Forecasting output growth by the yield curve: the role of structural breaks

He, Zhongfang

Department of Economics, University of Toronto, Bank of Canada

April 2009

Forecasting Output Growth by the Yield Curve: The Role of Structural Breaks*

Zhongfang He

Department of Economics

University of Toronto

April 2009

Abstract

This paper proposes a new structural-break vector autoregressive (VAR) model for predicting real output growth by the nominal yield curve information. We allow for the possibility of both in-sample and out-of-sample breaks in parameter values and use information in historical regimes to make inference on out-of-sample breaks. A Bayesian estimation and forecasting procedure is provided which accounts for the uncertainty of structural breaks and model parameters. We discuss dynamic consistency when forecasting recursively with structural break models, which has been ignored in the existing literature, and provide a solution. Applied to monthly US data from 1964 to 2006, we find strong evidence of structural breaks in the predictive relation between the yield curve and output growth in late 1979 and early 1983. The short rate has more predictive power for output growth than the term spread before 1979 while the term spread becomes more significant since the break

of 1983. Incorporating the possibility of structural breaks improves out-of-sample forecasts of output growth from 1 to 12 months ahead.

1 Introduction

Forecasting real economic activity such as output growth is an importance issue in empirical economics. Research over the last few decades has found that the nominal yield curve contains important predictive information for subsequent real economic growth. Examples of this voluminous literature include Harvey (1989), Laurent (1988,1989), Stock and Watson (1989), Chen (1991), Estrella and Hardouvelis (1991), Plosser and Rouwenhorst (1994), Davis and Fagan (1997), Estrella and Mishkin (1997,1998), Hamilton and Kim (2002) and Ang, Piazzesi, and Wei (2006), among many others.

Recently there has been growing evidence that the relationship between the yield curve and subsequent economic growth may be unstable over time; see, for example, Stock and Watson (1999,2003), Estrella, Rodrigues, and Schich (2003) and Giacomini and Rossi (2006). This forecasting instability poses a challenge for predicting output growth by the yield curve. If the predictive relation experiences structural breaks in the past, it may change in the future as well. Ignoring the possibility of future structural breaks could result in biased and poor forecasts of output growth. This paper is the first in the literature that studies the problem of forecasting output growth with the yield curve information in the presence of both in-sample and out-of-sample structural instability.

In this paper, we take into account the possibility that structural breaks in the predictive relation between the yield curve and output growth have occurred in the past as well as the possibility that they may occur in the future. In contrast to the existing work on the stability of the yield curve's predictive ability for output growth using univariate models, we jointly model the dynamics of the output growth and the yield curve by a tri-variate vector autoregressive (VAR) model of output growth, the short rate and the term spread in conjunction with structural breaks. The short rate is the nominal interest rate on short maturity government debt while the term spread is the difference between nominal interest rates on long and short maturity government debts. They are commonly

used in the literature to capture the predictive information in the yield curve for real economic activity. Stock and Watson (2003) provides some economic motivations for the use of these two interest rate variables for forecasting output growth. Ang et al. (2006) shows that VAR forecasts of quarterly GDP growth using the short rate and term spread are more accurate than univariate regressions of output growth on these two interest rate variables at all horizons considered in their paper. Given the well-documented evidence of structural changes in interest rates, e.g. Gray (1996), Ang and Bekaert (2002), Bansal and Zhou (2002) and Pesaran, Pettenuzzo, and Timmermann (2006), jointly modeling the structural breaks of the yield curve variables and output growth in a VAR model helps avoid attributing structural changes in interest rates to the predictive relation between the yield curve and output growth.

Built on the work of Chib (1998) and Pesaran et al. (2006), we model the structural break process as a hierarchical hidden Markov chain[1]. The parameters of the VAR model may take different values in different break segments and are assumed to be drawn from a common meta distribution. As data in a new regime becomes available, the meta distribution is updated by Bayes rule. Hence information in the parameters of previous break segments is used to learn about parameters in the new regime in an efficient way. Forecasts are made by integrating out the uncertainty about both the in-sample and out-of-sample breaks and parameters. A Markov chain Monte Carlo (MCMC) algorithm is developed to estimate the structural break VAR model, which extends the Pesaran et al. (2006) method for univariate settings to multivariate models. We provide a careful discussion of the prior on the number of in-sample regimes implied by the hierarchical structure of this type of model. This is in contrast to Pesaran et al. (2006) which imposes a uniform prior on the number of in-sample regimes and hence is inconsistent with hierarchical priors on other model parameters. We also discuss the issue of dynamic consistency, that is, the compatibility of assumptions through time concerning the possible number of structural breaks, when forecasting recursively with structural break models in

---

[1]A partial sample of the alternative Bayesian models that allow for parameter shifts of random magnitude and timing includes McCulloch and Tsay (1993), Giordani and Kohn (2006), Koop and Potter (2007), Maheu and Gordon (2008) and Maheu and McCurdy (2007), among others.

the way of Pesaran et al. (2006). A new forecasting approach is proposed that guarantees the dynamic consistency in recursive forecasting with structural break models.

The proposed model is applied to the monthly US data from January 1964 to December 2006. We consider a full-break specification in which all parameters of the VAR model are subject to structural breaks as well as a partial-break specification in which only the intercept and covariance matrix of the VAR model have structural breaks. We find that the full-break model is favored by the data despite the greater parsimony of the partial-break model. A full-break specification with 2 in-sample breaks provides the best description of the data among the models considered in this paper.

The break dates are identified at October 1979 and January 1983, which coincides closely with the change in monetary policy regime with the advent of the US Fed chairman Volcker in late 1979. Before 1979, the short rate predicts output growth while the term spread is largely insignificant. The regime between 1979 and 1983 is marked by exceptionally high volatilities of all three variables, during which neither the short rate nor the term spread is able to predict output growth. The most recent regime since 1983 has much lower volatility. During this period, the predictive power of the short rate largely disappears while the term spread becomes significant. This new finding is in contrast to the studies that do not consider the possibility of structural breaks. Most of the studies in the literature find that the short rate has little marginal predictive content for output growth once spreads are included, e.g. Plosser and Rouwenhorst (1994) and Stock and Watson (2003). In contrast, Ang et al. (2006) finds that the short rate has more predictive power for quarterly GDP growth than term spreads. This paper, by taking into account the possibility of parameter shifts, finds that the relative importance of the short rate and term spread is changing over time and the spread has more predictive power for output growth than the short rate in the most recent regime.

We perform recursive out-of-sample density forecasting exercises from January 2002 to December 2006 to compare the performance of the proposed structural break VAR model with the conventional no-break VAR model. The results show that incorporating the possibility of structural breaks significantly improves the forecasting accuracy of

output growth from 1 to 12 months ahead. The improvements in forecasting accuracy are steady and almost continuous throughout the forecasting period. We also experiment with imposing informative priors on the structural break VAR model for forecasting since a number of studies (Litterman (1980,1986),Kadiyala and Karlsson (1997)) have advocated their use in Bayesian VAR forecasting. We find that in our forecasting exercise, informative priors do not necessarily lead to better forecasts than the conventional diffuse priors.

The rest of the paper is organized as follows. Section 2 describes the structural break VAR model and the estimation method. Section 3 explains the forecast procedure. The empirical estimates are presented in Section 4. Section 5 is the conclusions. Technical details of the estimation algorithm are presented at the appendices.

2 The Model

We consider a tri-variate VAR model of the output growth, the short rate and the term spread for forecasting output growth, which has been found to produce superior out-of-sample forecasts of GDP growth than univariate OLS regressions (Ang et al. (2006)).

The VAR model is assumed to be subject to a random number of structural breaks which separate different regime segments. A subset of the VAR coefficients may change their values in different regimes. Formally, the structural break VAR model is

$$y_t = \mu_{s_t} + \Phi_{s_t} y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_{s_t}) \tag{1}$$

where $y_t \equiv (g_t, r_t, x_t)'$, g_t is the output growth, r_t is the short rate and x_t is the term spread at time t, $t = 1, 2, ..., T$.

The regimes are indexed by a state variable $s_t \in \{1, 2, \ldots, K\}$ following the transition

matrix

$$
\pi = \begin{pmatrix}
\pi_{1,1} & 1 - \pi_{1,1} & 0 & \ldots & & 0 \\
0 & \pi_{2,2} & 1 - \pi_{2,2} & \ldots & & 0 \\
& & & \ldots & & \\
0 & 0 & \ldots & \pi_{K-1,K-1} & 1 - \pi_{K-1,K-1} & \\
0 & 0 & \ldots & & 0 & 1
\end{pmatrix} \tag{2}
$$

where $\pi_{i,j}$ is the probability of moving to regime j given that the current regime is i. Note that at each point of time, the state variable s_t can either stay in the current regime or jump to the next one. The transition terminates in regime K. We will denote $\pi_i \equiv \pi_{i,i}$ for notational simplicity. A break occurs at time t if $s_t \neq s_{t-1}$. Note $s_t = K$ implies that K regimes, or, $K - 1$ breaks, have occurred in the data up to time t.

This formulation of structural break model is originally proposed by Chib (1998) and is used extensively in many subsequent studies of the literature, e.g. Pastor and Stambaugh (2001), Kim, Morley, and Nelson (2005), Pesaran et al. (2006), Liu and Maheu (2008) and He and Maheu (2008). It has two major benefits. First, it automatically imposes an ordering of the regime segments and hence solves the identification problem of regimes. Moreover, this formulation of structural breaks can be viewed as a hidden Markov model (HMM), facilitating the marriage with the existing large literature on HMM and hence development of efficient estimation methods (Scott (2002)). The regime-switching model of Hamilton (1988) can be viewed as a special case of this setup if identical states are assumed to recur (Pesaran et al. (2006)).

