THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 19-08

Gary Koopy
University of Strathclyde, Glasgow, UK
and
The Rimini Centre for Economic Analysis, Italy

Markus Jochmann
University of Strathclyde, Glasgow, UK
and
The Rimini Centre for Economic Analysis, Italy

Rodney W. Strachan
University of Queensland, UK
and
The Rimini Centre for Economic Analysis, Italy

"BAYESIAN FORECASTING USING STOCHASTIC SEARCH VARIABLE SELECTION IN A VAR SUBJECT TO BREAKS"

Copyright belongs to the author. Small sections of the text, not exceeding three paragraphs, can be used provided proper acknowledgement is given.

The *Rimini Centre for Economic Analysis* (RCEA) was established in March 2007. RCEA is a private, non-profit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *Small Open Economies in the Globalized World* (SOEGW). Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers series.

The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Rimini Centre for Economic Analysis.

The Rimini Centre for Economic Analysis
Legal address: Via Angherà, 22 – Head office: Via Patara, 3 - 47900 Rimini (RN) – Italy
www.rcfea.org - secretary@rcfea.org

Bayesian Forecasting using Stochastic Search Variable Selection in a VAR Subject to Breaks*

Markus Jochmann
University of Strathclyde

Gary Koop†
University of Strathclyde

Rodney W. Strachan
University of Queensland

June 2008

Abstract

This paper builds a model which has two extensions over a standard VAR. The first of these is stochastic search variable selection, which is an automatic model selection device which allows for coefficients in a possibly over-parameterized VAR to be set to zero. The second allows for an unknown number of structual breaks in the VAR parameters. We investigate the in-sample and forecasting performance of our model in an application involving a commonly-used US macroeconomic data set. We find that, in-sample, these extensions clearly are warranted. In a recursive forecasting exercise, we find moderate improvements over a standard VAR, although most of these improvements are due to the use of stochastic search variable selection rather than the inclusion of breaks.

*All authors are Fellows at the Rimini Centre for Economic Analysis.

†Corresponding author: Gary Koop, Department of Economics, University of Strathclyde, Glasgow, G4 0GE, U.K., email: Gary.Koop@strath.ac.uk

1 Introduction

Since the so-called Minnesota revolution associated with the work of Chris Sims and colleagues at the University of Minnesota [see Doan, Litterman and Sims (1984), Litterman (1986) and Sims (1980)], Bayesian vector autoregressions (VARs) have become a popular and successful forecasting tool [see, among a myriad of others, Andersson and Karlsson (2007) and Kadiyala and Karlsson (1993)]. Geweke and Whiteman (2006) provide an excellent survey of the development of Bayesian VAR forecasting.

Despite the success of VARs relative to many alternatives, two problems continue to undermine their forecast performance. These are the problems of structural breaks and over-parameterization. The purpose of this paper is to investigate whether extending a VAR to allow for structural breaks and using stochastic search variable selection will help overcome these problems.

With regards to the problem of over-parameterization, VARs have so many parameters that over-fitting is a serious risk. Good in-sample model fit may not necessarily lead to good forecasting performance. Furthermore, in a typical VAR, many coefficients are close to zero and/or are imprecisely estimated. This can lead to a great deal of forecast uncertainty (e.g. large predictive standard deviations). As a result, most Bayesian VARs involve informative priors (e.g. the Minnesota prior) which allow for shrinkage. Indeed a wide variety of Bayesian and non-Bayesian approaches have emphasized the importance of shrinkage to improve forecast performance. Geweke and Whiteman (2006) provide a general discussion of this issue and empirical examples include Diebold and Pauly (1990) and Koop and Potter (2004).

In this paper, we investigate an alternative way of treating the over-parameterization problem: stochastic search variable selection (SSVS). SSVS can be thought of as a hierarchical prior where each of the parameters in the VAR is drawn from one of two Normal distributions. The first of these has a zero mean and a variance very near to zero (i.e. the parameter is virtually zero and the corresponding variable is effectively excluded from the model). The second of these is a relatively noninformative prior (i.e. the parameter is non-zero and the corresponding variable is included in the model). Such "spike and slab" priors have become popular for model selection or model averaging in regression models [see, e.g., Ishwaran and Rao (2005)]. George, Sun and Ni (2008) develop a particular way of implementing SSVS in VAR models and Korobilis (2008) presents an application in a factor model. In this paper we use a VAR with SSVS (extended to allow for structural breaks as

described below) in a recursive forecasting exercise. Note that this approach can be thought of as a way of doing shrinkage (i.e. some coefficients are shrunk virtually to zero, while others are left relatively unconstrained and estimated in a data based fashion). In our recursive forecasting exercise it also allows us to see if the set of variables which are useful predictors changes over time.

With regards to the second problem, several recent papers have highlighted the fact that structural instability seems to be present in a wide variety of macroeconomic and financial time series [e.g. Ang and Bekaert (2002) and Stock and Watson (1996)]. The negative consequences of ignoring this instability for inference and forecasting has been stressed by, among many others, Clements and Hendry (1998, 1999), Koop and Potter (2001, 2007) and Pesaran, Pettenuzzo and Timmerman (2006). If a structural break occurs, then pre-break data can potentially contaminate forecasts. In this paper, we adopt an approach similar to Chib (1998) to modelling the break process. This approach assumes that there is a constant probability of a break in every time period and implies that our forecasts use only data since the most recent break. Furthermore, as outlined in Koop and Potter (2007) and Pesaran, Pettenuzzo and Timmerman (2006), such an approach allows for the possibility that a structural break might hit during the forecast period.

After building our model, which extends the VAR to allow for structural breaks and uses the SSVS hierarchical prior, we use it in a recursive forecasting exercise using US data. In particular, we work with three variables: the unemployment rate, the interest rate and the inflation rate. The original data runs from 1953Q1 through 2006Q3. We compare our full model (the VAR with SSVS plus breaks) to two restricted versions: a standard VAR with SSVS and a standard VAR without SSVS. We present in-sample results which indicate that the full model outperforms both restricted variants. We then present a battery of evidence from a recursive forecasting exercise. This evidence shows that SSVS offers an appreciable improvement in forecast performance. Evidence that the inclusion of structural breaks improves forecast performance in this data set is weaker.

2 The VAR with SSVS Prior and Structural Breaks

The models used in this paper all begin with an unrestricted VAR which we write as:

$$y_{t+h} = X_t\alpha + \varepsilon_t, \tag{1}$$

where y_{t+h} is an $n \times 1$ vector of observations on the dependent variables at time $t+h$, α a $K \times 1$ vector of VAR coefficients, ε_t are independent $N(0, \Sigma)$ errors for $t = 1, .., T$. X_t is the $n \times K$ matrix containing the dependent variables at time t through time $t - p + 1$ (i.e. X_t contains information available at time t) and an intercept [arranged appropriately to define a VAR(p)]. The forecast horizon is h.

To this foundation, we add the SSVS prior and allow for structural breaks. In this section, we describe how this is done, with full details of our Bayesian econometric methods, including the Markov chain Monte Carlo algorithm, described in the appendix.

2.1 Stochastic Search Variable Selection

SSVS is usually motivated as a method for selecting variables in a regression model. However, it can also be interpreted as a hierarchical prior for (1). In this section, we will describe the basic ideas of SSVS as it relates to the VAR coefficients, $\alpha = (\alpha_1, .., \alpha_K)'$. However, we also apply SSVS to the off-diagonal elements of Σ, allowing for each of these to be shrunk virtually to zero (we do not allow for the diagonal elements of Σ to be shrunk to zero since then it would be singular). Full details are given in the appendix.

The SSVS prior for each VAR coefficient is a mixture of two Normals:

$$\alpha_j|\gamma_j \sim \left(1 - \gamma_j\right) N\left(0, \kappa_{0j}^2\right) + \gamma_j N\left(0, \kappa_{1j}^2\right), \tag{2}$$

where γ_j is a dummy variable which equals 1 if α_j is drawn from the first Normal and equals 0 if it is drawn from the second. The prior is hierarchical since $\gamma = (\gamma_1, .., \gamma_K)'$ is treated as an unknown vector of parameters and estimated in a data-based fashion. The SSVS aspect of this prior arises by choosing the first prior variance, κ_{0j}^2, to be "small" (so that the coefficient is virtually zero) and the second prior variance, κ_{1j}^2, to be "large" (implying a relatively noninformative prior for the corresponding coefficient). Just how

"small" and "large" prior variances can be chosen is discussed in, e.g., George and McCulloch (1993, 1997). Basically, the small prior variance is chosen to be small enough so that the corresponding explanatory variable is, to all intents and purposes, deleted from the model. In this paper, we use what George, Sun and Ni (2008) call a "default semi-automatic approach" which requires no subjective prior information from the researcher (see the appendix for details).[1]

To understand how SSVS can be used when forecasting, note that γ_j can be interpreted as an indicator for whether the j^{th} VAR coefficient is in the model or not. SSVS methods allow, at each point in time in a recursive forecasting exercise, for the calculation of $\Pr\left(\gamma_j = 0|Data\right)$. SSVS can either be used as a model selection device (e.g. we can choose to forecast using the restricted VAR which includes only the coefficients for which $\Pr\left(\gamma_j = 0|Data\right) > \frac{1}{2}$), or it can be used to do model averaging. That is, our MCMC algorithm provides us with draws of γ_j for $j = 1, .., K$. Each draw implies a particular restricted VAR which we can use for forecasting. By averaging over all MCMC draws we are doing Bayesian model averaging. But we stress that $\Pr\left(\gamma_j = 0|Data\right)$ can be different at different points in time in the recursive forecasting exercise and, thus, the weights attached to different models are changing over time.

