as β , where $\gamma_i = 1$ indicates $\beta_i \neq 0$ and $\gamma_i = 0$ indicates $\beta_i = 0$. Let β_γ indicate the subset of β for which $\gamma_i = 1$, and let σ^2 be the residual variance from the regression model.

A spike-and-slab prior for the joint distribution of $(\beta, \gamma, \sigma^2)$ can be factored in the usual way.

$$(5) \quad p(\beta, \gamma, \sigma^2) = p(\beta_\gamma | \gamma, \sigma^2) p(\sigma^2 | \gamma) p(\gamma).$$

There are several ways to specify functional forms for these prior distributions. Here, we describe a particularly convenient choice.

The “spike” part of a spike-and-slab prior refers to the point mass at zero, for which we assume a Bernoulli distribution for each i , so that the prior is a product of Bernoulli’s:

$$(1) \quad \gamma \sim \prod_i \pi_i^{\gamma_i} (1 - \pi_i)^{1 - \gamma_i}.$$

When detailed prior information is unavailable, it is convenient to set all π_i equal to the same number, π . Alternatively, the researcher might expect that on average only k out of K possible coefficients are nonzero. In this case it is natural to set $\pi = k/K$. In our particular application, where we will use Google Trends categories as predictors, we may have prior beliefs about which categories are likely to be relevant.

More complex choices of $p(\gamma)$ can be made as well. For example, a non-Bernoulli model could be used to encode rules such as the hierarchical principle (no high order interactions without lower order interactions). The MCMC methods described below are robust to the specific choice of the prior.

The “slab” component is a prior for the values of the nonzero coefficients, conditional on knowledge of which coefficients are nonzero. Let b be a vector of prior guesses for regression coefficients, let Ω^{-1} be a prior precision matrix, and let Ω_γ^{-1} denote rows and columns of Ω^{-1} for which $\gamma_i = 1$. A conditionally conjugate “slab” prior is:

$$(7) \quad \beta_\gamma | \gamma, \sigma^2 \sim \mathcal{N}(b_\gamma, \sigma^2 (\Omega_\gamma^{-1})^{-1}), \quad \frac{1}{\sigma^2} \sim \Gamma\left(\frac{df}{2}, \frac{ss}{2}\right).$$

It is conventional to assume $b = 0$ (with the possible exception of the intercept term) and $\Omega^{-1} \propto \mathbf{X}^T \mathbf{X}$, in which case equation (7) is known as Zellner’s g -prior (Chipman et al. 2001). Because $\mathbf{X}^T \mathbf{X} / \sigma^2$ is the total Fisher information in the full data, it is reasonable to parametrize $\Omega^{-1} = \kappa (\mathbf{X}^T \mathbf{X}) / n$, the average information available from κ observations.

One issue with Zellner’s g -prior is that when the design matrix contains truly redundant predictors (as is the case when the number of possible predictors exceeds the number of observations), then $\mathbf{X}^T \mathbf{X}$ is rank deficient, which means that for some values of γ , $p(\beta, \sigma | \gamma)$ is improper. We can restore propriety by averaging $\mathbf{X}^T \mathbf{X}$ with its diagonal, so that

$$\Omega^{-1} = \frac{\kappa}{n} [w \mathbf{X}^T \mathbf{X} + (1 - w) \text{diag}(\mathbf{X}^T \mathbf{X})].$$

The final values that need to be chosen are df and ss . These can be elicited by asking the modeler for the R^2 statistic he expects to obtain from the regression, and the weight he would like to assign to that guess, measured in terms of the equivalent number of observations. The df parameter is the equivalent number of observations, and $ss = df(1 - R^2)s_y^2$.

Software implementing the spike-and-slab prior can make reasonable default choices for expected model size, κ , expected R^2 , and df , thereby giving the modeler the option either to accept the defaults, or to provide his own inputs.

4.2.3 Bayesian Model Averaging

Bayesian inference with spike-and-slab priors is an effective way to implement Bayesian model averaging over the space of time series regression models. As described below, we estimate the model by sampling from the posterior distribution of the parameters in the model. Each draw of parameters from the posterior can be combined with the available data to yield a forecast of y_{t+1} for that particular draw. Repeating these draws many times gives us an estimate of the posterior distribution of the forecast y_{t+1} .

This approach is motivated by the Madigan and Raftery (1994) proof that averaging over an ensemble of models does no worse than using the best single model in the ensemble. See Volinsky (2012) for links to tools and applications of Bayesian model averaging.

4.3 Estimating the Model

The Kalman filter, spike-and-slab regression, and model averaging all have natural Bayesian interpretations and tend to play well together. The basic parameters we need to estimate are γ (which variables are in the regression), β (the regression coefficients), and the variances of the error terms (V, W_1, W_2, W_3).

As the appendix describes in detail, we specify priors for each of these parameters and then sample from the posterior distribution using Markov Chain Monte Carlo (MCMC) techniques. There are a number of attractive short cuts available that make this sampling process quite efficient. These are described in more detail in the appendix and in a companion paper, Scott and Varian (2014).