Chib (1998) has developed an efficient Bayesian MCMC algorithm for estimating this class of structural break models. But the Chib method can not handle the possibility of out-of-sample breaks during forecasting horizons and hence is not well suited for out-of-sample forecasts unless one is willing to assume that no new breaks could occur out of sample.

To perform forecasts while taking into account possible out-of-sample breaks, we need to model the underlying process of the parameters in different regimes. In this paper, we follow Pesaran et al. (2006) and posit hierarchical priors for the regime parameters.

Parameters in each regime are assumed to be drawn from a common meta distribution. As data from new regimes become available, the meta distribution is updated by Bayes rule. Hence information in the parameters of previous regimes is used to learn about parameters in the new regime in an efficient way. This method has the attractive feature that it retains the sampling efficiency of the Chib method while conveniently modeling the underlying process of the regime parameters and hence is well suited for out-of-sample forecasts. We develop a new MCMC algorithm to estimate this class of models, extending the Pesaran et al. (2006) method for univariate settings to multivariate models.

Let $\phi_k = vec\left([\mu_k \; \Phi_k]'\right)$ be the vector containing the elements of μ_k and Φ_k. We assume that the linear coefficients ϕ_k, the covariance matrices Σ_k and the transition probabilities π_k are independently drawn from the following distributions respectively.

$$\phi_k \sim N(b_0, B_0)$$

$$\Sigma_k \sim IW(\Omega_0, v_0 + 3)$$

for $k = 1, 2, ..., K$, where IW denotes the inverse Wishart distribution, and

$$\pi_k \sim Beta(\alpha_0, \beta_0)$$

for $k = 1, 2, ..., K - 1$.

At the next level of the hierarchy, we assume that

$$b_0 \sim N(a_0, A_0)$$

$$B_0 \sim IW(D_0, d_0)$$

$$\Omega_0 \sim IW(\Psi_0, f_0)$$

$$v_0 \sim Gamma(\rho_0, \lambda_0)$$

$$\alpha_0 \sim Gamma(q_0, \gamma_0)$$

$$\beta_0 \sim Gamma(r_0, \delta_0)$$

where a_0, A_0, D_0, d_0, Ψ_0, f_0, ρ_0, λ_0, q_0, γ_0, r_0 and δ_0 are hyper-parameters and are specified *a priori*. This hierarchical structure creates dependence between parameters in different

regimes. Given an estimation sample, model parameters from different in-sample regimes ϕ_k, Σ_k and π_k are used to update the distributions of the hierarchical prior parameters b_0, B_0, Ω_0, v_0, α_0 and β_0 by Bayes rule. Inference on parameters of possible out-of-sample regimes is then based on the updated distributions of these hierarchical prior parameters. Hence information contained in in-sample estimates is efficiently used to produce forecasts outside the estimation sample.

To conduct Bayesian estimation, we divide the parameters into 3 blocks for a given number of in-sample regimes K: the latent states $S = (s_1, s_2, ..., s_T)$, parameters of the hierarchical priors $\Theta_0 = (b_0, B_0, \Omega_0, v_0, \alpha_0, \beta_0)$ and the other model parameters $\Theta = (\phi_1, ..., \phi_K, \Sigma_1, ..., \Sigma_K, \pi_1, ..., \pi_{K-1})$. A Gibbs sampler is developed to estimate this hierarchical structural break model, which iterates sampling from the following conditional distributions

- $S|\Theta_0, \Theta$

- $\Theta_0|S, \Theta$

- $\Theta|S, \Theta_0$

The details of the algorithm are provided at Appendix A.

Inference on the number of in-sample regimes K is conducted based on the posterior distribution $p(s_T = K|Y_T)$, $K = 1, 2, ..., \overline{K}$, since according to the transition matrix of Equation (2), the number of in-sample regimes K equals the state variable at the end of data sample s_T. In theory, the possible number of in-sample regimes $\overline{K}$ can be as large as the number of observations T, i.e. a structural break at every period of time. But in practice, it can often be set to be a relatively small number $\overline{K} < T$ provided that the posterior distribution $p(s_T = K|Y_T)$ does not support going beyond $\overline{K}$ in-sample regimes.

Applying Bayes rule, the posterior distribution $p(s_T = K|Y_T)$ can be decomposed as

$$p(s_T = K|Y_T) \propto p(s_T = K)p(Y_T|s_T = K)$$

where $p(s_T = K)$ is the prior probability and $p(Y_T|s_T = K)$ is the marginal likelihood of Y_T given K in-sample regimes. There is a significant Bayesian literature on methods of

computing the marginal likelihoods, e.g. Gelfand and Dey (1994), Newton and Raftery (1994), Chib (1995), Fruhwirth-Schnatter (1995, 2004), Meng and Wong (1996), and Chib and Jeliazkov (2001). Miazhynskaia and Dorffner (2006) provides a nice comparison of the various methods of computing marginal likelihoods. In this paper, we adopt the modified harmonic mean method of Gelfand and Dey (1994) which has been found to be accurate (Miazhynskaia and Dorffner (2006)) while computationally convenient. The details of implementing the modified harmonic mean method for the structural break VAR model are provided in Appendix B.

The prior $p(s_T = K)$ on the number of in-sample regimes K is implied by the hierarchical prior on transition probabilities π_k, $k = 1, 2, ..., \overline{K} - 1$, as

$$p(s_T) = \int p(s_T | \pi_1, ..., \pi_{\overline{K}-1}) \prod_{k=1}^{\overline{K}-1} p(\pi_k | \alpha_0, \beta_0) p(\alpha_0, \beta_0) d\pi_1 \cdots d\alpha_0 d\beta_0$$

This distribution has no closed form but can be simulated by using the approximation

$$p(s_T) \approx \frac{1}{n} \sum_{i=1}^{n} \widehat{p}(s_T | \pi_1^{(i)}, ..., \pi_{\overline{K}-1}^{(i)})$$

where $\pi_k^{(i)} \sim Beta\left(\alpha_0^{(i)}, \beta_0^{(i)}\right)$, $\alpha_0^{(i)} \sim Gamma\left(q_0, \gamma_0\right)$ and $\beta_0^{(i)} \sim Gamma\left(r_0, \delta_0\right)$. One can sample a path $\{1, s_2^{(i)}, s_3^{(i)}, \cdots, s_T^{(i)}; s_T^{(i)} \leq \overline{K}\}$ conditional on each draw of $\pi_k^{(i)}$, $\alpha_0^{(i)}$, $\beta_0^{(i)}$ and keep $s_T^{(i)}$ as a draw from $p(s_T)$. The prior probability $p(s_T = K)$ can be computed as $\sum_{i=1}^{n} I\{s_T^{(i)} = K\}/n$, where the indicator function $I\{s_T^{(i)} = K\} = 1$ if $s_T^{(i)} = K$ and 0 otherwise. This is in contrast to Pesaran et al. (2006) which imposes $p(s_T = K) = 1/\overline{K}$ and hence is inconsistent with the actual prior $p(s_T = K)$ implied by the hierarchical priors on other model parameters.

3 Out-of-Sample Forecasts

In this section, we show how to produce out-of-sample forecasts of output growth from the model proposed in Section 2. Given the posterior draws of parameters and latent states based on time T information $Y_T \equiv \{y_1, y_2, ..., y_T\}$, we forecast future values of output growth, which is the first element of the vector y, by taking into account the uncertainty about both in-sample and out-of-sample breaks and parameters.

Let h be the forecasting horizon. The predictive distribution of output growth g_{T+h} integrates out uncertainty about the number of in-sample breaks by Bayesian model averaging

$$p(g_{T+h}|Y_T) = \sum_{K=1}^{\overline{K}} p(g_{T+h}|Y_T, s_T = K)p(s_T = K|Y_T) \tag{3}$$

where K is the number of in-sample regimes in the data Y_T and has the upper limit $\overline{K}$. The weight used in the averaging, $p(s_T = K|Y_T)$, is the posterior probability of the structural break VAR model with K in-sample regimes. The other ingredient $p(g_{T+h}|Y_T, s_T = K)$ is the predictive distribution of output growth conditional on K in-sample regimes.

To integrate out the uncertainty about out-of-sample breaks, the conventional method, e.g. Pesaran et al. (2006), applies the decomposition

$$p(g_{T+h}|Y_T, s_T = K) = \sum_{j=K}^{K+h} p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j)p(s_{T+h} = j|Y_T, s_T = K) \tag{4}$$

that is, conditional on K in-sample regimes at time T, it is assumed that up to $K + h$ regimes could occur at time $T + h$. This approach is reasonable when viewed statically. But once being put in a recursive forecasting context as in practice, it becomes logically inconsistent if the upper limit of in-sample regimes $\overline{K}$ is kept fixed throughout. To see this, consider the example $K = \overline{K}$ and $h = 1$. When making forecasts at time T, one assumes that $s_{T+1} = \overline{K} + 1$ is possible according to Equation (4). But after arriving at $T + 1$, one assumes $s_{T+1} \le \overline{K}$ as the maximum number of in-sample regimes in the data Y_{T+1} is fixed at $\overline{K}$, which is inconsistent with the assumptions made at time T. This creates a dynamic inconsistency problem: assumptions concerning the possible number of structural breaks are not consistent through time when forecasting recursively. Increasing the upper limit $\overline{K}$ by 1 when moving 1 period forward is not an attractive solution since the extra computation cost would soon become too high to be practical and the number of possible structural breaks entertained should not be unboundedly increasing.

In this paper, we propose to set an upper limit on the total number of *both in-sample*

and out-of-sample regimes

$$p(g_{T+h}|Y_T, s_T = K) = \sum_{j=K}^{min\{\overline{K}, K+h\}} p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j)p(s_{T+h} = j|Y_T, s_T = K)$$

(5)

in order to be dynamically consistent. When forecasting h-periods ahead at any time T, the new method guarantees that the number of regimes assumed for time $T + h$ satisfies $s_{T+h} \leq \overline{K}$ regardless of the current regime s_T. This will be consistent with the assumption after one actually arrives at time $T + h$ that the upper limit of the number of in-sample regimes s_{T+h} is $\overline{K}$. This simple modification solves the dynamic inconsistency problem suffered by the conventional method while entertaining no extra computation cost.