2.2 Allowing For Structural Breaks

Given the finding of widespread structural instability in many macroeconomic time series models, it is important to allow for structural breaks (in both the conditional mean and the conditional variance) of the VAR. Accordingly, we extend the VAR with the SSVS prior to allow for breaks to occur. Our approach will allow for a constant probability of a break in each time period. With such an approach, the number of breaks, $M-1$, is unknown (and estimated in the model). This leads to a model with structural breaks at times $\tau_1, .., \tau_{M-1}$ (and, thus, M regimes exist). Thus, (1) is replaced by:

[1] Previously, we have referred to SSVS as allowing for coefficients to be "virtually zero" or explanatory variables being "to all intents and purposes zero" since κ_{0j}^2 is not precisely zero. In the remainder of this paper, we will omit such qualifying words and phrases like "virtually" or "to all intents and purposes".

$$
\begin{aligned}
&y_{t+h} = X_t\alpha^{(1)} + \varepsilon_t,\ \varepsilon_t \sim N(0, \Sigma^{(1)}) \text{ for } t = 1, .., \tau_1 \\
&y_{t+h} = X_t\alpha^{(2)} + \varepsilon_t,\ \varepsilon_t \sim N(0, \Sigma^{(2)}) \text{ for } t = \tau_1 + 1, .., \tau_2 \\
&. \\
&. \\
&y_{t+h} = X_t\alpha^{(M)} + \varepsilon_t,\ \varepsilon_t \sim N(0, \Sigma^{(M)}) \text{ for } t > \tau_{M-1}.
\end{aligned} \tag{3}
$$

There are many models which could be used for the break process [see, among many others, Chib (1998), Kim, Nelson and Piger (2004), Koop and Potter (2007, 2008), Maheu and Gordon (2007), Maheu and McCurdy (2007), McCulloch and Tsay (1993), Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo and Timmerman (2006)]. In this paper, we adopt a specification closely related to that used in Chib (1998). It assumes that there is constant probability of a structural break occurring, q, in every time period. This is a simple, but attractive choice, since (unlike many other approaches) it does not impose a fixed number of structural breaks on the model.[2] Furthermore, it allows us to predict the probability that a structural break occurs during our forecast period. Thus, it helps address one of the major concerns of macroeconomic forecasters: how to forecast when structural breaks might be occurring. In practice, our method will involve (with probability $1 - q$) forecasting using the VAR model which holds at time τ (the period the forecast is being made) and (with probability q) forecasting assuming a break has occurred. Similar in spirit to Maheu and McCurdy (2007) and Maheu and Gordon (2007), we assume that if a break occurs then past data provides no help with forecasting and, accordingly, forecasts are made using only prior information.[3]

Complete details are provided in the appendix.

3 Empirical Results

In this section, we present results working with a standard set of variables. In particular, our US data set runs from 1953Q1 through 2006Q3 and contains

[2] Formally, our algorithm requires the choice of a maximum number of breaks (with the actual number of breaks occuring in-sample being estimated from the data). We choose this maximum to be large enough so as not to reasonably constrain the number of breaks.

[3] In contrast, Koop and Potter (2007) and Pesaran, Pettenuzzo and Timmerman (2006) develop different models where, even after a structural break occurs, past data has some predictive power.Which alternative is preferable depends on the nature of the break process in the empirical problem under study.

the unemployment rate (seasonally adjusted civilian unemployment rate, all workers over age 16), interest rate (yield on three month Treasury bill rate) and inflation rate (the annual percentage change in a chain-weighted GDP price index).[4] This set of variables (possibly transformed) has been used by, among many others, Cogley and Sargent (2005), Koop, Leon-Gonzalez and Strachan (2007) and Primiceri (2005). Figure 1 shows graphs of the three variables.

[4]The data were obtained from the Federal Reserve Bank of St. Louis website, http://research.stlouisfed.org/fred2/.

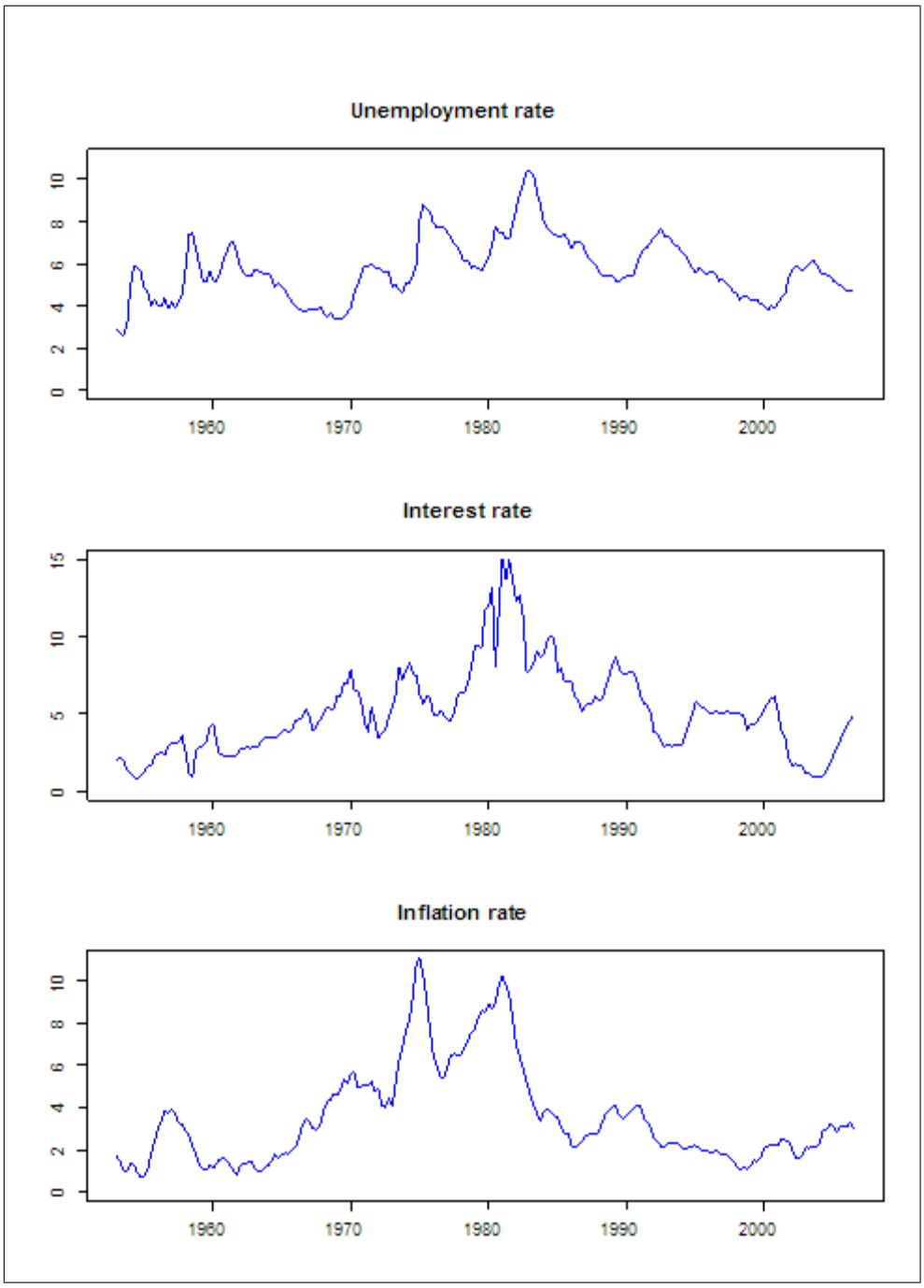

Figure 1: The Data

We allow for 4 lags in our VAR. This large lag length choice is motivated by our use of SSVS. That is, even if the true lag length is less than 4, the use of SSVS means that coefficients on longer lags can be set to zero. We use forecast horizons $h = 1, 4$ and 8. Note that y_{t+h} is our dependent variable and we repeat our entire modeling exercise three times, once for each forecasting horizon. Our in-sample results are for the $h = 1$ case.