These techniques yield a sample from the posterior distribution for the parameters that can then be used to construct a posterior distribution for forecasts of time series of interest.

4.4 Fun with Priors

We have already indicated that it is possible to use an informative prior to describe beliefs about the expected number of predictors. It is also possible

to use a prior in the regression to indicate likely relationships. For example, one might expect that automobile purchases are likely to be correlated with automotive-related queries.

A less obvious example involves using data-based priors for estimating the state and observation variances, (V, W_1, W_2, W_3) . Even though the Google Trends data only goes back to January 2004, economic time series are often much longer. One can estimate posterior distribution of the parameters in the univariate Kalman filter using the long series, then use this posterior distribution as the prior distribution for the shorter series where the Google Trends data are available.

4.5 Nowcasting Consumer Sentiment

To illustrate the use of BSTS for nowcasting, we examine the University of Michigan monthly survey of Consumer Sentiment from January 2004 to April 2012. We focus on “nowcasting” since we expect that queries at time t could be related to sentiment at time t but are not necessarily predictive of sentiment in the more distant future.

Our data from Google Trends starts at January 2004, and our sample ends in April 2012, giving us 100 observations. For predictors, we use 151 categories from Google Trends that have some connection with economics. These potential predictors were chosen from the roughly 300 query categories using the authors’ judgment.

Our problem is to find a good set of predictors for 100 observations chosen from a set of 151 possible predictors. This qualifies as a mildly obese, if not truly fat, regression.

The Consumer Sentiment index is not highly seasonal, but many of the potential predictors are seasonal, so we first deseasonalize the data by using the R command `stl`. We then detrend the predictors by regressing each predictor on a simple time trend. A visual inspection of the time series of the predictors indicated that these techniques were sufficient to “whiten” the data.

We then applied the BSTS estimation procedure described earlier. Figure 4.1 shows the inclusion probability for the top five predictors. A white bar indicates that the predictor has a positive relationship with consumer sentiment and a black bar indicates a negative relationship. The length of the bar measures the proportion of the estimated models in which that predictor was present.

The top predictor is financial planning, which is included in almost all of the models explored. The top queries in this category in the United States can be found on the Google Trends web page. They are “schwab,” “401k,” “charles schwab,” “ira,” “smith barney,” “fidelity 401k,” “john hancock,” “403b,” “401k withdrawl,” and “roth ira.”

The second most probable predictor is investing, which tends to have a

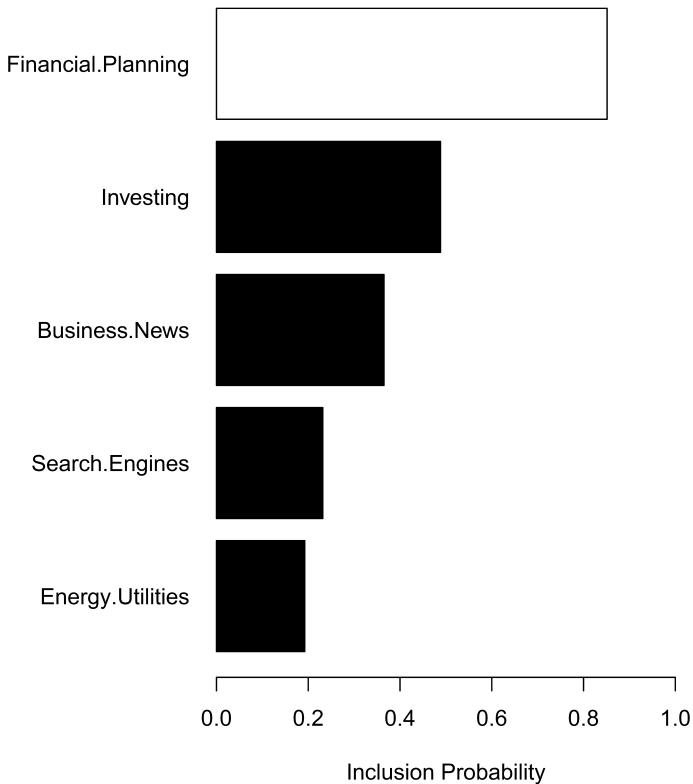


Fig. 4.1 Top five predictors for consumer sentiment

Note: Bars show the probability of inclusion. Shading indicates the sign of the coefficient.

negative relationship with confidence. The top queries in this category are “stock,” “gold,” “fidelity,” “stocks,” “stock market,” “silver,” “gold price,” “mutual,” “scottrade,” and “finance.”

The inclusion of the energy category is likely due to gasoline prices, which are known to have a negative relationship with consumer sentiment in the United States. We have no explanation for the search engine inclusion, though a visual inspection of the series shows that it does change direction at about the time the recession started. We speculate that the financial crisis influenced queries relating to economic conditions, which were classified as being related to both business news and search engines.

Figure 4.2 shows the posterior distribution of the one-step-ahead forecast along with the actual observations.

Note that the regression parameters are estimated using the entire sample of data, but the forecasts for period t are made using the value of consumer sentiment at $t - 1$ and the observed query categories at time t (for the included categories).