When viewed as a mixture distribution, sampling from the predictive distribution of output growth $p(g_{T+h}|Y_T)$ in Equation (3) is straightforward:

Step 1. Compute the posterior probabilities $p(s_T = K|Y_T)$ for $K = 1, 2, ..., \overline{K}$.

Step 2. Sample indices $k \in \{1, 2, ..., \overline{K}\}$ from a multinomial distribution with the posterior probabilities $p(s_T = K|Y_T)$ as parameters.

Step 3. If $k = K$, then sample $g_{T+h}^{(i)}$ from the predictive distribution $p(g_{T+h}|Y_T, s_T = K)$ according to Equation (5).

The resulting sample of output growth $\{g_{T+h}^{(i)}\}_{i=1}^n$ will provide a complete distribution of the future output growth based on the current information set. The predictive mean of any function of the output growth $f(g_{T+h})$ can be consistently estimated as

$$E[f(g_{T+h})|Y_T] \approx \frac{1}{n} \sum_{i=1}^n f(g_{T+h}^{(i)})$$

There are two important ingredients for this forecasting method. One is the posterior probability of in-sample regimes $p(s_T = K|Y_T)$, whose computation has been discussed in detail in the preceding section of model description. The other ingredient is the predictive distribution

$$p(g_{T+h}|Y_T, s_T = K) = \sum_{j=K}^{min\{\overline{K}, K+h\}} p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j)p(s_{T+h} = j|Y_T, s_T = K)$$

11

which is used in Step 3 above. For ease of exposition, consider first the predictive distribution conditional on the relevant model parameters. Let θ denote the set containing the in-sample parameters ϕ_K, Σ_K and out-of-sample parameters π_K, π_{K+1}, ϕ_{K+1}, Σ_{K+1},..., $\pi_{min\{\overline{K},K+h\}}$, $\phi_{min\{\overline{K},K+h\}}$, $\Sigma_{min\{\overline{K},K+h\}}$, that is,

$$\theta = \left(\phi_K, \Sigma_K, \pi_K, \pi_{K+1}, \phi_{K+1}, \Sigma_{K+1}, ..., \pi_{min\{\overline{K},K+h\}}, \phi_{min\{\overline{K},K+h\}}, \Sigma_{min\{\overline{K},K+h\}}\right)$$

We have

$$p(g_{T+h}|Y_T, s_T = K, \theta)$$
$$= \sum_{j=K}^{min\{\overline{K},K+h\}} p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j, \theta)p(s_{T+h} = j|Y_T, s_T = K, \theta) \qquad (6)$$

Consider the component distribution

$$p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j, \theta)p(s_{T+h} = j|Y_T, s_T = K, \theta), \quad j \geq K$$

of Equation (6), which specifies a total of j regimes at time $T + h$ with K in-sample regimes and $j - K$ out-of-sample regimes. For the case $j > K$, i.e. $j - K$ out-of-sample breaks occurring during $T+1, ..., T+h$, it is needed to integrate over all possible locations of the out-of-sample breaks. Let τ_k be the location of the k-th break point. Applying the law of total probability, we have

$$p(g_{T+h}|Y_T, s_T = K, s_{T+h} = j, \theta)p(s_{T+h} = j|Y_T, s_T = K, \theta)$$
$$= \sum_{1\leq i_1 < i_2 ... < i_{j-K} \leq h} p(\tau_K = T + i_1, \tau_{K+1} = T + i_2, ..., \tau_{j-1} = T + i_{j-K}|Y_T, s_T = K, \theta)\cdot$$
$$p(g_{T+h}|Y_T, s_T = K, \tau_K = T + i_1, \tau_{K+1} = T + i_2, ..., \tau_{j-1} = T + i_{j-K}, \theta) \qquad (7)$$

This integrates over all possible locations of out-of-sample breaks. The probability of the out-of-sample break scenario in the first term of Equation (7) is given by

$$p(\tau_K = T + i_1, \tau_{K+1} = T + i_2, ..., \tau_{j-1} = T + i_{j-K}|Y_T, s_T = K, \theta)$$
$$= \pi_K^{i_1-1}(1 - \pi_K)\pi_{K+1}^{i_2-i_1-1}(1 - \pi_{K+1})\cdots\pi_j^{h-i_{j-K}} \qquad (8)$$

12

The clearest way of writing out the predictive densities under each scenario of out-of-sample breaks

$$p(g_{T+h}|Y_T, s_T = K, \tau_K = T + i_1, \tau_{K+1} = T + i_2, ..., \tau_{j-1} = T + i_{j-K}, \theta)$$

which is the second term in Equation (7), is to use recursive equations. Since the densities under specific scenarios of out-of-sample breaks are Gaussian, we only need to specify their means and variances. Let $\mu(i)$, $\Sigma(i)$ be the mean and covariance matrix of the predictive distribution of y_{T+i} from the structural break VAR model of Equation (1). We have

$$\mu(1) = \mu_{s_{T+1}} + \Phi_{s_{T+1}} y_T,$$

$$\Sigma(1) = \Sigma_{s_{T+1}},$$

$$\mu(i) = \mu_{s_{T+i}} + \Phi_{s_{T+i}} \mu(i-1),$$

$$\Sigma(i) = \Sigma_{s_{T+i}} + \Phi_{s_{T+i}} \Sigma(i-1) \Phi'_{s_{T+i}} \tag{9}$$

where s_{T+i} is the index of the regime at $T + i$, $i = 1, 2, ..., h$. Let $e \equiv (1 \ \ 0 \ \ 0)'$. It can be shown (Lutkepohl (2006)) that the density of output growth g_{T+h}, which is the first element of the vector y_{T+h}, is

$$g_{T+h}|Y_T, s_T = K, \tau_K, \tau_{K+1}, ..., \tau_{j-1}, \theta \sim N(e'\mu(h), e'\Sigma(h)e) \tag{10}$$

where $\mu(h)$ and $\Sigma(h)$ track the specific scenario of out-of-sample breaks $\tau_K, \tau_{K+1}, ..., \tau_{j-1}$ by tracking the out-of-sample states $s_{T+1}, ..., s_{T+h}$ according to Equation (9). For the case $j = K$, i.e. no out-of-sample breaks, it is straightforward to show

$$p(s_{T+h} = K|Y_T, s_T = K, \theta) = \pi_K^h \tag{11}$$

and the predictive density $p(g_{T+h}|Y_T, s_T = K, s_{T+h} = K, \theta)$ can be obtained from Equations (9) and (10) by setting $s_{T+1} = \cdots = s_{T+h} = K$.

When there is uncertainty surrounding the parameters, the predictive distribution

13

$p(g_{T+h}|Y_T, s_T = K)$ needs to integrate over the parameters.

$$p(g_{T+h}|Y_T, s_T = K)$$
$$= \int p(g_{T+h}|Y_T, s_T = K, \theta) p(\theta|Y_T, s_T = K) d\theta$$
$$\approx \frac{1}{n} \sum_{i=1}^{n} p(g_{T+h}|Y_T, s_T = K, \theta^{(i)}) \tag{12}$$

where the components in $\theta^{(i)}$ are $\pi_k^{(i)} \sim Beta(\alpha_0^{(i)}, \beta_0^{(i)})$ for $k = K, ..., min\{\overline{K}, K+h\}$, $\phi_k^{(i)} \sim N(b_0^{(i)}, B_0^{(i)})$ and $\Sigma_k^{(i)} \sim IW(\Omega_0^{(i)}, v_0^{(i)})$ for $k = K+1, ..., min\{\overline{K}, K+h\}$. The parameters $\{\alpha_0^{(i)}, \beta_0^{(i)}, b_0^{(i)}, B_0^{(i)}, \Omega_0^{(i)}, v_0^{(i)}, \phi_K^{(i)}, \Sigma_K^{(i)}\}_{i=1}^n$ are posterior draws from the structural break VAR model based on the information set Y_T and K in-sample regimes. Computing the conditional predictive density $p(g_{T+h}|Y_T, s_T = K, \theta^{(i)})$ of Equation (12) follows discussions in the preceding paragraph. To sample from this distribution, one can first simulate a path of out-of-sample regimes $\{s_{T+1}^{(i)}, ..., s_{T+h}^{(i)}\}$ based on the transition probabilities $\pi_K^{(i)}, ..., \pi_{min\{\overline{K}, K+h\}}^{(i)}$ and then sample $g_{T+h}^{(i)}$ from the predictive distribution under the simulated path of regimes according to Equations (9) and (10).

4   Empirical Results

We use monthly nominal zero-coupon yield data with maturities of 3 months and 5 years from the Center for Research in Security Prices (CRSP) spanning January 1964 to December 2006. Prior to this period, there were few traded long bonds. So data on long yields before 1964 may be unreliable (Fama and Bliss (1987)). The short rate is the 3-month yield from the Fama-Bliss risk-free rate file. The term spread is constructed as the 5-year yield minus the 3-month yield, with the 5-year yield data derived from the Fama-Bliss discount bond file. Ang et al. (2006) finds that this term spread has the best predictive power for quarterly GDP growth among all the maturities of spreads considered in their paper. All yield data are continuously compounded. Data on real output growth is the log growth rate of industrial production index from the FRED dataset of the US Federal Reserve. A plot of the data is presented in Figure 1. All data are scaled up by 100 in estimation.