Note also that we are working with VARs as opposed to models which allow for cointegration. All of our variables are rates so unit root issues are likely of little importance. Ignoring unit root and cointegration issues is common practice in Bayesian macroeconomic studies. For instance, most of the work associated with Christopher Sims works with VARs [see, e.g., the citations at the beginning of the introduction or, more recently, Sims and Zha (2006)] treating cointegration as a particular parametric restriction that may or may not hold and likely to have little relevance for forecasting at short to medium horizons. Similar considerations presumably motivate recent influential recent empirical macroeconomic papers such as Cogley and Sargent (2001, 2005) and Primiceri (2005) who work with VARs.

Our "default semi-automatic approach" to SSVS (see appendix) specifies a prior for the VAR coefficients. The prior for the break probability we use is given in the appendix.

We divide our discussion of empirical results into two sub-sections. In the first of these, we briefly present in-sample estimation results using the full sample. In the second, we present results for our recursive forecasting exercise.

We present empirical results for three models: the VAR with SSVS and breaks, the VAR with SSVS but no breaks and the VAR. The second model is obtained by restricting $q = 0$ and the third model additionally imposes $\gamma_j = 1$ for all parameters. The prior for any restricted model is identical to the prior in the unrestricted model, conditional on the restriction being imposed.

3.1 In-Sample Results

Figure 2 plots the probability that a break occurs in each time period. There is evidence of two breaks, one in the mid to late 1960's and one in the mid 1980s. That is, Figure 2 implies that the cumulative probability that a break occurs sometime in 1964-1970 is virtually one and the same holds for the 1983-1987 period. The BICs also indicate support for the model with

breaks. The BICs for the VAR, the VAR with SSVS but no breaks and the VAR with SSVS and breaks are 506.426, 388.685 and 374.904, respectively, indicating strong support for the model with breaks.[5]

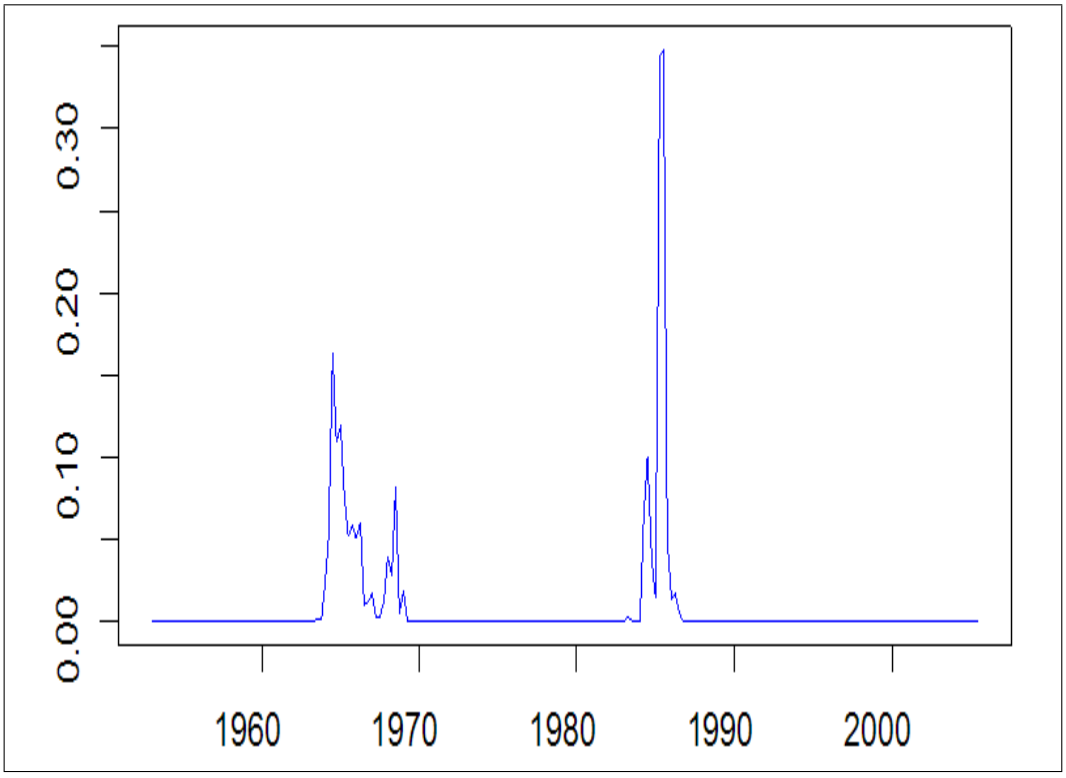

Figure 2: Posterior Probability of a Break Occurring

Appendix B contains additional empirical results using the full sample. In particular, it contains the posterior probability of inclusion of each parameter (i.e. the probability that $\gamma_j = 1$) as well as posterior means and standard

[5]In models with hierarchical priors such as ours, issues arise with how you count the number of parameters in the penalty for complexity term in BIC (see, e.g., Carlin and Louis, 2000, pages 220-223). To avoid these complications, in models with SSVS priors we count all coefficients which are included in the model with at least 16% probability.

deviations for every coefficient in every one of the three regimes in the VAR with SSVS and breaks. The interested reader can look through these tables in detail. Here we note two points. Firstly, it is the case that the SSVS prior is excluding (with high probability) many of the VAR coefficients in each regime. This indicates that it is an effective way of ensuring parsimony in the over-parameterized VAR (although, interestingly, SSVS is not deleting off-diagonal elements of the error covariance matrices). Secondly, in the model with breaks, although there is some evidence that some of the VAR coefficients differ across regimes, most of the change is coming in the error covariance matrices. That is, change in the error covariance matrix is causing the breaks, not change in the VAR coefficients.

3.2 Forecasting

In this section, we present results from a recursive forecasting exercise. The recursive exercise will involve using data through time τ to forecast $\tau + h$ for $h = 1, 4$ and 8. This will be carried out for $\tau = \tau_0, .., T - h$ where τ_0 is 1974Q4.

Our end goal is the predictive density using information through time τ. We use notation where $y^*_{\tau+h}$ is the random variable we are wishing to forecast (and $y_{\tau+h}$ is the actual realization of $y^*_{\tau+h}$). Thus, $p\left(y^*_{\tau+h}|X_\tau, .., X_1\right)$ is the predictive density of interest. The properties of this density can be obtained using our MCMC algorithm. We let θ denote all the parameters of the VAR with breaks and θ_m be the VAR parameters in the m^{th} regime and $\theta_m^{(r)}$ for $r = 1, .., R$ denote MCMC replications (subsequent to the burn-in replications). Then, if no structural break occurs between $\tau + 1$ and $\tau + h$, standard MCMC theory and the structure of our model imply:

$$p\left(y^*_{\tau+h}|X_\tau, .., X_1\right) = \frac{1}{R}\sum_{r=1}^{R} p\left(y^*_{\tau+h}|X_\tau, .., X_1, \theta_m^{(r)}\right) \tag{4}$$

as $R \to \infty$. Note that $p\left(y^*_{\tau+h}|X_\tau, .., X_1, \theta_m^{(r)}\right)$ has a standard analytical form:

$$p\left(y^*_{\tau+h}|X_\tau, .., X_1, \theta_m^{(r)}\right) = f_N(y^*_{\tau+h}|X_\tau\alpha_m^{(r)}, \Sigma_m^{(r)}), \tag{5}$$

where f_N denotes the normal density.

When we allow for structural breaks to occur with probability q, a slight extension is required. At each draw of our MCMC algorithm, we have the

VAR coefficients and error covariance matrix in the last regime (i.e. using data since the last structural break) and a draw of q. With probability $1-q$, the predictive density is given by (5). With probability q a break has occurred and the predictive is based on the SSVS prior (see the Technical Appendix for details).

Once we have $p\left(y_{\tau+h}^{*}|X_{\tau},..,X_{1}\right)$ we have to choose some methods for evaluating forecast performance. We use mean squared forecast error, mean absolute forecast error, predictive likelihoods and hit rates for this purpose. The first two of these are based on point forecasts and are defined as:

$$MSFE=\frac{\sum_{\tau=\tau_{0}}^{T-h}\left[y_{\tau+h}-E\left(y_{\tau+h}|X_{\tau},..,X_{1}\right)\right]^{2}}{T-h-\tau_{0}+1}$$

and

$$MAFE=\frac{\sum_{\tau=\tau_{0}}^{T-h}\left|y_{\tau+h}-E\left(y_{\tau+h}|X_{\tau},..,X_{1}\right)\right|}{T-h-\tau_{0}+1}.$$

Mean squared forecast error can be decomposed into the variance of forecast errors plus their bias squared. To aid in interpretation, we also present this bias squared (with the variance term being MSFE minus this).