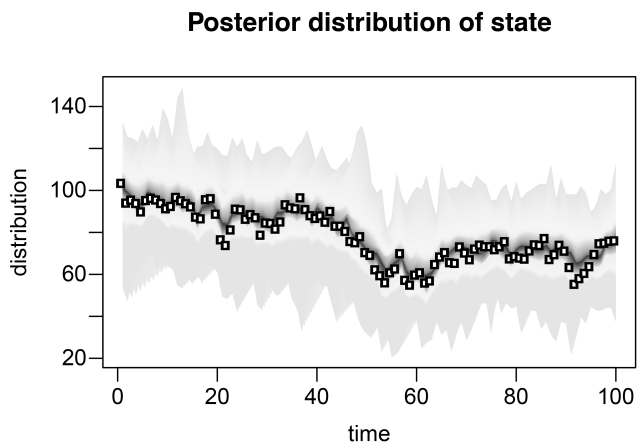


Fig. 4.2 Posterior distribution of forecast and the observations

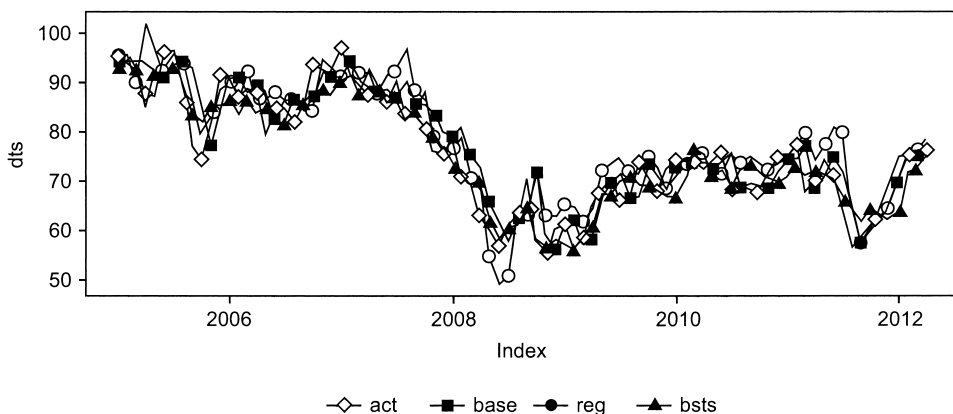


Fig. 4.3 Actual, base AR(1), regression, and BSTS one-step-ahead predictions

The model predicts reasonably well with a mean absolute one-step-ahead prediction error of about 4.5 percent. A naive AR(1) model has a mean absolute one-step-ahead prediction error of 5.2 percent, indicating an improvement of about 14 percent. See figure 4.3 for a time series plot of the actual, AR(1), and BSTS one-step-ahead predictions.

As we have seen, the BSTS system can decompose the forecast into the trend and regression components. The trend component is basically the univariate Kalman filter forecast, while the regression component uses the predictors from the query categories. Figure 4.4 illustrates the contribution of each state variable and regressor to the fit. The faint line in each panel is the previous fit.