4.1 In-Sample Estimates

First, we conduct an in-sample study of the full-sample data from January 1964 through December 2006. The purpose is to identify if there are structural breaks in the predictive content of the yield curve for output growth and, if so, how the predictive relations change over time.

Let $m = 3$ be the number of variables in the structural break VAR model. The priors are set to be: $b_0 \sim N(0, 100I_{m(m+1)})$, $B_0 \sim IW(I_{m(m+1)}, m(m+1)+4)$, $\Omega_0 \sim IW(0.001I_{m(m+1)}, m+4)$, $v_0 \sim Gamma(2,3)$, $\alpha_0 \sim Gamma(20,1)$ and $\beta_0 \sim Gamma(2, 0.05)$, where $I_{m(m+1)}$ denotes a $m(m+1) \times m(m+1)$ identity matrix. These priors are diffuse over realistic ranges of values for the parameters. We set the upper limit on the number of in-sample regimes $\overline{K}$ to be 5. As will be seen below, the posterior distribution on the number of in-sample regimes does not support going beyond 5. Table 1 contains the simulated prior distribution on the number of in-sample regimes implied by the hierarchical prior of transition probabilities. It can be seen that the prior probability $p(s_T = 1) = 0.733$ and hence the prior strongly favors no break in the data. We discard 5,000 initial draws and retain the next 80,000 for posterior analysis. The chain mixes well. The acceptance rates of the Metropolis-Hastings steps are all in the range between 0.3 and 0.5. As an example, the posterior draws of the predictive coefficient of the term spread in the first regime is presented in Figure 2.

We estimate the no-break VAR model as well as the full-break VAR model of Equation (1) with 1 to 4 in-sample breaks, that is, $K = 1, 2, ..., 5$. The marginal likelihoods peak at 2 in-sample breaks and diminishes as more breaks are introduced. The resulting log marginal likelihoods are presented in Table 2. It can be seen that the full-break VAR model with 2 in-sample breaks has the largest marginal likelihood. Table 2 provides the posterior distribution of the number of in-sample regimes K along with its simulated prior distribution. The posterior probability $p(s_T = 3|Y_T)$ is 0.998. So there is overwhelming evidence supporting 3 in-sample regimes or 2 in-sample breaks. The large difference in the prior and posterior distributions of K suggests that the data is informative about the number of in-sample regimes. Table 3, 4 and 5 provide the parameter estimates

of the full-break VAR model with 3 in-sample regimes. The slope coefficients of the full-break VAR model exhibit smaller differences across regimes than the intercepts and covariance matrices. So a partial-break VAR model in which only the intercept and the covariance matrix have structural breaks is also investigated. We find that the log marginal likelihoods of the partial break VAR model with 1, 2 and 3 in-sample breaks are -1044.06, -1004.06 and -1000.45 respectively, significantly below those of the corresponding full-break VAR models.

Based on the estimates of marginal likelihoods, the full-break VAR model with 2 in-sample breaks is found to provide the best description of the data among the models considered. The posterior median of the first break date is October 1979 with the 95% credible set between July 1979 and October 1979 while the posterior median of the second break date is January 1983 with the 95% credible set between November 1982 and March 1985. The posterior distributions of the break dates are plotted in Figure 3. These estimates of break dates are broadly consistent with the findings of previous studies using univariate models. For example, Estrella et al. (2003) suggests a break around September 1983 when forecasting monthly industrial production growth by term spread. Pesaran et al. (2006) finds evidence of breaks in an AR(1) model of monthly 3-month T-bill rates at September 1979 and September 1982. Maheu and Gordon (2008) finds evidence of break in an AR(2) model of quarterly GDP growth at the 3rd quarter of 1983.

There seems to be a compelling connection of the break dates to the change in the monetary policy regime with the advent of the US Fed chairman Volcker in the late 1979. Some fundamental changes in the Fed's operating procedure took place beginning at October of that year. The move to a monetary policy regime targeting money growth was essentially over by 1983. These dates coincide closely with the estimates of break dates. As will be seen below, the behavior and relation of the data series vary greatly across regimes identified by these break dates.

Estimates of the parameters for the full-break VAR model with 3 in-sample regimes are presented in Table 3, 4 and 5. Parameters significant at the 95% level, that is, the value of 0 is outside the 95% credible set, are marked by "*" for visual convenience. For

comparison, estimates of the corresponding no-break VAR model are presented in Table 6. It can be seen that, for the structural break VAR model, the covariances between the innovations of the variables change over the 3 regimes and the second regime is marked by high volatility of all three variables. Before this regime, it is the short rate that significantly predicts output growth while the term spread is insignificant. During the high-volatility regime, none of these variables can predict the output growth. After the high-volatility regime, the predictive power of the short rate disappears while the term spread becomes significant. Figure 4 plots the posterior distributions of the predictive coefficients of the short rate and term spread in the 3 in-sample regimes. It can be seen that the posterior distributions of these coefficients have undergone noticeable changes during the sample period. In contrast, the no-break VAR estimates show that both the short rate and term spread are significant over the whole sample period. Note that most of the studies of predicting output growth involving short rates and term spreads found that the short rate has little marginal predictive power once spreads are included, e.g. Plosser and Rouwenhorst (1994) and Stock and Watson (2003). In contrast, Ang et al. (2006) found that the short rate has more predictive power for GDP growth than term spreads. Our approach differs from the existing studies by explicitly modeling the possible structural instability of the predictive relation and finds that the relative importance of short rates and term spreads is changing over time. In the most recent regime, the term spread has more predictive power for output growth than the short rate.

4.2   Performance of Out-of-Sample Forecasts

To assess the usefulness of incorporating structural breaks in the predictive relation of the yield curve and output growth, a recursive out-of-sample forecasting exercise is conducted to compare the performance of the structural break VAR model with the no-break VAR model, that is, at each point of time, only historically available information is used to estimate the models and make forecasts. We focus on the forecast of output growth as it is the major interest of this paper. Forecasts of all 3 variables from the VAR models can be done in a similar way. We use the same diffuse priors as in the in-sample study

17

of Section 4.1. The estimates are based on 80,000 posterior draws after discarding 5,000 initial draws.

We consider 4 forecasting horizons: 1-month-ahead, 3-months-ahead, 6-months-ahead and 12-months-ahead, which are commonly used in practice. In forecasting, we assume that there are at most 2 out-of-sample breaks during these forecasting horizons. This assumption is plausible since an out-of-sample regime shorter than the considered forecasting horizons seems unlikely based on historical estimates. As a practical matter, scenarios of more than 2 out-of-sample breaks have numerically negligible probabilities as estimates of the transition probabilities π_k are uniformly close to 1.

The log predictive Bayes factor for output growth is used to assess the models' forecasting performances. Given a sample of data y_1, y_2,...,y_T and the starting forecast date τ, $\tau \leq T - h$, the log cumulative predictive likelihood of output growth is

$$\sum_{t=\tau}^{T-h} log\left(p(g_{t+h}|Y_t)\right) \tag{13}$$

where h is the forecast horizon. Note that the starting forecast date $\tau \leq T-h$ since $T-h$ is the last point in which a h-period ahead forecast can be evaluated given T data points. For two competing models M_1 and M_2, the difference in their log cumulative predictive likelihoods of output growth

$$\sum_{t=\tau}^{T-h} log\left(\frac{p(g_{t+h}|Y_t; M_1)}{p(g_{t+h}|Y_t; M_2)}\right) \tag{14}$$

is the log predictive Bayes factor of output growth. Model M_1 is favored by the data if the log predictive Bayes factor is positive. This measure keeps the cumulative record of out-of-sample forecasting performance of the models and is the central quantity of interest for Bayesian model comparison (Geweke and Whiteman (2005)). As can be seen from Section 3, the predictive distribution of output growth $p(g_{t+h}|Y_t)$ under the structural break VAR model is a mixture of normal distributions with different means and variances and hence is likely to be highly non-Gaussian. Compared with traditional measures such as the Root Mean Squared Error (RMSE), the predictive Bayes factor can provide a more complete view of the forecasting performance in such cases.

The model averaging approach of Equation (3) is used to compute the predictive likelihoods of output growth $p(g_{t+h}|Y_t)$ according to the formulas in Section 3, which integrate out the uncertainty about both in-sample and out-of-sample breaks and parameters. For each forecasting horizon h, the component models of Equation (3) are the full-break VAR models with $K = 1$ to 5 in-sample regimes. For each component model with K in-sample regimes, we compute the posterior probability $p(s_t = K|Y_t)$ and the predictive density $p(g_{t+h}|Y_t, s_t = K)$ which integrates over all possible scenarios of out-of-sample breaks according to Equations (6) and (12). The sum of the product of these predictive densities $p(g_{t+h}|Y_t, s_t = K)$ and posterior probabilities $p(s_t = K|Y_t)$ is computed as in Equation (3) and is recorded as the predictive likelihood of output growth $p(g_{t+h}|Y_t)$. This procedure is repeated recursively from January 2002 to December 2006 for a 5-year period. Similarly we compute the predictive likelihoods of output growth $p(g_{t+h}|Y_t)$ recursively for the no-break VAR model. The log predictive Bayes factor of output growth is then computed based on the predictive likelihoods according to Equations (13) and (14).

Figure 5 plots the log predictive Bayes factors of the full-break VAR model against the no-break VAR model. It can be seen that the full-break VAR model outperforms the no-break VAR model in all of the 4 forecasting horizons, despite the fact that the no-break model is more parsimonious. The predictive Bayes factors increase steadily throughout the forecasting period[2]. At the end of the 5-year forecasting period, the log predictive Bayes factors are more than 4.5 for all of the horizons, which translate into cumulative predictive likelihoods of output growth from the structural break VAR model more than 90 times higher than those from the no-break VAR model. This forecasting exercise illustrates the precision gain of moving from a no-break VAR model to the structural break VAR representation.