Predictive likelihoods are motivated and described in many places such as Geweke and Amisano (2007). The basic predictive likelihood is the predictive density for $y_{\tau+h}^{*}$ evaluated at the actual outcome $y_{\tau+h}$. We use the sum of log predictive likelihoods for forecast evaluation:

$$\sum_{\tau=\tau_{0}}^{T-h}\log\left[p\left(y_{\tau+h}^{*}=y_{\tau+h}|X_{\tau},..,X_{1}\right)\right]$$

where each term in the summation can be approximated using MCMC output in a method similar to (4).

Table 1 presents MSFE and bias squared, Table 2 presents MAFE and Table 3 presents the forecast performance metric based on predictive likelihoods for our three models, three variables and three forecast horizons. They tell a similar story. Adding SSVS to the benchmark VAR typically improves forecast performance, but adding breaks does not. We elaborate on these points next.

Tables 1 and 2 indicate that (with the exception of inflation at longer horizons), the SSVS prior is helping to improve forecast performance. Relative to results found in other forecasting studies, a reduction in the square

root of MSFE of 10% indicates a substantial improvement in forecast performance. For the unemployment rate, SSVS is leading to improvements of this magnitude (particularly at short and medium horizons). For the interest rate, there are also appreciable improvements in forecast performance. Even for the inflation rate (a variable which is notoriously difficult to forecast), at short horizons, including SSVS does lead to some forecast improvement.

However, the inclusion of breaks does not improve forecast performance in this data set. In fact, with one exception, it leads to a deterioration of forecast performance. This sort of finding is common in forecasting studies (see, e.g., Dacco and Satchell, 1999). Models with high dimensional parameter spaces which fit well in-sample, often do not forecast well. Even use of the SSVS prior, which is intended to help overcome the problems caused by over-parameterization, does not fully overcome this problem in this data set.

The results relating to forecast bias in Table 1 shed more light on the poor forecasting performance of the model with breaks. It can be seen that all of our models are producing forecasts which are on average approximately unbiased. It is clearly the variance component that is driving the MSFE. This is consistent with the idea that the predictive density for the SSVS-with-breaks model is too dispersed. After a break occurs, our model begins estimating a new VAR using only data since the break. The advantages and disadvantages of such a strategy have been discussed in the literature. A particularly lucid exposition is provided in Pastor and Stambaugh (2001) in an empirical exercise involving the equity premium:

> In standard approaches to models that admit structural breaks, estimates of current parameters rely on data only since the most recent estimated break. Discarding the earlier data reduces the risk of contaminating an estimate ... with data generated under a different [process]. That practice seems prudent, but it contends with the reality that shorter histories typically yield less precise estimates. Suppose a shift in the equity premium occurred a month ago. Discarding virtually all of the historical data on equity returns would certainly remove the risk of contamination by pre-break data, but it hardly seems sensible in estimating the current equity premium. Completely discarding the pre-break data is appropriate only when the premium might have shifted to such a degree that the pre-break data are no more useful ..., than, say, pre-break rainfall data, but such a view almost surely ignores

economics [Pastor and Stambaugh, 2001, pages 1207-1208].

Our approach, by discarding all pre-break data, reduces the risk that our forecasts are contaminated by data from a different data generating process, but our resulting predictives are less precisely estimated. At least in the present application, this latter factor is clearly causing problems. Note that there do exist other Bayesian approaches which allow for pre-break data to play at least some role in post-break estimation (thus potentially leading to more precise forecasts). Examples include Pesaran, Pettenuzzo and Timmerman (2006) and Koop and Potter (2007).

Table 1: Forecast Evaluation Using Mean Squared Forecast Error

Forecast horizon	Dep. variable	Model		
		VAR	SSVS	SSVS + breaks
		MSFE		
1	unemp	0.104	0.086	0.113
	interest	1.005	0.975	1.077
	inflation	0.112	0.106	0.126
4	unemp	0.620	0.484	1.343
	interest	4.197	3.745	7.483
	inflation	1.777	1.799	2.196
8	unemp	1.469	1.331	1.248
	interest	8.232	8.152	12.984
	inflation	3.286	4.410	5.286
		Bias squared		
1	unemp	0.00063	0.00003	0.02984
	interest	0.00019	0.00006	0.01659
	inflation	0.00364	0.00434	0.00040
4	unemp	0.00001	0.00001	0.00035
	interest	0.01350	0.01044	0.00152
	inflation	0.15163	0.19365	0.65499
8	unemp	0.00278	0.01953	0.01365
	interest	0.01592	0.08554	0.58556
	inflation	0.26425	0.64445	1.14304

Table 2: Forecast Evaluation Using Mean Absolute Forecast Error				
Forecast horizon	Dep. variable	Model		
		VAR	SSVS	SSVS + breaks
1	unemp	0.210	0.200	0.255
	interest	0.608	0.606	0.657
	inflation	0.248	0.240	0.251
4	unemp	0.633	0.577	0.843
	interest	1.509	1.446	1.945
	inflation	0.898	0.911	1.179
8	unemp	0.905	0.883	0.924
	interest	2.166	2.215	2.850
	inflation	1.274	1.485	2.096

The poor forecast performance of the model with breaks exhibited in Tables 1 and 2, could partly be due to the fact that they are based on point forecasts.[6] As we saw in the preceding section, most of the evidence for breaks seems to arise due to breaks in the conditional variance (i.e. the error covariances are changing over time), not the conditional mean (i.e. the VAR coefficients are exhibiting little change over time). Loosely speaking, point forecasts largely reflect the conditional mean, not the conditional variance. Hence, the breaks in conditional variance are not contaminating our forecasts from the VARs without breaks. However, the log predictive likelihoods incorporate all aspects of the predictive distribution and, hence, Table 3 indicates that this argument does not tell the whole story. That is, even using predictive likelihoods, we are finding the model which allows for breaks does not forecast well.

As a final exercise, to see how well our predictive densities are fitting in the tails of the distribution, we calculate hit rates for a rare but important event. This event is that the unemployment and inflation rate both rise. A "Hit" is defined as occurring if the event occurs and the point forecast also lies in the region defined by the even. The "Hit Rate" is the proportion of Hits. Table 4 presents the Hit Rates for our various models and forecast horizons. These Hit Rates are all reasonably high, being in excess of 0.70. At very short horizons, the VAR using SSVS and breaks does have an appreciably higher hit rate, although this finding does not occur at longer horizons.

[6]Boxplots of the point forecasts are given in Appendix B.

Table 3: Log Predictive Likelihoods			
Forecast horizon	Model		
	VAR	SSVS	SSVS + breaks
1	-1.457	-1.250	-1.572
4	-4.658	-4.406	-7.205
8	-7.395	-7.320	-10.540

Table 4: Hit Rates			
Forecast horizon	Model		
	VAR	SSVS	SSVS + breaks
1	0.729	0.725	0.754
4	0.743	0.736	0.769
8	0.765	0.769	0.720

4 Conclusions

In this paper, we develop a model which has two major extensions over a standard VAR. The first of these is SSVS and the second is breaks in all the model parameters. The motivation for the first is that VARs have a large number of parameters (with many of them being insignificant in empirical applications). This can lead to over-parameterization problems: over-fitting in-sample along with imprecise inference. The hope is that SSVS will mitigate these problems. The motivation for the second is that structural breaks are often observed to occur in macroeconomic data sets and a forecasting procedure which ignores them could be seriously misleading.

We investigate the in-sample and forecasting performance of our model in an application involving a standard trivariate VAR with three popular macroeconomic variables. Our findings are mixed, but are moderately encouraging. In-sample, the VAR with SSVS and breaks clearly performs better than either a VAR with SSVS or a standard VAR. In our recursive forecasting exercise, the use of SSVS in the VAR does bring substantial improvements relative to a standard VAR. However, there is little evidence that the further addition of breaks improves forecasting performance.

References

Andersson, M. and Karlsson, S. (2007). "Bayesian forecast combination for VAR models," forthcoming in S. Chib, W. Griffiths, G. Koop and D. Terrell (eds.), Advances in Econometrics, Volume 23: Bayesian Econometrics, (Elsevier: Amsterdam). Manuscript available at http://www.bus.lsu.edu/hill/aie/karlsson.pdf.

Ang, A. and Bekaert, G. (2002). "Regime switches in interest rates," *Journal of Business and Economic Statistics*, 20, 163-182.

Carlin, B. and Louis, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, second edition. (Chapman and Hall: Boca Raton).

Chib, S. (1998). "Estimation and comparison of multiple change-point models," *Journal of Econometrics,* 86, 221-241.

Clements, M. and Hendry, D. (1998). *Forecasting Economic Time Series.* (Cambridge University Press: Cambridge).

Clements, M. and Hendry, D. (1999). *Forecasting Non-stationary Economic Time Series.* (The MIT Press: Cambridge).