Incremental Plot for Fitting y_1

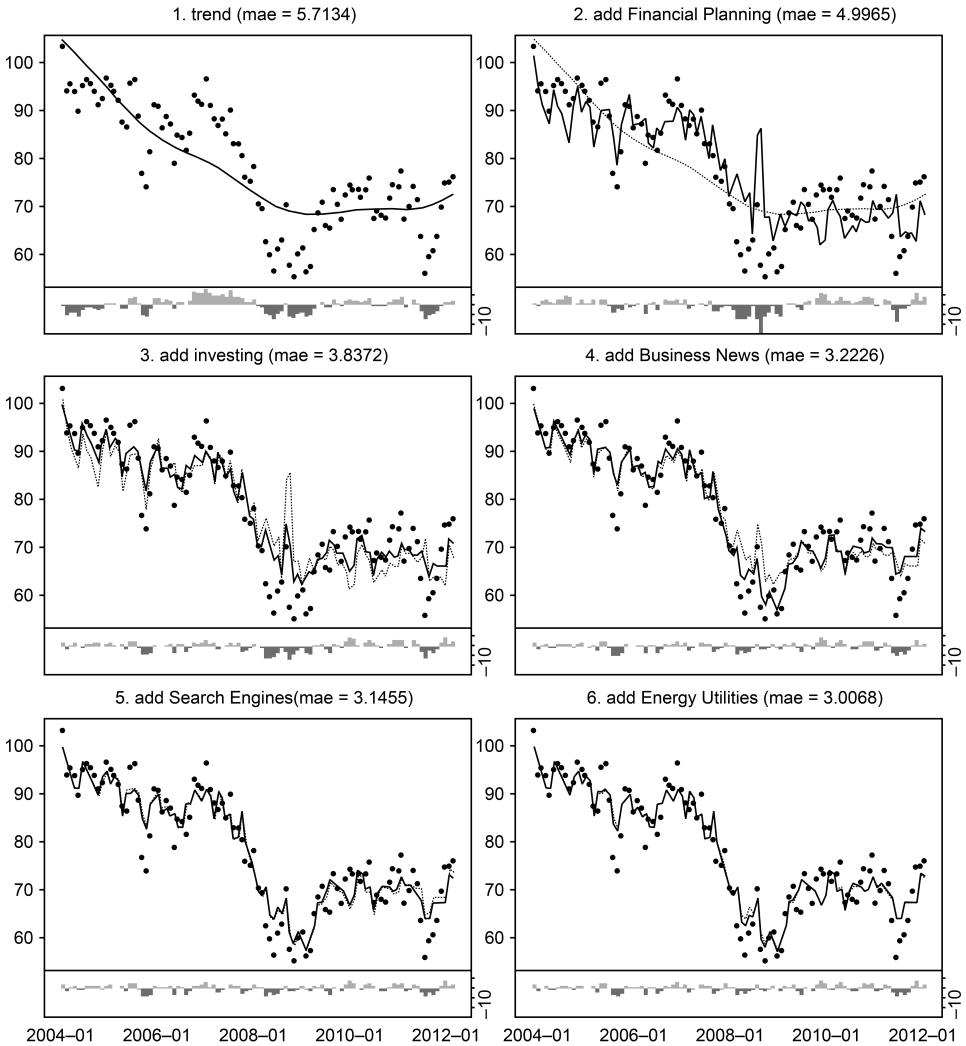


Fig. 4.4 Decomposition of forecast for consumer sentiment using Google Trends data

Note: Variables are ordered by probability of inclusion, mean absolute error is given in title, and residuals are shown at the bottom of each panel.

4.6 Nowcasting Gun Sales

The National Instant Criminal Background Check (NICS) is a service offered by the FBI to federal firearms licensees that can quickly determine whether a prospective buyer is eligible to buy firearms or explosives. A monthly report on the number of checks conducted is available on the Web.¹

We downloaded the NICS data and fed it to Google Correlate, which produced 100 queries that were highly correlated with this series. The first ten were “stack on,” “bread,” “44 mag,” “buckeye outdoors,” “mossberg,” “g star,” “ruger 44,” “baking,” “.308,” and “savage 22.” Most of these queries are related to weapons; the exceptions (bread and baking) have to do with the fact that hunting season starts at about the same time as Thanksgiving in many states.

We used BSTS to find the best predictors from this set for the NICS background check data. Since the data was highly seasonal, we used both a local linear trend and seasonal state variables. The best predictor by far was “gun stores” which, interestingly, only ranked 36th on the list of correlates. The in-sample MAE of the simple model using only trend + seasonal was 0.34, but adding “gun stores” cut the MAE to 0.15, a substantial reduction. Figure 4.5 shows how adding trend, seasonal, and query data improves the in-sample fit.

We also ran BSTS using all 585 verticals produced by Google Trends to fit the 107 observations of monthly NICS data. The two most probable predictors are shown in table 4.1. The category “Recreation::Outdoors::Hunting:and:Shooting” is by far the most probable predictor. The forecast decomposition is shown in figure 4.6, which indicates a substantial contribution by the regression component.

4.7 Summary

We have described a Bayesian approach to variable selection for time series that combines Kalman filtering, spike-and-slab regressions, and model averaging. Although the system was developed for nowcasting using Google Trends data, there are many other possible applications.

For example, Brodersen et al. (2013) describe how to use BSTS to estimate ad effectiveness. The basic idea is to build a BSTS model that predicts website visits using trend, seasonal, and regression components. When an ad campaign is initiated, this model can be used to predict the counterfactual—what would have happened in the absence of the campaign. The difference between the actual and the counterfactual is the causal effect of the ad campaign.

We have focused on nowcasting since, in most cases, the action taken by

1. http://www.fbi.gov/about-us/cjis/nics/reports/080112_1998_2012_Monthly_Yearly_Totals.pdf.