---

[2]There is a drop in the predictive Bayes factors at the forecast of output growth in Sep 2005. This can be explained by a sharp change of output growth around this period. The output growth rate is -1.64% in August and jumps to 1.14% in September 2005, which are the lowest and highest output growth rates during the forecasting period from Jan 2002 to Dec 2006. The average output growth rate during the forecasting period is only 0.21%. Nevertheless forecasts of the break VAR model pick up momentum shortly after this period.

We experiment with imposing informative priors on the structural break VAR model since in the Bayesian literature of VAR forecasting, the use of informative priors has been frequently advocated, e.g. Kadiyala and Karlsson (1997). One popular choice is the Litterman prior (Litterman (1980,1986)). Kadiyala et al. (1997) specifies the Litterman prior as a Gaussian prior distribution for parameters in the intercept and slope matrix of VAR models. The mean of the prior is 1 for diagonal elements of the slope matrix and equals 0 for the intercept and off-diagonal elements of the slope matrix. The variances of the prior recommended in Kadiyala et al. (1997) are set as: $10^5 \sigma_i^2$ for the i-th element in the intercept, 0.05 for the diagonal elements and $\frac{0.005 \sigma_i^2}{\sigma_j^2}$ for the off-diagonal (i,j) element of the slope matrix, where σ_i is the residue standard error of an autoregression for variable i. The covariances in the prior are set to be 0.

In the context of the structural break VAR model, the relevant hyper-parameters in the hierarchical priors are a_0, A_0, d_0 and D_0 as defined in Section 2. We set a_0 to be the mean of the Litterman prior[3], $A_0 = 0.1 I_{m(m+1)}$, $d_0 = m(m+1) + 50$ and D_0 the covariance matrix of the Litterman prior scaled by $d_0 - m(m+1) - 1$, where $m = 3$ is the number of variables in the VAR. These values of hyper-parameters calibrate the distribution for parameters in the intercepts and slope matrices of the structural break VAR model to center around the Litterman priors with small variation. Table 7 reports the log cumulative predictive likelihood from Jan 2002 to Dec 2006 for the structural break VAR model with this informative prior along with those from the structural break VAR model with diffuse prior and no-break VAR model. Among the structural break VAR models, the informative prior produces better 1-month-ahead forecasts while being outperformed in 6-months-ahead forecasts by the diffuse prior. For the horizons 3 and 12 months, performances of the two priors are close. Nevertheless both produce superior forecasts than the no-break VAR model.

---

[3]For the diagonal element in the slope matrix of VAR corresponding to output growth, the mean is set to be 0.3 instead of 1 since, unlike the short rate and term spread which are highly persistent, the 1st order autocorrelation of output growth is about 0.3

5 Conclusion

In this paper, we present a new structural break VAR model for predicting real output growth by the nominal yield curve which allows for the possibility of both in-sample and out-of-sample breaks in parameter values. We jointly model structural breaks in the output growth and yield curve. A Bayesian estimation approach is provided which extends the method of Pesaran et al. (2006) for univariate models to multivariate settings. We provide a discussion of the prior on the number of in-sample regimes implied by the hierarchical structure of this type of model. This is in contrast to Pesaran et al. (2006) which imposes a uniform prior on the number of in-sample regimes and hence is inconsistent with hierarchical priors on other model parameters. A new forecasting method is proposed which guarantees dynamic consistency when forecasting output growth recursively in real time. The empirical application focuses on the monthly US data from 1964 to 2006. We find strong evidence of structural breaks in the predictive relation between the yield curve and output growth in late 1979 and early 1983. Before 1979, the short rate has more predictive power for output growth than the term spread while the term spread becomes more significant since the break of 1983. In the forecasting exercise, we find that incorporating the possibility of structural breaks produces more accurate out-of-sample forecasts of output growth than the no-break VAR model at all horizons considered in this paper.

Appendix A: Gibbs Sampler for the Structural Break VAR Model

For a given number of in-sample regimes K, let $\Theta = (\phi_1, ..., \phi_K, \Sigma_1, ..., \Sigma_K, \pi_1, ..., \pi_{K-1})$, $\Theta_0 = (b_0, B_0, \Omega_0, v_0, \alpha_0, \beta_0)$, $S = (s_1, s_2, ..., s_T)$ and $Y_T = (y_1, ..., y_T)$. The posterior distribution of interest is $p(S, \Theta, \Theta_0 | Y_T)$. A Gibbs sampler is used to sample from this posterior distribution. Detailed discussions on Markov chain Monte Carlo methods of which Gibbs sampling is a special case can be found in Chib (2001), Koop (2003) and Geweke (2005).

1. Sample S from $p(S|Y_T, \Theta, \Theta_0)$.

Sampling the latent states S follows Chib (1998). The steps are as follows:

(1). Compute the filtering density $p(s_t|Y_t, \Theta)$ for $t = 1, 2, ..., T$ by the Hamilton filter (Hamilton (1989)). This involves repeatedly applying a prediction and a filtering step.

$$p(s_t = k|Y_{t-1}, \Theta) = p(s_{t-1} = k|Y_{t-1}, \Theta)\pi_k + p(s_{t-1} = k - 1|Y_{t-1}, \Theta)(1 - \pi_{k-1})$$

$$p(s_t = k|Y_t, \Theta) \propto p(s_t = k|Y_{t-1}, \Theta)p(y_t|Y_{t-1}, s_t = k, \Theta)$$

for $k = 1, 2, ..., K$, where the likelihood $p(y_t|Y_{t-1}, s_t = k, \Theta) = N(y_t|\mu_k + \Phi_k y_{t-1}, \Sigma_k)$. The filter starts at $p(s_t = 1|Y_0, \Theta) = 1$.

(2). Set $s_T = K$.

(3). Given $s_{t+1} = k$, sample s_t as

$$s_t = \begin{cases} \text{k,} & \text{with probability } c_t; \\ \text{k-1,} & \text{with probability } 1 - c_t. \end{cases}$$

where $c_t \propto p(s_t = k|Y_t, \Theta)p(s_{t+1} = k|s_t = k, \Theta)$. Note $s_1 = 1$.

2. Sample π_k from $p(\pi_k|Y_T, \Theta_{-\pi_k}, \Theta_0, S)$ for $k = 1, ..., K - 1$, where $\Theta_{-\pi_k}$ is the subset of Θ excluding the parameter π_k.

$$\pi_k \sim Beta(\alpha_0 + n_{kk}, \beta_0 + 1)$$

where n_{kk} is the number of one-step transition from state k to state k in the sequence S.

3. Sample ϕ_k from $p(\phi_k|Y_T, \Theta_{-\phi_k}, \Theta_0, S)$ and Σ_k from $p(\Sigma_k|Y_T, \Theta_{-\Sigma_k}, \Theta_0, S)$, $k = 1, 2, ..., K$, where $\Theta_{-\phi_k}$ and $\Theta_{-\Sigma_k}$ are the subsets of Θ excluding the parameter ϕ_k and Σ_k respectively.

The conditional posterior densities of ϕ_k and Σ_k depend only on the data in regime k. Therefore, let $\{y_t : t = n_{k-1} + 1, n_{k-1} + 2, ..., n_k\}$ be the data in regime k, $\widehat{Y}_k$ be a $(n_k - n_{k-1}) \times 3$ matrix of data with each row as an observation y_t, $\widehat{y}_k = vec(\widehat{Y}_k)$ be a vector stacking the columns of $\widehat{Y}_k$. Let $\widehat{x}_k$ be a $(n_k - n_{k-1}) \times 3$ matrix stacking observations $y_{n_{k-1}}, y_{n_{k-1}+1}, ..., y_{n_k-1}$, $\widehat{X}_k$ be a $(n_k - n_{k-1}) \times 4$ matrix concatenating a $(n_k - n_{k-1}) \times 1$ vector of 1's horizontally with $\widehat{x}_k$ and $\widehat{Z}_k = I_3 \otimes \widehat{X}_k$, where $\otimes$ denotes the Kronecker product. The conditional posterior densities of ϕ_k and Σ_k are

$$\phi_k \sim N(b_k, B_k)$$

$$\Sigma_k \sim IW(\Omega_k, v_k)$$

where

$$B_k = \left(\widehat{Z}'_k (\Sigma_k^{-1} \otimes I_{n_k - n_{k-1}}) \widehat{Z}_k + B_0^{-1} \right)^{-1}$$

$$b_k = B_k \left(\widehat{Z}'_k (\Sigma_k^{-1} \otimes I_{n_k - n_{k-1}}) \widehat{y}_k + B_0^{-1} b_0 \right)$$

$$\Omega_k = \Omega_0 + \sum_{t=n_{k-1}+1}^{n_k} (y_t - \mu_k - \Phi_k y_{t-1})(y_t - \mu_k - \Phi_k y_{t-1})'$$

$$v_k = v_0 + (n_k - n_{k-1})$$

4. Sample b_0 from $p(b_0|Y_T, \Theta, \Theta_{0\_b_0}, S)$, where $\Theta_{0\_b_0}$ denotes the subset of Θ_0 excluding the parameter b_0.

$$b_0 \sim N(a, A)$$

where

$$A = \left(K B_0^{-1} + A_0^{-1} \right)^{-1}$$

$$a = A \left(B_0^{-1} \sum_{k=1}^{K} \phi_k + A_0^{-1} a_0 \right)$$

5. Sample B_0 from $p(b_0|Y_T, \Theta, \Theta_{0\_B_0}, S)$, where $\Theta_{0\_B_0}$ denotes the subset of Θ_0 excluding the parameter B_0.

$$B_0 \sim IW \left(D_0 + \sum_{k=1}^{K} (\phi_k - b_0)(\phi_k - b_0)', d_0 + K \right)$$

6. Sample Ω_0 from $p(\Omega_0|Y_T, \Theta, \Theta_{0\_\Omega_0}, S)$, where $\Theta_{0\_\Omega_0}$ denotes the subset of Θ_0 excluding the parameter Ω_0, (Metropolis step).