Cogley, T. and Sargent, T. (2001). "Evolving post-World War II inflation dynamics," *NBER Macroeconomic Annual*, 16, 331-373.

Cogley, T. and Sargent, T. (2005). "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S," *Review of Economic Dynamics*, 8, 262-302.

Dacco, R. and Satchell, S. (1999). "Why do regime-switching models forecast so poorly?" *International Journal of Forecasting*, 18, 1-16.

Diebold, F. and Pauly, P. (1990). "The use of prior information in forecast combination," *International Journal of Forecasting*, 6, 503-508.

Doan, T., Litterman, R. and Sims, C. (1984). "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 1-100.

George, E. and McCulloch, R. (1993). "Variable selection via Gibbs sampling," *Journal of the American Statistical Association,* 85, 398-409.

George, E. and McCulloch, R. (1997). "Approaches for Bayesian variable selection," *Statistica Sinica,* 7, 339-373.

George, E., Sun, D. and Ni, S. (2008). "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142, 553-580.

Gerlach, R., Carter, C and Kohn, R. (2000). "Efficient Bayesian inference in dynamic mixture models," *Journal of the American Statistical Association*, 95, 819-828.

Geweke, J. and Amisano, J. (2007). "Hierarchical Markov normal mixture models with applications to financial asset returns," manuscript available at

http://www.biz.uiowa.edu/faculty/jgeweke/papers/paperA/paper.pdf.

Geweke, J. and Whiteman, C. (2006). "Bayesian forecasting," Chapter 1 in G. Elliott, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*. (Elsevier: Amsterdam).

Giordani, P. and Kohn, R. (2006). "Efficient Bayesian inference for multiple change-point and mixture innovation models," *Journal of Business and Economic Statistics*, forthcoming.

Ishwaran, H. and Rao, J.S. (2005). "Spike and slab variable selection: Frequentist and Bayesian strategies," *Annals of Statistics*, 33, 730-773.

Kadiyala, K. and Karlsson, S. (1993). "Forecasting with generalized Bayesian vector autoregressions," *Journal of Forecasting*, 12, 365-378.

Kim, C.-J., Nelson, C. and Piger, J. (2004). "The less volatile U.S. economy: A Bayesian investigation of timing, breadth, and potential explanations," *Journal of Business and Economic Statistics*, 22, 80-93.

Koop, G., Leon-Gonzalez, R. and Strachan, R. (2007). "On the evolution of monetary policy," manuscript available at

http://personal.strath.ac.uk/gary.koop/koop_leongonzalez_strachan_kls5.pdf.

Koop, G. and Potter, S. (2001). "Are apparent findings of nonlinearity due to structural instability in economic time series?," *The Econometrics Journal* 4, 37-55.

Koop, G. and Potter, S. (2004). "Forecasting in dynamic factor models using Bayesian model averaging," *The Econometrics Journal*, 7, 550-565.

Koop, G. and Potter, S. (2007). "Estimation and forecasting in models with multiple breaks," *Review of Economic Studies*, 74, 763-789.

Koop, G. and Potter, S. (2008). "Prior elicitation in multiple change-point models," *International Economic Review*, forthcoming, manuscript available at http://personal.strath.ac.uk/gary.koop/kp12rev.pdf

Korobilis, D. (2008). "Bayesian estimation, model selection and forecasting in multiple change-point models," forthcoming in *Advances in Econometrics*, volume 23 (Bayesian Econometric Methods), edited by S. Chib, B. Griffiths, G. Koop and D. Terrell (Elsevier Science: Amsterdam).

Litterman, R. (1986). "Forecasting with Bayesian vector autoregressions - 5 years of experience," *Journal of Business and Economic Statistics*, 4, 25-38.

Maheu, J. and Gordon, S. (2007). "Learning, forecasting and structural breaks," *Journal of Applied Econometrics*, forthcoming.

Maheu, J. and McCurdy, T. (2007). "How useful are historical data for forecasting the long-run equity return distribution," *Journal of Business and*

Economic Statistics, forthcoming.

McCulloch, R and Tsay, R. (1993). "Bayesian inference and prediction for mean and variance shifts in autoregressive time series," *Journal of the American Statistical Association,* 88, 968-978.

Pastor, L. and Stambaugh, R. (2001). "The equity premium and structural breaks," *Journal of Finance*, 56, 1207-1239.

Pesaran, M.H., Pettenuzzo, D. and Timmerman, A. (2006). "Forecasting time series subject to multiple structural breaks," *Review of Economic Studies*, 73, 1057-1084.

Primiceri. G. (2005). "Time varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, 72, 821-852.

Sims, C. (1980). "Macroeconomics and reality," *Econometrica*, 48, 1-80.

Sims, C. and Zha, T. (2006). "Were there regime switches in macroeconomic policy?" *American Economic Review*, 96, 54-81.

Stock, J. and Watson, M. (1996). "Evidence on structural instability in microeconomic time series relations," *Journal of Business and Economic Statistics*, 14, 11-30.

Appendix A: Technical Details

The following are the models and associated Markov chain Monte Carlo algorithms used in the empirical section. We begin by establishing the notation relating to the data. y_t is an $n \times 1$ vector containing data on n dependent variables and x_t is a $k \times 1$ vector containing the accompanying explanatory variables. For the VAR(p) with an intercept, $x_t = \left(1, y'_{t-1}, .., y_{t-p}\right)'$ and $k = 1 + pn$.

The VAR with SSVS Prior

The Likelihood Function

The VAR model can be written in matrix form as:

$$Y = XA + \varepsilon, \tag{A.1}$$

where Y is a $T \times n$ matrix with t^{th} row given by y'_{t+h}, X is a $T \times k$ matrix with t^{th} row given by x'_t, A is a $k \times n$ matrix of coefficients and ε is a $T \times n$ matrix with t^{th} row given by ε'_t. The likelihood function is defined by assuming ε_t (an $n \times 1$ vector of errors) to be i.i.d. $N(0, \Sigma)$.

The Prior

SSVS can be interpreted as defining a hierarchical prior for all of the elements of A and Σ. Each element in the hierarchy is a mixture of two normals, one with a small variance (implying the coefficient is not in the model) and one with a large variance (implying the coefficient is included in the model). We adopt a particular way of applying SSVS based on George, Sun and Ni (2008) and much of the material in this part of the appendix is based on their paper.

With regards to the VAR coefficients, let $\alpha = vec(A)$ with elements α_j for $j = 1, .., kn$. The prior for α has the form:

$$\alpha | \gamma \sim N(0, DD), \tag{A.2}$$

where γ is a $kn \times 1$ vector of unknown parameters with typical element $\gamma_j \in \{0, 1\}$, and D is a diagonal matrix with $(j, j)^{th}$ element given by d_j where

$$d_j = \begin{cases} \kappa_{0j} \text{ if } \gamma_j = 0 \\ \kappa_{1j} \text{ if } \gamma_j = 1 \end{cases}. \tag{A.3}$$

Note that George, Sun and Ni (2008) extend (A.2) so that the prior covariance matrix is DRD where R is a correlation matrix which can be appropriately chosen to allow for prior correlations between the VAR coefficients.

In this paper, we assume $R = I$ and, thus, the VAR coefficients are *a priori* independent of one another.

Note that this prior implies a mixture of two Normals:

$$\alpha_j|\gamma_j \sim \left(1-\gamma_j\right) N\left(0, \kappa_{0j}^2\right) + \gamma_j N\left(0, \kappa_{1j}^2\right). \tag{A.4}$$

We use what George, Sun and Ni (2008) call the "default semi-automatic approach" to selecting the prior hyperparameters κ_{0j} and κ_{1j} and the reader is referred to this paper for additional justification for this approach. Basically, κ_{0j} should be selected so that α_j is essentially zero and κ_{1j} should be selected so that α_j is empirically substantive. The default semi-automatic approach involves $\kappa_{0j} = c_0\sqrt{\widehat{var\left(\alpha_j\right)}}$ and $\kappa_{1j} = c_1\sqrt{\widehat{var\left(\alpha_j\right)}}$ where $\widehat{var\left(\alpha_j\right)}$ is an estimate of the variance of the coefficient in an unrestricted VAR (e.g. the ordinary least squares quantity or an estimate based on a preliminary MCMC run of the VAR using a non-informative prior). The pre-selected constants c_0 and c_1 must have $c_0 << c_1$ and we set $c_0 = \frac{1}{10}$ and $c_1 = 10$.

For γ, the SSVS prior posits that each element has a Bernoulli form (independent of the other elements of γ) and, hence, for $j = 1, .., kn$, we have

$$\begin{array}{l} \Pr\left(\gamma_j = 1\right) = \underline{q}_j \\ \Pr\left(\gamma_j = 0\right) = 1 - \underline{q}_j \end{array}. \tag{A.5}$$

We set $\underline{q}_j = \frac{1}{2}$ for all j. This is a natural default choice, implying each coefficient is *a priori* equally likely to be included as excluded.