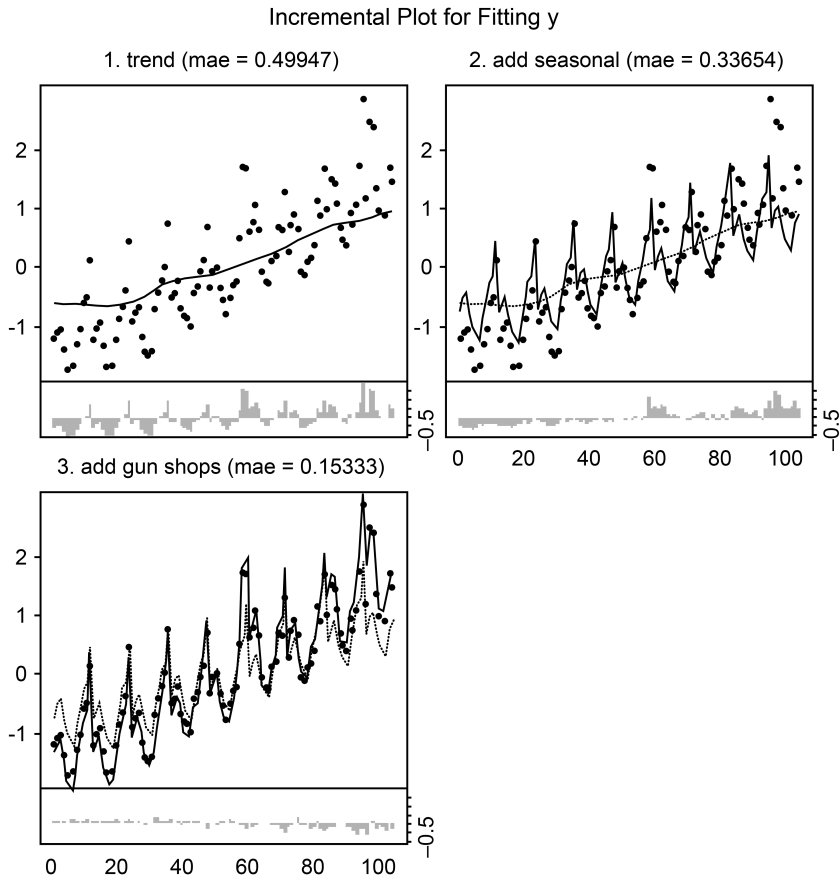


Fig. 4.5 Decomposition of forecast for NICS using correlate data

Note: Variables are ordered by probability of inclusion, mean absolute error is given in the title, and residuals are shown at the bottom of each panel.

Table 4.1 Google Trends predictors for NICS checks

Category	Mean	Inc. prob.
Recreation::outdoors::hunting:and:shooting	1,056,208	0.97
Travel::adventure:travel	-84,467	0.09

individuals is contemporaneous with the related queries. But in some cases, such as vacation planning or housing purchases, the relevant queries may precede the actions by several months. In such cases queries may help in longer-term forecasting. (See, e.g., Choi and Liu 2011.)

As more and more data becomes available the problem of “fat regressions” will arise in many other contexts and we anticipate there will be considerable interest in model selection. Given the widespread availability of “big data” it

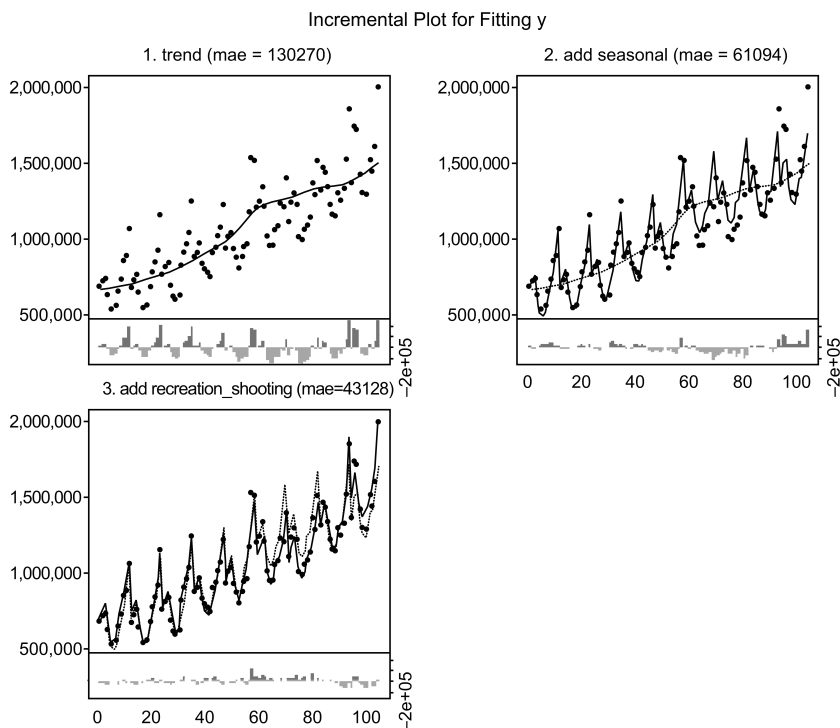


Fig. 4.6 Decomposition of forecast for NICS using Google Trends data

Note: Variables are ordered by probability of inclusion, mean absolute error is given in the title, and residuals are shown at the bottom of each panel.

seems strange that so much attention is paid to sampling uncertainty when the real issue in most cases is model uncertainty. We believe that Bayesian methods such as those we have described allow for better ways to describe model uncertainty. (See Varian [2014] for further discussion.)

It is widely recognized that averaging many small models tends to give better out-of-sample forecasting performance than using a single complex model. Bayesian methods give a principled way to perform such averaging which should, in turn, lead to better forecasts.

Appendix

Structural Time Series Models

Here we describe our Bayesian Structural Time Series model. More detail can be found in Scott and Varian (2014). We focus on structural time series models of the standard form

$$(8) \quad \begin{aligned} y_t &= Z_t^T \alpha_t + \epsilon_t & \epsilon_t &\sim \mathcal{N}(0, H_t) \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t & \eta_t &\sim \mathcal{N}(0, Q_t). \end{aligned}$$

Here y_t is time series to be modeled and the vector α_t is a latent variable indicating the state of the model; it contains any trend, seasonal, or other components deemed necessary by the modeler.