Since the conditional posterior density of Ω_0 is non-standard, a Metropolis-Hastings step is introduced to sample Ω_0. At iteration j, the proposal distribution for $\Omega_0^{(j)}$ is $IW\left((c_1 - m - 1)\Omega_0^{(j-1)}, c_1 \right)$, which calibrates the mean of the proposal at $\Omega_0^{(j-1)}$. The free parameter c_1 controls the variation of the proposal: a bigger value of c_1 implies smaller variation and hence higher acceptance rate of the Metropolis step. m is the dimension of y_t. Note the prior density is $p(\Omega_0) = IW(\Omega_0|\Psi_0, f_0)$ and the likelihood function is

$p(\Sigma_1, ..., \Sigma_K | \Omega_0, v_0) = \prod_{k=1}^{K} IW(\Sigma_k | \Omega_0, v_0)$. For any candidate draw Ω_0^* from the proposal distribution, the resulting acceptance probability is given by

$$\zeta_1 = min \left\{ 1, \frac{IW(\Omega_0^* | \Psi_0, f_0) \prod_{k=1}^{K} IW(\Sigma_k | \Omega_0^*, v_0) IW\left(\Omega_0^{(j-1)} | (c_1 - m - 1)\Omega_0^*, c_1\right)}{IW(\Omega_0^{(j-1)} | \Psi_0, f_0) \prod_{k=1}^{K} IW(\Sigma_k | \Omega_0^{(j-1)}, v_0) IW\left(\Omega_0^* | (c_1 - m - 1)\Omega_0^{(j-1)}, c_1\right)} \right\}$$

With probability ζ_1, the candidate draw Ω_0^* is accepted and $\Omega_0^{(j)} = \Omega_0^*$. Otherwise, $\Omega_0^{(j)} = \Omega_0^{(j-1)}$.

7. Sample v_0 from $p(v_0 | Y_T, \Theta, \Theta_{0\_v_0}, S)$, where $\Theta_{0\_v_0}$ denotes the subset of Θ_0 excluding the parameter v_0, (Metropolis step).

A Metropolis-Hastings step is used to sample v_0. At iteration j, the proposal distribution for $v_0^{(j)}$ is $Gamma\left(v_0^{(j-1)}/c_2, c_2\right)$, which calibrates the mean of the proposal at $v_0^{(j-1)}$. The free parameter c_2 controls the variation of the proposal: a bigger value of c_2 implies larger variation and hence lower acceptance rate of the Metropolis step. Note the prior density is $p(v_0) = Gamma(v_0 | \rho_0, \lambda_0)$ and the likelihood function is $p(\Sigma_1, ..., \Sigma_K | \Omega_0, v_0) = \prod_{k=1}^{K} IW(\Sigma_k | \Omega_0, v_0)$. For any candidate draw v_0^* from the proposal distribution, the resulting acceptance probability is given by

$$\zeta_2 = min \left\{ 1, \frac{Gamma(v_0^* | \rho_0, \lambda_0) \prod_{k=1}^{K} IW(\Sigma_k | \Omega_0, v_0^*) Gamma\left(v_0^{(j-1)} | v_0^*/c_2, c_2\right)}{Gamma(v_0^{(j-1)} | \rho_0, \lambda_0) \prod_{k=1}^{K} IW(\Sigma_k | \Omega_0, v_0^{(j-1)}) Gamma\left(v_0^* | v_0^{(j-1)}/c_2, c_2\right)} \right\}$$

With probability ζ_2, the proposed draw v_0^* is accepted and $v_0^{(j)} = v_0^*$. Otherwise, $v_0^{(j)} = v_0^{(j-1)}$.

8. Sample α_0 from $p(\alpha_0 | Y_T, \Theta, \Theta_{0\_\alpha_0}, S)$, where $\Theta_{0\_\alpha_0}$ denotes the subset of Θ_0 excluding the parameter α_0, (Metropolis step).

A Metropolis-Hastings step is used to sample α_0. At iteration j, the proposal distribution for $\alpha_0^{(j)}$ is $Gamma\left(\alpha_0^{(j-1)}/c_3, c_3\right)$, which calibrates the mean of the proposal at $\alpha_0^{(j-1)}$. The free parameter c_3 controls the variation of the proposal: a bigger value of c_3 implies larger variation and hence lower acceptance rate of the Metropolis step. Note the prior density is $p(\alpha_0) = Gamma(\alpha_0 | q_0, \gamma_0)$ and the likelihood function is $p(\pi_1, ..., \pi_{K-1} | \alpha_0, \beta_0) = \prod_{k=1}^{K-1} Beta(\pi_k | \alpha_0, \beta_0)$. For any candidate draw α_0^* from the proposal distribution, the re-

sulting acceptance probability is given by

$$\zeta_3 = min \left\{ 1, \frac{Gamma(\alpha_0^*|q_0, \gamma_0) \prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0^*, \beta_0) Gamma\left(\alpha_0^{(j-1)}|\alpha_0^*/c_3, c_3\right)}{Gamma(\alpha_0^{(j-1)}|q_0, \gamma_0) \prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0^{(j-1)}, \beta_0) Gamma\left(\alpha_0^*|\alpha_0^{(j-1)}/c_3, c_3\right)} \right\}$$

With probability ζ_3, the proposed draw α_0^* is accepted and $\alpha_0^{(j)} = \alpha_0^*$. Otherwise, $\alpha_0^{(j)} = \alpha_0^{(j-1)}$.

9. Sample β_0 from $p(\beta_0|Y_T, \Theta, \Theta_{0\_\beta_0}, S)$, where $\Theta_{0\_\beta_0}$ denotes the subset of Θ_0 excluding the parameter β_0, (Metropolis step).

A Metropolis-Hastings step is used to sample β_0. At iteration j, the proposal distribution for $\beta_0^{(j)}$ is $Gamma\left(\beta_0^{(j-1)}/c_4, c_4\right)$, which calibrates the mean of the proposal at $\beta_0^{(j-1)}$. The free parameter c_4 controls the variation of the proposal: a bigger value of c_4 implies larger variation and hence lower acceptance rate of the Metropolis step. Note the prior density is $p(\beta_0) = Gamma(\beta_0|r_0, \delta_0)$ and the likelihood function is $p(\pi_1, ..., \pi_{K-1}|\alpha_0, \beta_0) = \prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0, \beta_0)$. For any candidate draw β_0^* from the proposal distribution, the resulting acceptance probability is given by

$$\zeta_4 = min \left\{ 1, \frac{Gamma(\beta_0^*|r_0, \delta_0) \prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0, \beta_0^*) Gamma\left(\beta_0^{(j-1)}|\beta_0^*/c_4, c_4\right)}{Gamma(\beta_0^{(j-1)}|r_0, \delta_0) \prod_{k=1}^{K-1} Beta(\pi_k|\alpha_0, \beta_0^{(j-1)}) Gamma\left(\beta_0^*|\beta_0^{(j-1)}/c_4, c_4\right)} \right\}$$

With probability ζ_4, the proposed draw β_0^* is accepted and $\beta_0^{(j)} = \beta_0^*$. Otherwise, $\beta_0^{(j)} = \beta_0^{(j-1)}$.

This completes the algorithm for sampling from the structural break VAR model.

Appendix B: Marginal Likelihood of the Structural Break VAR Model

In this paper, we adopt the modified harmonic mean (MHM) method of Gelfand and Dey (1994) to compute the marginal likelihood. The basic idea of the MHM method is to utilize the simple identity

$$\frac{1}{p(Y_t|s_t = K)} = \frac{p(\Theta, \Theta_0|Y_t, S_t = K)}{p(\Theta, \Theta_0|S_t = K)p(Y_t|\Theta, \Theta_0, S_t = K)}$$

Note the priors of Θ and Θ_0 are independent of the number of in-sample regimes. So $p(\Theta, \Theta_0 | S_t = K) = p(\Theta_0)p(\Theta|\Theta_0)$. Also the likelihood of Y_t does not depend on Θ_0 given the value of Θ. So $p(Y_t|\Theta, \Theta_0, S_t = K) = p(Y_t|\Theta, S_t = K)$.

For any density function $h(\Theta, \Theta_0)$ whose support is contained in that of $p(\Theta, \Theta_0|Y_t, S_t = K)$, it is easy to show that

$$\frac{1}{p(Y_t|s_t = K)} = \int \frac{h(\Theta, \Theta_0)}{p(\Theta_0)p(\Theta|\Theta_0)p(Y_t|\Theta, S_t = K)} p(\Theta, \Theta_0|Y_t, S_t = K)d\Theta d\Theta_0$$

Given a sample of posterior draws $\{\Theta^{(j)}, \Theta_0^{(j)}\}_{j=1}^n$ from $p(\Theta, \Theta_0|Y_t, S_t = K)$, the log marginal likelihood can be computed as

$$log\left(p(Y_t|s_t = K)\right) \approx -log\left(\frac{1}{n}\sum_{j=1}^n \frac{h(\Theta^{(j)}, \Theta_0^{(j)})}{p(\Theta_0^{(j)})p(\Theta^{(j)}|\Theta_0^{(j)})p(Y_t|\Theta^{(j)}, S_t = K)}\right)$$

The likelihood function $p(Y_t|\Theta, S_t = K)$ can be obtained as a by-product from the Hamilton filter in the estimation process.