We can decompose the error covariance matrix as:

$$\Sigma^{-1} = \Psi\Psi', \tag{A.6}$$

where Ψ is upper-triangular. The SSVS prior involves using a standard Gamma prior for square of each of the diagonal elements of Ψ and the SSVS mixture of normals prior for each element above the diagonal. Note that this implies that the diagonal elements of Ψ are always included in the model, ensuring a positive definite error covariance matrix. This form for the prior also greatly simplifies posterior computation. Precise details are provided in the next paragraph.

Let the non-zero elements of Ψ be labelled as ψ_{ij} and define $\psi = \left(\psi_{11}, .., \psi_{nn}\right)'$, $\eta_j = \left(\psi_{1j}, .., \psi_{j-1,j}\right)'$ and $\eta = \left(\eta_2', .., \eta_n'\right)'$. For the diagonal elements, we assume prior independence with:

$$\psi_{jj}^2 \sim G\left(\underline{a}_j, \underline{b}_j\right), \tag{A.7}$$

where $G\left(\underline{a}_j, \underline{b}_j\right)$ denotes the Gamma distribution with mean $\frac{\underline{a}_j}{\underline{b}_j}$ and variance $\frac{\underline{a}_j}{\underline{b}_j^2}$. We specify $\underline{a}_j = 2.2$ and $\underline{b}_j = 0.24$.

The hierarchical prior for η takes the same mixture of normals form as α. In particular, the SSVS prior has

$$\eta_j|\omega_j \sim N\left(0, F_j F_j\right), \tag{A.8}$$

where $\omega_j = \left(\omega_{1j}, .., \omega_{j-1,j}\right)'$ is a vector of unknown parameters with typical element $\omega_{ij} \in \{0, 1\}$, and $F_j = diag\left(f_{1j}, .., f_{j-1,j}\right)$ where

$$f_{ij} = \begin{cases} \kappa_{0ij} \text{ if } \omega_{ij} = 0 \\ \kappa_{1ij} \text{ if } \omega_{ij} = 1 \end{cases}, \tag{A.9}$$

for $j = 2, .., n$ and $i = 1, .., j-1$.

Note that this prior implies a mixture of two Normals for each off-diagonal element of Ψ:

$$\psi_{ij}|\omega_{ij} \sim (1 - \omega_{ij}) N\left(0, \kappa_{0ij}^2\right) + \omega_{ij} N\left(0, \kappa_{1ij}^2\right). \tag{A.10}$$

As we did with κ_{0j} and κ_{1j} (the prior hyperparameters in the SSVS for the VAR coefficients), we use a semi-automatic default approach to selecting κ_{0ij} and κ_{1ij}.That is, we set $\kappa_{0ij} = c_0\sqrt{\widehat{var\left(\psi_{ij}\right)}}$ and $\kappa_{1ij} = c_1\sqrt{\widehat{var\left(\psi_{ij}\right)}}$ where $\widehat{var\left(\psi_{ij}\right)}$ is an estimate of the variance of the appropriate off-diagonal element of Σ from an unrestricted VAR (e.g. the ordinary least squares quantity or the posterior variance from a preliminary run of the unrestricted VAR using a noninformative prior). The pre-selected constants c_0 and c_1 are set (as before) to be $c_0 = \frac{1}{10}$ and $c_1 = 10$.

For $\omega = \left(\omega_2', .., \omega_n'\right)'$, the SSVS prior posits that each element has a Bernoulli form (independent of the other elements of ω) and, hence, we have

$$\begin{array}{l} \Pr\left(\omega_{ji} = 1\right) = \underline{q}_{ji} \\ \Pr\left(\omega_{ji} = 0\right) = 1 - \underline{q}_{ji} \end{array}. \tag{A.11}$$

We make the default choice of $\underline{q}_{ji} = \frac{1}{2}$ for all j and i.

This completes our discussion of the SSVS hierarchical prior for the VAR. Note that we also work with an unrestricted VAR. To ensure comparability

of priors across the two models, our unrestricted VAR is a limiting case of the VAR with SSVS prior. That is, our unrestricted VAR is as above except that $\gamma_j = \omega_{ji} = 1$ for all i and j.

Posterior Computation

Posterior computation can be carried out using the MCMC algorithm described in George, Sun and Ni (2008). This algorithm involves the following posterior conditional distributions. Note that, to keep the notation as simple as possible, we are suppressing the conditioning arguments, but stress that these are the full conditional posteriors necessary to describe a valid MCMC algorithm.

For the VAR coefficients we have

$$\alpha \sim N\left(\overline{\alpha}, \overline{V}_\alpha\right), \tag{A.12}$$

where

$$\overline{V}_\alpha = \left\{(\Psi\Psi') \otimes (X'X) + (DD)^{-1}\right\}^{-1},$$

$$\overline{\alpha} = \overline{V}_\alpha \left[\left\{(\Psi\Psi') \otimes (X'X)\right\} \widehat{\alpha}\right],$$

$$\widehat{\alpha} = vec\left(\widehat{A}\right)$$

and

$$\widehat{A} = (X'X)^{-1} X'Y.$$

The conditional posterior for γ has γ_j being independent Bernoulli random variables:

$$\begin{array}{l} \Pr\left(\gamma_j = 1\right) = \overline{q}_j \\ \Pr\left(\gamma_j = 0\right) = 1 - \overline{q}_j \end{array}, \tag{A.13}$$

where

$$\overline{q}_j = \frac{\frac{1}{\kappa_{1j}} \exp\left(-\frac{\alpha_j^2}{2\kappa_{1j}^2}\right) \underline{q}_j}{\frac{1}{\kappa_{1j}} \exp\left(-\frac{\alpha_j^2}{2\kappa_{1j}^2}\right) \underline{q}_j + \frac{1}{\kappa_{0j}} \exp\left(-\frac{\alpha_j^2}{2\kappa_{0j}^2}\right) \left(1 - \underline{q}_j\right)}.$$

The conditional posterior for ψ can be obtained by noting that the conditional posterior for ψ_{jj}^2 (for $j = 1, .., n$) are independent of one another with

$$\psi_{jj}^2 \sim G\left(a_j + \frac{T}{2}, \bar{b}_j\right), \tag{A.14}$$

where

$$\bar{b}_j = \begin{cases} b_1 + \dfrac{v_{11}}{2} \text{ if } j = 1 \\ b_1 + \dfrac{1}{2}\left\{v_{jj} - v_j' \left[V_{j-1} + (D_j D_j)^{-1}\right]^{-1} v_j\right\} \text{ if } j = 2, .., n \end{cases}.$$

The preceding equation uses a notation where

$$V = (Y - XA)'(Y - XA)$$

has elements v_{ij}, $v_j = (v_{1j}, .., v_{j-1,j})'$ and V_j is the upper left $j \times j$ block of V.

The conditional posterior of η has the conditional posteriors of η_j (for $j = 2, .., n$) being independent of one another with:

$$\eta_j \sim N\left(\overline{\eta}_j, \overline{V}_j\right), \tag{A.15}$$

where

$$\overline{V}_j = \left[V_{j-1} + (D_j D_j)^{-1}\right],$$

$$\overline{\eta}_j = -\psi_{jj} \overline{V}_j v_j.$$

Finally, the conditional posterior for ω has ω_{ji} being independent Bernoulli random variables:

$$\begin{array}{l} \Pr(\omega_{ji} = 1) = \overline{q}_{ji} \\ \Pr(\omega_{ji} = 0) = 1 - \overline{q}_{ji} \end{array}, \tag{A.16}$$

where

$$\overline{q}_{ji} = \frac{\frac{1}{\kappa_{1ji}} \exp\left(-\frac{\psi_{ji}^2}{2\kappa_{1ji}^2}\right) \underline{q}_{ji}}{\frac{1}{\kappa_{1ji}} \exp\left(-\frac{\psi_{ji}^2}{2\kappa_{1ji}^2}\right) \underline{q}_{ji} + \frac{1}{\kappa_{0ji}} \exp\left(-\frac{\psi_{ji}^2}{2\kappa_{0ji}^2}\right) \left(1 - \underline{q}_{ji}\right)}.$$

In summary, an MCMC algorithm for the VAR with SSVS involves sequentially drawing from (A.12), (A.13), (A.14), (A.15) and (A.16).