Z_t is a vector of coefficients applied to the state variables, ϵ_t is a normally distributed error term with mean zero, and H_t is its variance. Each state component contributes to the block diagonal transition matrix T_t , the rectangular block diagonal residual matrix R_t , and the observation vector Z_t . The error term η_t has covariance matrix Q_t .

The model matrices (Z, T, R, H, Q) can be used to construct the Kalman filter, which can then be used to forecast future values $y_{t+\tau}$ from current observations (y_1, \dots, y_t) . One attractive feature of the Kalman filter is that it has a natural Bayesian interpretation and can easily be combined with the variable selection and model averaging techniques we have chosen.

Regression

Regressors can be included in a structural time series model in either a static framework (where the regression coefficients are fixed) or dynamic framework (where the regression coefficients can change over time).

In a dynamic regression the coefficients are a component of the state vector, which evolve over time according to some stochastic process. In a static regression, by contrast, the coefficients are fixed, unknown parameters. A convenient way to include a static regression component in the model is to set $\alpha_t = 1, t_t = 1, q_t = 0$, and $z_t = \beta'x_t$. This specification adds $\beta'x_t$ to the contributions of the other state components in a computationally efficient way because it only adds one additional state to the model. A small dimension is helpful because the Kalman recursions are quadratic in the dimension of the state space.

Estimating the Model Using Markov Chain Monte Carlo

We estimate the posterior distribution of the model parameters using Markov Chain Monte Carlo. Let θ denote the collection of model parameters (β, σ, ψ) where ψ is the collection of all model parameters associated with state components other than the static regression. Then the complete data posterior distribution is:

$$(9) \quad p(\theta, \alpha | y) \propto p(\theta) p(\alpha_0) \prod_{t=1}^n p(y_t | \alpha_t, \theta) p(\alpha_t | \alpha_{t-1}, \theta).$$

In order to sample from the posterior distribution we use an efficient Gibbs sampling algorithm that alternates between draws of $p(\alpha | \theta, y)$ and $p(\theta | \alpha, y)$, which produces a sequence $(\theta, \alpha)_0, (\theta, \alpha)_1, \dots$ from a Markov chain with stationary distribution $p(\theta | \alpha, y)$.

The key point is that, conditional on α , the time series and regression

components of the model are independent. Thus the draw from $p(\theta|\alpha, \mathbf{y})$ decomposes into several independent draws from the different conditional posterior distributions of the state components. In particular, $p(\psi, \beta, \sigma^{-2}|\alpha, \mathbf{y}) = p(\psi|\alpha, \mathbf{y})p(\beta, \sigma^{-2}|\alpha, \mathbf{y})$.

Sampling α

The idea of using Kalman filtering to sample the state in a linear Gaussian structural time series model was independently proposed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994). Various improvements to the early algorithms have been made by de Jong and Shepard (1995), Rue (2001), and others. We use the method proposed by Durbin and Koopman (2002), who observed that the variance of $p(\alpha|\theta, \mathbf{y})$ does not depend on the numerical value of \mathbf{y} . Durbin and Koopman (2001) describe a fast smoothing method for computing $E(\alpha|\mathbf{y}, \theta)$ using the Kalman filter.

Thus one may simulate a fake data set $(\mathbf{y}^*, \alpha^*) \sim p(\mathbf{y}, \alpha|\theta)$ by simply iterating equation (8). Then the fast mean smoother can be used to subtract the conditional mean $E(\alpha^*|\theta, \mathbf{y}^*)$ from α^* , which is now mean zero with the correct variance. A second fast smoother can be used to add in $E(\alpha|\mathbf{y}, \theta)$, yielding a draw of α with the correct moments. Because $p(\alpha|\mathbf{y}, \theta)$ is Gaussian, the correct moments imply the correct distribution.

Sampling θ

Many of the usual models for state components are simple random walks, whose variance parameters are trivial to sample conditional on α . For example, consider the state variables for the local linear trend model described in equation (4):

$$\mu_{t+1} = \mu_t + \delta_t + \eta_{0t}$$

$$\delta_{t+1} = \delta_t + \eta_{1t}$$

where η_0 and η_1 are independent Gaussian error terms with variances ψ_0^2 and ψ_1^2 . With independent Gamma priors on $\psi_0^{-2} \sim \Gamma(df_0/2, SS_0/2)$ and $\psi_1^{-2} \sim \Gamma(df_1/2, SS_1/2)$, their full conditional is the product of two independent Gamma distributions

$$p(\psi_0^{-2}, \psi_1^{-2}|\alpha) = \Gamma\left(\frac{df_0 + n - 1}{2}, \frac{SS_0}{2}\right)\Gamma\left(\frac{df_1 + n - 1}{2}, \frac{SS_1}{2}\right),$$

where $SS_0 = ss_0 + \sum_{t=2}^n (\mu_t - \mu_{t-1} - \delta_{t-1})^2$ and $SS_1 = ss_1 + \sum_{t=2}^n (\delta_t - \delta_{t-1})^2$. These complete data sufficient statistics are observed given α , so drawing ψ_0^{-2} and ψ_1^{-2} from their full conditional distribution is trivial. Most of the traditional state models can be handled similarly, including the seasonal component of the BSM and dynamic regression coefficients.