One condition for this numerical integration is that the function $\frac{h(\Theta, \Theta_0)}{p(\Theta_0)p(\Theta|\Theta_0)p(Y_t|\Theta, S_t = K)}$ needs to be bounded above for the rate of convergence to be practical (Geweke(1999)). Geweke (1999) proposes a convenient implementation of $h(\cdot)$ which satisfies the above condition. The function $h(\cdot)$ is chosen to be a truncated Gaussian density with the mean and covariance matrix constructed from posterior draws of $p(\Theta, \Theta_0|Y_t, S_t = K)$. Specifically, let $\widetilde{\Theta} = \{\Theta, \Theta_0\}$, $\overline{\Theta} = \frac{1}{n}\sum_{j=1}^n \widetilde{\Theta}^{(j)}$ and $\overline{\Omega} = \frac{1}{n}\sum_{j=1}^n \left(\widetilde{\Theta}^{(j)} - \overline{\Theta}\right)\left(\widetilde{\Theta}^{(j)} - \overline{\Theta}\right)'$. Define $\widehat{\Theta} \equiv \left\{\widetilde{\Theta} : \left(\widetilde{\Theta} - \overline{\Theta}\right)'\overline{\Omega}^{-1}\left(\widetilde{\Theta} - \overline{\Theta}\right) \leq \chi_p^2(m)\right\}$, where m is the dimension of $\widetilde{\Theta}$ and $\chi_p^2(m)$ is the $p \times 100$ percent critical value of a χ^2 distribution with m degrees of freedom. The implementation of $h(\cdot)$ is given by

$$h\left(\widetilde{\Theta}\right) = \frac{1}{p(2\pi)^{0.5m}\left|\overline{\Omega}\right|^{0.5}}exp\left(-0.5\left(\widetilde{\Theta} - \overline{\Theta}\right)'\overline{\Omega}^{-1}\left(\widetilde{\Theta} - \overline{\Theta}\right)\right)I_{\widehat{\Theta}}$$

where

$$I_{\widehat{\Theta}} = \begin{cases} 1, & \text{if } \widetilde{\Theta} \in \widehat{\Theta}; \\ 0, & \text{otherwise.} \end{cases}$$

A smaller value of the truncation probability p will likely result in better behavior of the numerical integration since more tail draws are discarded. But greater simulation error may occur as fewer draws are retained in the set $\widehat{\Theta}$. In practice, it is usually chosen to be in the range $(0.9, 1)$.

References

ANG, A., AND G. BEKAERT (2002): "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20, 163–182.

ANG, A., M. PIAZZESI, AND M. WEI (2006): "What Does the Yield Curve Tell Us About GDP Growth?," *Journal of Econometrics*, 131, 359–403.

BANSAL, R., AND H. ZHOU (2002): "Term Structure of Interest Rates with Regime Shifts," *Journal of Finance*, 57, 1997–2043.

CHEN, N. (1991): "Financial Investment Opportunities and the Macroeconomy," *Journal of Finance*, 46, 529–554.

CHIB, S. (1995): "Marginal Likelihood from the Gibbs Sampler," *Journal of the American Statistical Association*, 90, 1313–1321.

CHIB, S. (1998): "Estimation and Comparison of Multiple Change Point Models," *Journal of Econometrics*, 86, 221–241.

——— (2001): "Markov Chain Monte Carlo Methods: Computation and Inference," in *Handbook of Econometrics*, ed. by Heckman, and Leamer. Elsevier Science.

CHIB, S., AND I. JELIAZKOV (2001): "Marginal Likelihood from the Metropolis-Hasting Output," *Journal of the American Statistical Association*, 96, 270–281.

DAVIS, P., AND G. FAGAN (1997): "Are Financial Spreads Useful Indicators of Future Inflation and Output Growth in EU Countries?," *Journal of Applied Econometrics*, 12, 701–714.

ESTRELLA, A., AND G. HARDOUVELIS (1991): "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance*, 46, 555–576.

ESTRELLA, A., AND F. MISHKIN (1997): "The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank," *European Economic Review*, 41, 1375–1401.

——— (1998): "Predicting U.S. Recessions: Financial Variables as Leading Indicators," *Review of Economics and Statistics*, 80, 45–61.

ESTRELLA, A., A. RODRIGUES, AND S. SCHICH (2003): "How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States," *Review of Economics and Statistics*, 85, 629–644.

FAMA, E., AND R. BLISS (1987): "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

FRUHWIRTH-SCHNATTER, S. (1995): "Bayesian Model Discrimination and Bayes Factor for Linear Gaussian State Space Models," *Journal of the Royal Statistical Society, Series B*, 57, 237–246.

——— (2004): "Estimating Marginal Likelihoods for Mixture and Markov Switching Models Using Bridge Sampling Techniques," *The Econometrics Journal*, 7, 143–167.

GELFAND, A., AND D. DEY (1994): "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of The Royal Statistical Society*, B, 56, 501–514.

GEWEKE, J. (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley and Sons Ltd.

GEWEKE, J., AND C. WHITEMAN (2005): "Bayesian Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, vol. forthcoming.

GIACOMINI, R., AND B. ROSSI (2006): "How Stable is the Forecasting Performance of the Yield Curve for Output Growth?," *Oxford Bulletin of Economics and Statistics*, 68, 783–795.

GIORDANI, P., AND R. KOHN (2006): "Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models," Sveriges Riksbank Working Paper 196.

GRAY, S. F. (1996): "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.

HAMILTON, J., AND D. KIM (2002): "A Re-Examination of the Predictability of Economic Activity using the Yield Spread," *Journal of Money, Credit and Banking*, 34, 340–360.

HAMILTON, J. D. (1988): "Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates," *Journal of Economic Dynamics and Control*, 12, 385–423.

HARVEY, C. R. (1989): "The Real Term Structure and Consumption Growth," *Journal of Financial Economics*, 22, 305–333.

HE, Z., AND J. MAHEU (2008): "Real Time Detection of Structural Breaks in GARCH Models," Working Paper 336, Department of Economics, University of Toronto.

KADIYALA, R., AND S. KARLSSON (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR Models," *Journal of Applied Econometrics*, 12, 99–132.

KIM, C., J. MORLEY, AND C. NELSON (2005): "The structural breaks in the equity premium," *Journal of Business and Economic Statistics*, 23, 181–191.

KOOP, G. (2003): *Bayesian Econometrics*. Wiley, Chichester, England.

KOOP, G., AND S. POTTER (2007): "Estimation and Forecasting in Models with Multiple Breaks," *Review of Economic Studies*, 74, 763–789.

LAURENT, R. (1988): "An Interest Rate-Based Indiator of Monetary Policy," *Federal Reserve Bank of Chicago Economic Perspectives*, 12, 3–14.

——— (1989): "Testing the Spread," *Federal Reserve Bank of Chicago Economic Perspectives*, 13, 22–34.

LITTERMAN, R. (1980): "A Bayesian Procedure for Forecasting with Vector Autoregression," Mimeo, Massachusetts Institute of Technology.

——— (1986): "Forecasting with Bayesian Vector Autoregression: Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25–38.

LIU, C., AND J. MAHEU (2008): "Are There Structural Breaks in Realized Volatility ?," *Journal of Financial Econometrics*, 6, 326–360.

LUTKEPOHL, H. (2006): *New Introduction to Multiple Time Series Analysis*. Springer.

MAHEU, J., AND T. MCCURDY (2007): "How Useful are Historical Data for Forecasting the Long-Run Equity Return Distribution ?," *Journal of Business and Economic Statistics*, forthcoming.

MAHEU, J. M., AND S. GORDON (2008): "Learning, Forecasting and Structural Breaks," *Journal of Applied Econometrics*, 23, 553–583.

MCCULLOCH, R., AND R. TSAY (1993): "Bayesian inference and prediction for mean and variance shifts in autoregressive time series," *Journal of the American Statistical Association*, 88, 968–978.

MENG, X., AND W. WONG (1996): "Simulating Ratios of Normalizing Constants via a Simple Identity," *Statistical Sinica*, 6, 831–860.

MIAZHYNSKAIA, T., AND G. DORFFNER (2006): "A Comparison of Bayesian Model Selection Based on MCMC with an Application to GARCH-Type Models," *Statistical Papers*, 47, 525–549.

NEWTON, M. A., AND A. RAFTERY (1994): "Approximate Bayesian inference by the weighted likelihood bootstrap (with Discussion).," *Journal of The Royal Statistical Society*, B, 56, 3–48.

PASTOR, L., AND R. F. STAMBAUGH (2001): "The Equity Premium and Structural Breaks," *Journal of Finance*, 4, 1207–1231.

PESARAN, H., D. PETTENUZZO, AND A. TIMMERMANN (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *Review of Economic Studies*, 73, 1057–1084.

PLOSSER, C., AND G. ROUWENHORST (1994): "International Term Structure and Real Economic Growth," *Journal of Monetary Economics*, 33, 133–156.

SCOTT, S. L. (2002): "Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century," *Journal of the American Statistical Association*, 97(457), 337–351.

STOCK, J., AND M. WATSON (1989): "New Indexes of Coincident and Leading Economic Indicators," in *NBER Macroeconomics Annual*, ed. by O. Blanchard, and S. Fisher, pp. 352–394.

STOCK, J. H., AND M. W. WATSON (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293–335.

STOCK, J. H., AND M. W. WATSON (2003): "Forecasting Output and Inflation: The Role of Asset Prices," *Journal of Economic Literature*, 41(3), 788–829.

Table 1: Distributions of the Number of In-Sample Regimes

	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$
prior $p(s_T = K)$	0.733	0.181	0.054	0.019	0.013
posterior $p(s_T = K|Y_T)$	0	0	0.998	0.002	0

This table reports the prior and posterior distributions of the number of in-sample regimes in the full-break VAR model for the monthly US data from January 1964 through December 2006.

Table 2: Model Comparison by Marginal Likelihoods

	No-Break	1-Break	2-Break	3-Break	4-Break
log ML	-1131.81	-1008.69	-978.27	-983.65	-992.68

This table reports the log marginal likelihoods of the no-break VAR model and the full-break VAR models with 1 to 4 in-sample breaks for the monthly US data from January 1964 through December 2006.