The VAR with SSVS Prior and Structural Breaks at Known Points in Time

As an intermediate step to using the SSVS prior in a VAR subject to structural breaks, suppose that the breaks occur at known points in time. After each break occurs, a new regime applies and both the VAR coefficients and error covariance matrix can be completely different than before the break. Regimes will be characterized by discrete random variables, s_t (for $t = 1, .., T$) which takes on values $\{1, 2, \ldots, M\}$. Let $S_i = (s_1, \ldots, s_i)'$ and $S^{i+1} = (s_{i+1}, \ldots, s_T)'$. Using our previously-given notation for the dependent and explanatory variables, the likelihood function for our model is defined by assuming $p(y_t|x_t, s_t = m) = p(y_t|x_t, \theta_m)$ for $m = 1, \ldots, M$. We are using the generic notation θ_m to denote the parameters in regime m. In our VAR, $\theta_m = \left\{\alpha^{(m)}, \gamma^{(m)}, \psi^{(m)}, \eta^{(m)}, \omega^{(m)}\right\}$ where the parameters have the same definitions as above except that we are using (m) superscripts to denote parameters in each regime.

Bayesian inference, including posterior simulation, in this model is a straightforward extension of our previous results. That is, conditional on S_T, the data is broken into M regimes. Using only the data in regime m, we can use the algorithm described in the previous section. This can be repeated for $m = 1, .., M$. Formally, we could write out this algorithm by putting (m) superscripts on everything (i.e. on all data quantities, parameters, prior hyperparameters) in the previous section, but for the sake of brevity we do not do so here.

The VAR with SSVS Prior and Structural Breaks at Unknown Points in Time

Our previous discussion of the model with structural breaks at known points in time suggests that a wide variety of hierarchical priors can be used for S_T. An MCMC algorithm will involve taking our previous algorithm for

the VAR with SSVS prior and adding a block which draws from S_T. This is indeed the case. In this paper, we use an approach that is similar to the popular hierarchical prior for S_T developed in Chib (1998). Chib's approach to structural break modeling has been used by many including Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo and Timmerman (2006).

To implement this setup, we begin by selecting M, the maximum number of regimes. We assume that s_t is Markovian. That is,

$$\Pr\left(s_t = j | s_{t-1} = i\right) = \begin{cases} 1-q & \text{if } j = i \neq M, \\ q & \text{if } j = i+1, \\ 1 & \text{if } i = j = M, \\ 0 & \text{otherwise.} \end{cases} \tag{A.17}$$

In words, the time series variable goes from regime to regime. Once it has gone through the m^{th} regime, there is no returning to this regime. It goes through regimes sequentially, so it is not possible to skip from regime i to regime $i+2$. Once it reaches the M^{th} regime it stays there. The difference between the prior used in this paper and Chib's (1998) approach is, that we choose $M = T$ and do not assume that every regime is visited.

Equation (A.17) defines a hierarchical prior for the states. To complete the model, a prior for q is required. A popular choice is a Beta prior: $B\left(\underline{\delta}_1, \underline{\delta}_2\right)$. We select $\underline{\delta}_1 = 1$ and $\underline{\delta}_2 = 0.01$ which is a relatively noninformative choice.

Bayesian inference in our model and the model of Chib (1998) is based on an MCMC algorithm. If $\Theta = \left(\theta_1', \ldots, \theta_M'\right)'$ then the algorithm proceeds by sequentially drawing from

$$\Theta | Y, S_T, q, \tag{A.18}$$

$$S_T | Y, \Theta, q \tag{A.19}$$

and

$$p | Y, \Theta, S_T. \tag{A.20}$$

We have already described how to draw from (A.18). Simulation from (A.19) is done using a method developed in Chib (1996). This involves noting that:

$$p\left(S_T|Y,\Theta,q\right)=p\left(s_T|Y,\Theta,q\right)p\left(s_{T-1}|Y,S^T,\Theta,q\right) \tag{A.21}$$
$$\cdots p\left(s_t|Y,S^{t+1},\Theta,q\right)\cdots p\left(s_1|Y,S^2,\Theta,q\right).$$

Draws from s_t can be obtained using the fact [see Chib (1996)] that

$$p\left(s_t|Y,S^{t+1},\Theta,q\right)\propto p\left(s_t|y_t,\Theta,q\right)p\left(s_{t+1}|s_t,q\right). \tag{A.22}$$

Since $p\left(s_{t+1}|s_t,q\right)$ is the transition probability and the integrating constant can be easily obtained (conditional on the value of s_{t+1}, s_t can take on only two values), we need only to worry about $p\left(s_t|y_t,\Theta,q\right)$. Chib (1996) recommends the following recursive strategy. Given knowledge of $p\left(s_{t-1}=m|x_t,\Theta,q\right)$ (and remembering that, in the VAR case, x_t contains an intercept and lags of the dependent variable), we can obtain:

$$p\left(s_t=j|y_t,\Theta,q\right)=\frac{p\left(s_t=j|x_t,\Theta,q\right)p\left(y_t|x_t,\theta_j\right)}{\sum_{m=j-1}^{j}p\left(s_t=m|x_t,\Theta,q\right)p\left(y_t|x_t,\theta_m\right)}, \tag{A.23}$$

using the fact that

$$p\left(s_t=j|x_t,\Theta,q\right)=(1-q)\cdot p\left(s_{t-1}=j-1|x_t,\Theta,q\right)+q\cdot p\left(s_{t-1}=j|x_t,\Theta,q\right), \tag{A.24}$$

for $j=1,..,M$.

Thus, the algorithm proceeds by calculating (A.22) for every time period using (A.23) and (A.24) beginning at $t=1$ and going forward in time (the so-called *forward iteration*). Then the states themselves are drawn using (A.21), beginning at period T and going backwards in time (we will refer to this as the *backward iteration*).

To be precise, the algorithm begins with $p\left(s_1=1|x_1,\Theta,q\right)=1$ and then uses (A.24) to calculate $p\left(s_t=k|x_t,\Theta,q\right)$ for $t=2,..,T$. This allows for everything in (A.23) to be calculated except for $p\left(y_t|x_t,\theta_m\right)$. But this is just the normal p.d.f. from the SSVS-VAR likelihood for the t^{th} observation (evaluated at the MCMC draw of θ_m). Thus, (A.23) can be evaluated and plugged into (A.22) to obtain the discrete p.d.f. $p\left(s_t|Y,S^{t+1},\Theta,q\right)$ for $t=1,..,T$. Draws from these p.d.f.s can be used in the ordering of (A.21) to draw the states (conditional on all the other parameters of the model).

Finally, derivations in McCulloch and Tsay (1993) imply that the conditional posterior for q is Beta:

$$q \sim B\left(\overline{\delta}_1, \overline{\delta}_2\right),$$

where

$$\overline{\delta}_1 = \underline{\delta}_1 + T - s_T$$

and

$$\overline{\delta}_2 = \underline{\delta}_2 + s_T.$$

With regards to prediction, with the exception of the model with structural breaks, the strategy outlined in the body of the text, using (4) and (5) applied. When breaks are present, the predictive density, conditional on the MCMC draws of the parameters, is a mixture of (5) (with probability $1-q$) and the prior based quantity (with probability q). Since q is small, this prior quantity has little effect but cannot be evaluated analytically due to the hierarchical structure of the prior. For ease of computation, we approximate it using

$$f_N(y^*_{\tau+h}|\tilde{\mu}, \tilde{\Sigma}),$$

where $\tilde{\mu}$ and $\tilde{\Sigma}$ are the mean vector and covariance matrix of a large number of draws from the prior predictive density.

Appendix B: Additional Empirical Results

Table B1: Probability of Inclusion of VAR coefficients (VAR with SSVS, but no breaks)

	Dependent variable		
	unemp. rate	interest rate	inflation
constant	1.000	0.102	0.796
unemp(-1)	1.000	0.982	0.863
unemp(-2)	1.000	0.913	0.479
unemp(-3)	0.114	0.082	0.181
unemp(-4)	0.123	0.225	0.260
interest(-1)	0.255	1.000	0.060
interest(-2)	0.154	0.091	0.056
interest(-3)	0.772	0.762	0.047
interest(-4)	0.143	0.117	0.055
inflation(-1)	0.044	0.913	1.000
inflation(-2)	0.032	0.557	1.000
inflation(-3)	0.033	0.148	0.063
inflation(-4)	0.065	0.284	0.049

Table B2: Probability of Inclusion of Off-Diagonal Elements of Error Cov. (VAR with SSVS, but no breaks)

	unemployment rate	interest rate
unemp		
interest	1.000	
inflation	1.000	0.907

Table B3: Probability of Inclusion of VAR coefficients (VAR with SSVS and breaks)