The full conditional for (β, σ^{-2}) is likewise independent of the other state components, with $\tilde{y}_t = y_t - Z_t^T \alpha_t + \beta^T x_t \sim \mathcal{N}(\beta^T x_t, \sigma^2)$. Thus, by subtracting

the contributions from the other state components from each y_t we are left with a standard spike-and-slab regression. The posterior distribution can be simulated efficiently by drawing from $p(\gamma|\alpha, \mathbf{y})$ using a sequence of Gibbs sampling steps, and then drawing from the well-known closed form $p(\beta_\gamma, \sigma^{-2}|\gamma, \alpha, \mathbf{y})$. This technique is known as “stochastic search variable selection” (George and McCulloch 1997). There have been many suggested improvements to the SSVS algorithm (notably Ghosh and Clyde 2011), but we have obtained satisfactory results with the basic algorithm.

The conditional posteriors for β_γ and σ^{-2} can be found in standard texts (e.g., Gelman et al. 2002). They are:

$$(10) \quad p(\beta|\mathbf{y}, \alpha, \gamma, \sigma^{-2}) = \mathcal{N}(\tilde{\beta}_\gamma, \sigma^2 V_\gamma), \text{ and } p(\sigma^{-2}|\mathbf{y}, \alpha, \gamma) = \Gamma\left(\frac{df+n}{2}, \frac{ss+\tilde{S}}{2}\right)$$

where the complete data sufficient statistics are $V_\gamma^{-1} = \mathbf{X}^T \mathbf{X}_\gamma + \Omega_\gamma^{-1}$, $\tilde{\beta}_\gamma = V_\gamma^{-1}(\mathbf{X}^T \tilde{\mathbf{y}}_\gamma + \Omega_\gamma^{-1} b_\gamma)$, and $\tilde{S} = \sum_{t=1}^n (\tilde{y}_t - \mathbf{x}_t^T \tilde{\beta}_\gamma)^2 + (\tilde{\beta}_\gamma - b_\gamma)^T \Omega_\gamma^{-1} (\tilde{\beta}_\gamma - b_\gamma)$. The distribution for $p(\gamma|\alpha, \mathbf{y})$ can be shown to be

$$(11) \quad p(\gamma|\mathbf{y}, \alpha) \propto \frac{|\Omega_\gamma^{-1}|^{-U/2}}{|V_\gamma^{-1}|^{-U/2}} \tilde{S}^{-(df+n)/2}.$$

Let $|\gamma|$ denote the number of included components. Under Zellner’s g -prior it is easy to see that

$$\frac{|\Omega_\gamma^{-1}|}{|V_\gamma|} = \left(\frac{\kappa/n}{1 + \kappa/n}\right)^{|\gamma|}$$

is decreasing in $|\gamma|$. It is true in general that $|\Omega_\gamma^{-1}| \leq |\Omega_\gamma^{-1} + \mathbf{X}^T \mathbf{X}_\gamma|$, which implies that $p(\gamma|\mathbf{y}, \alpha)$ prefers models with few predictors and small residual variation.

Equation (11) can be used in a Gibbs sampling algorithm that draws each γ_t given γ_{-t} (the elements of γ other than γ_t). Each full conditional distribution is proportional to equation (11), and γ_t can only assume two possible values. Notice that $p(\gamma|\mathbf{y}, \alpha)$ only requires matrix computations for those variables that are actually included in the model. Thus if the model is sparse, the Gibbs sampler involves many inexpensive decompositions of small matrices, which makes SSVS computationally tractable even for problems with a relatively large number of predictors.

References

Arola, Concha, and Enrique Galan. 2012. “Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches.” Technical Report, Bank of Spain. <http://www.bde.es/webbde/SES/Secciones/Publicaciones/Publicaciones/Seriadas/DocumentosOcasionales/12/Fich/do1203e.pdf>.

- Brodersen, Kay, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. 2013. "Inferring Causal Impact Using Bayesian Structural Time-series Models." Technical Report, Google, Inc. <http://research.google.com/pubs/pub41854.html>.
- Carrière-Swallow, Yan, and Felipe Labbé. 2011. "Nowcasting with Google Trends in an Emerging Market." *Journal of Forecasting* 32 (4): 289–98. doi: 10.1002/for.1252.
- Carter, Chris K., and Robert Kohn. 1994. "On Gibbs Sampling for State Space Models." *Biometrika* 81 (3): 541–53.
- Castle, Jennifer L., Nicholas W. P. Fawcett, and David F. Hendry. 2010. "Evaluating Automatic Model Selection." Technical Report no. 474, Department of Economics, University of Oxford. <http://economics.ouls.ox.ac.uk/14734/1/paper474.pdf>.
- Castle, Jennifer L., Xiaochuan Qin, and W. Robert Reed. 2009. "How to Pick the Best Regression Equation: A Review and Comparison of Model Selection Algorithms." Technical Report no. 13/2009, Department of Economics, University of Canterbury. <http://www.econ.canterbury.ac.nz/RePEc/cbt/econwp/0913.pdf>.
- Chipman, Hugh, Edward I. George, Robert E. McCulloch, Merlise Clyde, Dean P. Foster, and Rober A. Stine. 2001. "The Practical Implementation of Bayesian Model Selection." In *Lecture Notes-Monograph Series*, vol. 38, 65–134. Beachwood, OH: Institute of Mathematical Statistics.
- Choi, Hyunyoung, and Paul Liu. 2011. "Reading Tea Leaves in the Tourism Industry: A Case Study in the Gulf Oil Spill." Technical Report, Google, Inc. http://www.google.com/url?q=http%3A%2F%2Fwww.google.com%2Fgoogleblogs%2Fpdfs%2Fgoogle_gulf_tourism_march2011.pdf.
- Choi, Hyunyoung, and Hal Varian. 2009a. "Predicting Initial Claims for Unemployment Insurance Using Google Trends." Technical report, Google, Inc. <http://research.google.com/archive/papers/initialclaimsUS.pdf>.
- . 2009b. "Predicting the Present with Google Trends." Technical report, Google, Inc. http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf.
- . 2011. "Using Search Engine Data for Nowcasting—An Illustration." In *Actes des Rencontres Économiques*, pages 535–38. Rencontres Économiques d'Aix-en-Provence, Le Cercle des économistes. Aix-en-Provence, France. http://www.lecercledeconomistes.asso.fr/IMG/pdf/Actes_Rencontres_Economiques_d_Aix-en-Provence_2011.pdf.
- . 2012. "Predicting the Present with Google Trends." *Economic Record* 88, 2–9. <http://onlinelibrary.wiley.com/doi/10.1111/j.1475-4932.2012.00809.x/pdf>.
- de Jong, Piet, and Neil Shepard. 1995. "The Simulation Smoother for Time Series Models." *Biometrika*, 82 (2): 339–50.
- Durbin, James, and Siem Jan Koopman. 2001. *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- . 2002. "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis." *Biometrika* 89 (3): 603–16.
- Frühwirth-Schnatter, Sylvia. 1994. "Data Augmentation and Dynamic Linear Models." *Journal of Time Series Analysis* 15 (2): 183–202.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2002. *Bayesian Data Analysis*, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC.
- George, Edward I., and Robert E. McCulloch. 1997. "Approaches for Bayesian Variable Selection." *Statistica Sinica* 7: 339–73. <http://www3.stat.sinica.edu.tw/statistica/oldpdf/A7n26.pdf>.
- Ghosh, Joyee, and Merlise A. Clyde. 2011. "Rao-Blackwellization for Bayesian Variable Selection and Model Averaging in Linear and Binary Regression: A Novel Data Augmentation Approach." *Journal of the American Statistical Association* 106 (495): 1041–52.

- Harvey, Andrew. 1991. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*, 2nd ed. New York: Springer Science+Business Media.
- Hellerstein, Rebecca, and Menno Middeldorp. 2012. "Forecasting with Internet Search Data." Liberty Street Economics Blog of the Federal Reserve Bank of New York, January 4. <http://libertystreeteconomics.newyorkfed.org/2012/01/forecasting-with-internet-search-data.html>.
- Madigan, David M., and Adrian E. Raftery. 1994. "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window." *Journal of the American Statistical Association* 89:1335–46.
- McLaren, Nick, and Rachana Shanbhoge. 2011. "Using Internet Search Data as Economic Indicators." Bank of England Quarterly Bulletin, June. <http://www.bankofengland.co.uk/publications/quarterlybulletin/qb110206.pdf>.
- Petris, Giovanni, Sonia Petrone, and Patrizia Campagnoli. 2009. *Dynamic Linear Models with R*. New York: Springer Science+Business Media.
- Rue, Håvard. 2001. "Fast Sampling of Gaussian Markov Random Fields." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (2): 325–38.
- Scott, Steven L., and Hal R. Varian. 2014. "Predicting the Present with Bayesian Structural Time Series." *International Journal of Mathematical Modelling and Numerical Optimisation* 5 (1/2): 4–23.
- Stock, James, and Mark Watson. 2010. "Dynamic Factor Models." In *Oxford Handbook of Economic Forecasting*, edited by M. Clements and D. Hendry. Oxford: Oxford University Press.
- Suhoy, Tanya. 2009. "Query Indices and a 2008 Downturn: Israeli Data." Technical Report, Bank of Israel. <http://www.bankisrael.gov.il/deptdata/mehkar/papers/dp0906e.pdf>.
- Varian, Hal R. 2010. "Computer Mediated Transactions." *American Economic Review Papers & Proceedings* 100 (2): 1–10.
- . 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28 (2): 3–28.
- Volinsky, Chris. 2012. "Bayesian Model Averaging Home Page." Technical Report, Bell Labs. <http://www2.research.att.com/~volinsky/bma.html>.