	Regime I	Regime II	Regime III
	Dep. variable: unemp		
constant	0.998	0.143	0.436
unemp(-1)	1.000	1.000	1.000
unemp(-2)	0.363	0.329	0.104
unemp(-3)	0.353	0.215	0.128
unemp(-4)	0.202	0.147	0.171
interest(-1)	0.704	0.292	0.739
interest(-2)	0.359	0.103	0.321
interest(-3)	0.424	0.883	0.628
interest(-4)	0.624	0.212	0.379
inflation(-1)	0.268	0.400	0.633
inflation(-2)	0.255	0.100	0.250
inflation(-3)	0.345	0.069	0.140
inflation(-4)	0.546	0.075	0.127
	Dep. variable: interest		
constant	0.159	0.168	0.109
unemp(-1)	0.168	0.799	0.139
unemp(-2)	0.089	0.404	0.061
unemp(-3)	0.084	0.165	0.211
unemp(-4)	0.382	0.669	0.336
interest(-1)	1.000	1.000	1.000
interest(-2)	0.455	0.158	0.430
interest(-3)	0.262	0.539	0.466
interest(-4)	0.258	0.606	0.351
inflation(-1)	0.050	0.819	0.099
inflation(-2)	0.037	0.226	0.042
inflation(-3)	0.037	0.120	0.046
inflation(-4)	0.054	0.219	0.074
	Dep. variable: inflation		
constant	0.985	0.694	0.199
unemp(-1)	0.931	0.149	0.152
unemp(-2)	0.127	0.105	0.088
unemp(-3)	0.140	0.062	0.109
unemp(-4)	0.112	0.092	0.083
interest(-1)	0.148	0.083	0.080
interest(-2)	0.120	0.097	0.074
interest(-3)	0.139	30.074	0.071
interest(-4)	0.229	0.176	0.057
inflation(-1)	1.000	1.000	1.000
inflation(-2)	0.426	0.930	0.346
inflation(-3)	0.368	0.161	0.343
inflation(-4)	0.300	0.091	0.264

Table B4: Probability of Inclusion of Off-Diagonal Elements of Error Cov. (VAR with SSVS and breaks)		
	unemployment rate	interest
	Regime I	
unemp		
interest	1.000	
inflation	0.999	0.957
	Regime II	
unemp		
interest	1.000	
inflation	1.000	0.944
	Regime III	
unemp		
interest	0.436	
inflation	1.000	0.104

Table B5: Posterior Means (standard devs) of VAR coefficients (VAR with SSVS and breaks)

	Regime I		Regime II		Regime III	
			Dep. variable: unemp			
constant	1.443	(0.465)	0.018	(0.088)	0.114	(0.155)
unemp(-1)	0.836	(0.180)	0.966	(0.133)	0.949	(0.089)
unemp(-2)	-0.095	(0.155)	-0.068	(0.122)	-0.004	(0.043)
unemp(-3)	-0.090	(0.157)	-0.030	(0.069)	-0.015	(0.059)
unemp(-4)	0.034	(0.095)	-0.005	(0.037)	-0.009	(0.048)
interest(-1)	-0.177	(0.159)	-0.015	(0.030)	-0.084	(0.062)
interest(-2)	0.029	(0.135)	-0.001	(0.015)	-0.023	(0.071)
interest(-3)	0.096	(0.145)	0.088	(0.044)	0.084	(0.079)
interest(-4)	0.141	(0.154)	0.011	(0.031)	0.027	(0.047)
inflation(-1)	-0.034	(0.112)	0.032	(0.040)	0.071	(0.068)
inflation(-2)	-0.063	(0.166)	0.010	(0.027)	0.028	(0.062)
inflation(-3)	-0.064	(0.240)	0.004	(0.022)	0.009	(0.041)
inflation(-4)	0.175	(0.215)	0.002	(0.016)	0.002	(0.029)
			Dep. variable: interest			
constant	0.030	(0.195)	0.038	(0.254)	0.026	(0.117)
unemp(-1)	-0.039	(0.108)	-0.902	(0.636)	-0.037	(0.126)
unemp(-2)	0.024	(0.093)	0.354	(0.541)	0.003	(0.072)
unemp(-3)	0.015	(0.062)	0.045	(0.253)	-0.099	(0.256)
unemp(-4)	0.078	(0.102)	0.379	(0.334)	0.167	(0.271)
interest(-1)	0.906	(0.173)	0.574	(0.175)	1.219	(0.131)
interest(-2)	-0.156	(0.219)	0.005	(0.052)	-0.123	(0.181)
interest(-3)	0.052	(0.139)	0.144	(0.165)	-0.099	(0.133)
interest(-4)	0.045	(0.107)	0.199	(0.195)	-0.055	(0.092)
inflation(-1)	0.004	(0.030)	0.372	(0.253)	0.017	(0.058)
inflation(-2)	0.005	(0.045)	-0.088	(0.225)	0.007	(0.052)
inflation(-3)	0.005	(0.039)	0.014	(0.170)	0.004	(0.057)
inflation(-4)	0.005	(0.033)	-0.057	(0.144)	-0.013	(0.060)
			Dep. variable: inflation			
constant	1.042	(0.346)	0.310	(0.283)	0.033	(0.087)
unemp(-1)	-0.167	(0.070)	-0.020	(0.065)	-0.023	(0.077)
unemp(-2)	0.009	(0.044)	0.002	(0.055)	0.011	(0.065)
unemp(-3)	0.013	(0.034)	-0.000	(0.035)	0.021	(0.064)
unemp(-4)	0.006	(0.023)	0.006	(0.033)	0.003	(0.028)
interest(-1)	-0.001	(0.022)	-0.000	(0.010)	0.001	(0.011)
interest(-2)	0.000	(0.023)	0.003	(0.016)	-0.000	(0.014)
interest(-3)	0.007	(0.038)	0.001	(0.012)	-0.000	(0.014)
interest(-4)	-0.019	(0.052)	-0.007	(0.020)	0.000	(0.009)
inflation(-1)	1.103	(0.162)	1.521	(0.125)	1.134	(0.122)
inflation(-2)	-0.137	(0.198)	-0.524	(0.190)	-0.089	(0.146)
inflation(-3)	-0.076	(0.121)	-0.031	(0.101)	-0.058	(0.097)
inflation(-4)	-0.033	(0.075)	0.000	(0.029)	-0.030	(0.066)

Table B6: Posterior Means (standard devs) of Error Covariance (VAR with SSVS and breaks)

	unemployment rate	interest rate	inflation
		Regime I	
unemp	2.296		
	(0.257)		
interest	1.416	2.642	
	(0.406)	(0.290)	
inflation	0.259	-0.020	4.532
	(0.433)	(0.387)	(0.482)
		Regime II	
unemp	3.046		
	(0.272)		
interest	2.015	1.087	
	(0.428)	(0.111)	
inflation	-0.152	-0.150	2.307
	(0.438)	(0.132)	(0.194)
		Regime III	
unemp	5.231		
	(0.442)		
interest	2.896	2.825	
	(0.681)	(0.231)	
inflation	0.615	-0.336	4.728
	(0.746)	(0.316)	(0.386)

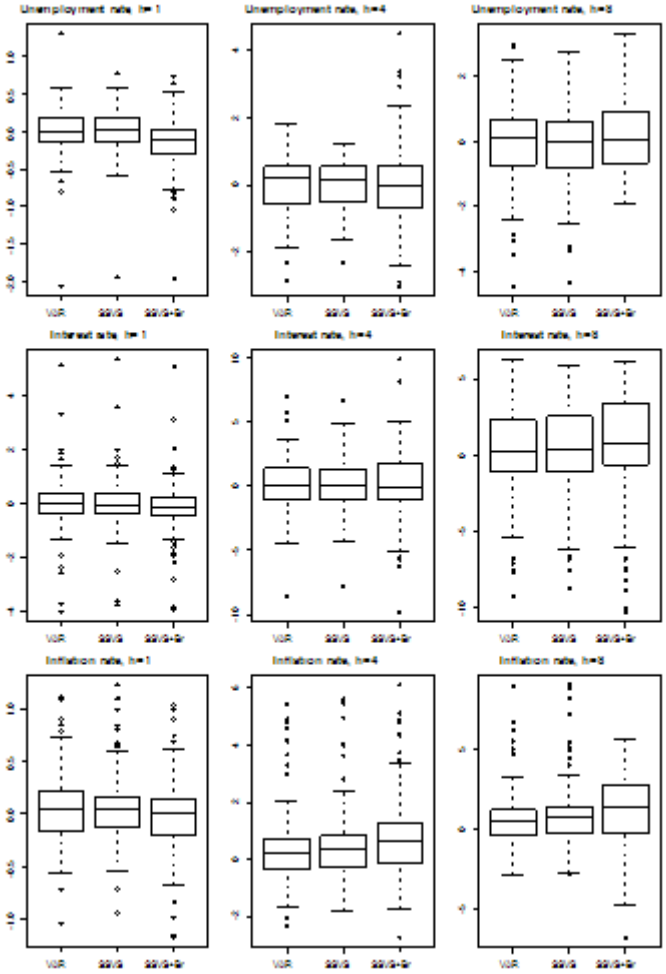

Figure B1: Boxplots of Point Forecasts for All Variables, Models and Forecast Horizons