Table 3: Parameter Estimates of the Full-Break VAR Model: Regime 1

$$
\mu_1 =
\begin{pmatrix}
0.567^* \\
(0.106, 1.040) \\
0.039 \\
(-0.259, 0.335) \\
0.114 \\
(-0.127, 0.355)
\end{pmatrix}
$$

$$
\Phi_1 =
\begin{pmatrix}
0.321^* & -0.072^* & 0.084 \\
(0.188, 0.454) & (-0.145, -0.001) & (-0.050, 0.218) \\
0.035 & 0.993^* & 0.042 \\
(-0.044, 0.116) & (0.947, 1.038) & (-0.041, 0.124) \\
-0.004 & -0.012 & 0.917^* \\
(-0.069, 0.060) & (-0.049, 0.025) & (0.850, 0.983)
\end{pmatrix}
$$

$$
\Sigma_1 =
\begin{pmatrix}
0.621^* & & \\
(0.505, 0.762) & & \\
-0.013 & 0.218^* & \\
(-0.068, 0.040) & (0.178, 0.268) & \\
-0.005 & -0.134^* & 0.142^* \\
(-0.049, 0.038) & (-0.169, -0.105) & (0.115, 0.174)
\end{pmatrix}
$$

This table reports the posterior means and 95% credible sets for regime 1 parameters of the full-break VAR model with 2 in-sample breaks $y_t = \mu_{s_t} + \Phi_{s_t} y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma_{s_t})$, where y_t=(output growth, short rate, term spread), for the monthly US data from January 1964 through December 2006. Parameters significant at the 95% level are marked by "*".

Table 4: Parameter Estimates of the Full-Break VAR Model: Regime 2

$$
\mu_2 = \begin{pmatrix}
0.414 \\
(-0.364, 1.233) \\
0.196 \\
(-0.511, 0.959) \\
0.153 \\
(-0.510, 0.823)
\end{pmatrix}
$$

$$
\Phi_2 = \begin{pmatrix}
0.351^* & -0.046 & 0.041 \\
(0.064, 0.632) & (-0.115, 0.021) & (-0.131, 0.210) \\
0.449^* & 0.981^* & 0.093 \\
(0.061, 0.852) & (0.911, 1.049) & (-0.141, 0.337) \\
-0.295^* & -0.008 & 0.832^* \\
(-0.577, -0.025) & (-0.068, 0.049) & (0.663, 0.994)
\end{pmatrix}
$$

$$
\Sigma_2 = \begin{pmatrix}
0.860^* & & \\
(0.526, 1.387) & & \\
0.508^* & 1.949^* & \\
(0.102, 1.091) & (1.121, 3.196) & \\
-0.268 & -1.082^* & 0.865^* \\
(-0.632, 0.003) & (-1.830, -0.604) & (0.526, 1.407)
\end{pmatrix}
$$

This table reports the posterior means and 95% credible sets for regime 2 parameters of the full-break VAR model with 2 in-sample breaks $y_t = \mu_{s_t} + \Phi_{s_t} y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma_{s_t})$, where y_t=(output growth, short rate, term spread), for the monthly US data from January 1964 through December 2006. Parameters significant at the 95% level are marked by "*".

Table 5: Parameter Estimates of the Full-Break VAR Model: Regime 3

$$
\mu_3 = \begin{pmatrix}
0.112 \\
(-0.086, 0.331) \\
-0.001 \\
(-0.010, 0.097) \\
0.091 \\
(-0.023, 0.214)
\end{pmatrix}
$$

$$
\Phi_3 = \begin{pmatrix}
0.115 & 0.003 & 0.075^* \\
(-0.005, 0.233) & (-0.030, 0.032) & (0.006, 0.144) \\
0.077^* & 0.990^* & 0.016 \\
(0.018, 0.135) & (0.976, 1.004) & (-0.018, 0.050) \\
0.056 & -0.007 & 0.942^* \\
(-0.011, 0.123) & (-0.026, 0.010) & (0.903, 0.981)
\end{pmatrix}
$$

$$
\Sigma_3 = \begin{pmatrix}
0.282^* & & \\
(0.238, 0.334) & & \\
0.032^* & 0.071^* & \\
(0.015, 0.050) & (0.056, 0.085) & \\
0.001 & -0.028^* & 0.094^* \\
(-0.019, 0.020) & (-0.039, -0.017) & (0.079, 0.112)
\end{pmatrix}
$$

This table reports the posterior means and 95% credible sets for regime 3 parameters of the full-break VAR model with 2 in-sample breaks $y_t = \mu_{s_t} + \Phi_{s_t} y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma_{s_t})$, where y_t=(output growth, short rate, term spread), for the monthly US data from January 1964 through December 2006. Parameters significant at the 95% level are marked by "*".

Table 6: Parameter Estimates of the No-Break VAR Model

$$
\mu = \begin{pmatrix}
0.307^* \\
(0.115, 0.495) \\
0.021 \\
(-0.127, 0.167) \\
0.068 \\
(-0.049, 0.185)
\end{pmatrix}
$$

$$
\Phi = \begin{pmatrix}
0.294^* & -0.032^* & 0.058^* \\
(0.212, 0.376) & (-0.055, -0.008) & (0.001, 0.117) \\
0.138^* & 0.989^* & 0.013 \\
(0.074, 0.204) & (0.970, 1.008) & (-0.033, 0.059) \\
-0.042 & 0.001 & 0.935^* \\
(-0.094, 0.010) & (-0.014, 0.016) & (0.898, 0.971)
\end{pmatrix}
$$

$$
\Sigma = \begin{pmatrix}
0.458^* & & \\
(0.405, 0.515) & & \\
0.055^* & 0.288^* & \\
(0.024, 0.087) & (0.255, 0.326) & \\
-0.027^* & -0.160^* & 0.180^* \\
(-0.052, -0.002) & (-0.185, -0.137) & (0.160, 0.204)
\end{pmatrix}
$$

This table reports the posterior means and 95% credible sets for parameters of the no-break VAR model $y_t = \mu + \Phi y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, \Sigma)$, where y_t=(output growth, short rate, term spread), for the monthly US data from January 1964 through December 2006. Parameters significant at the 95% level are marked by "*".

Table 7: Comparing Log Cumulative Predictive Likelihoods of Output Growth

Forecasting Horizon	No-Break VAR	Full-Break VAR with Diffuse Prior	Full-Break VAR with Informative Prior
1 month	-55.78	-48.86	-43.72
3 months	-54.83	-46.64	-46.96
6 months	-52.50	-46.42	-47.90
12 months	-45.49	-40.81	-40.57

This table reports log cumulative predictive likelihoods of output growth $\sum_{t=T-60}^{T-h} log(p(g_{t+h}|Y_t))$, $h = 1, 3, 6, 12$, for the no-break VAR model and the full-break VAR models with diffuse and informative priors during the period Jan 2002-Dec 2006.
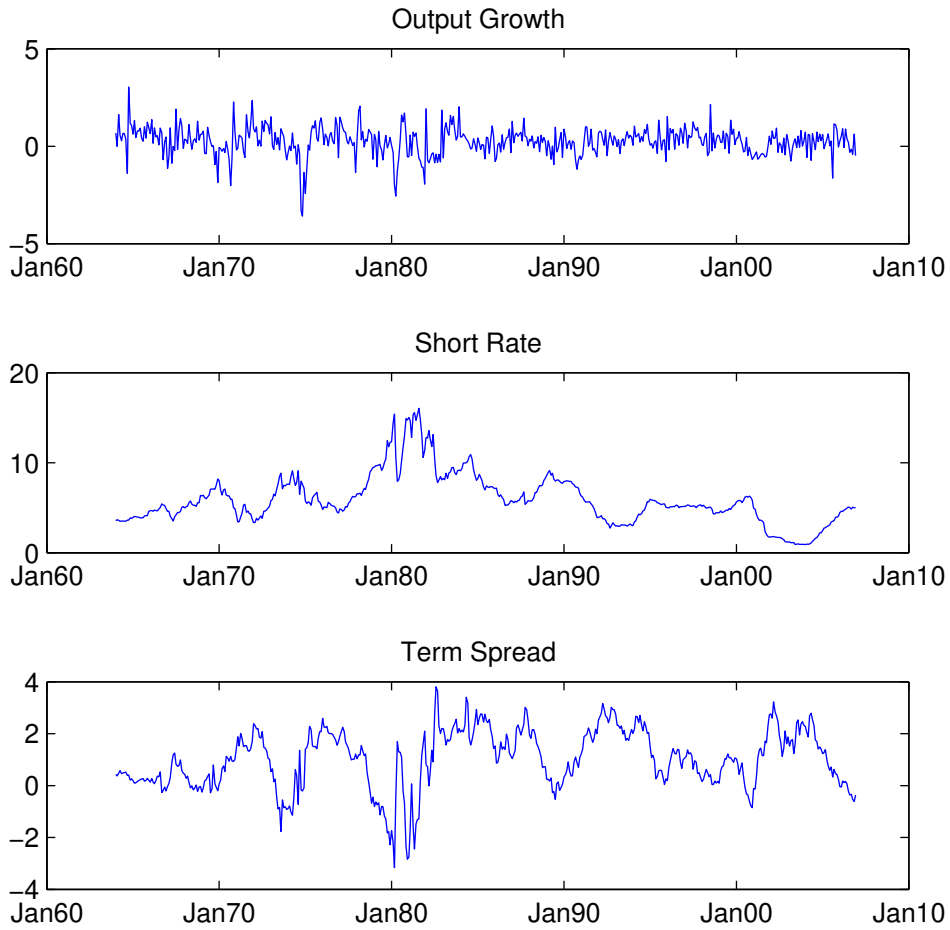
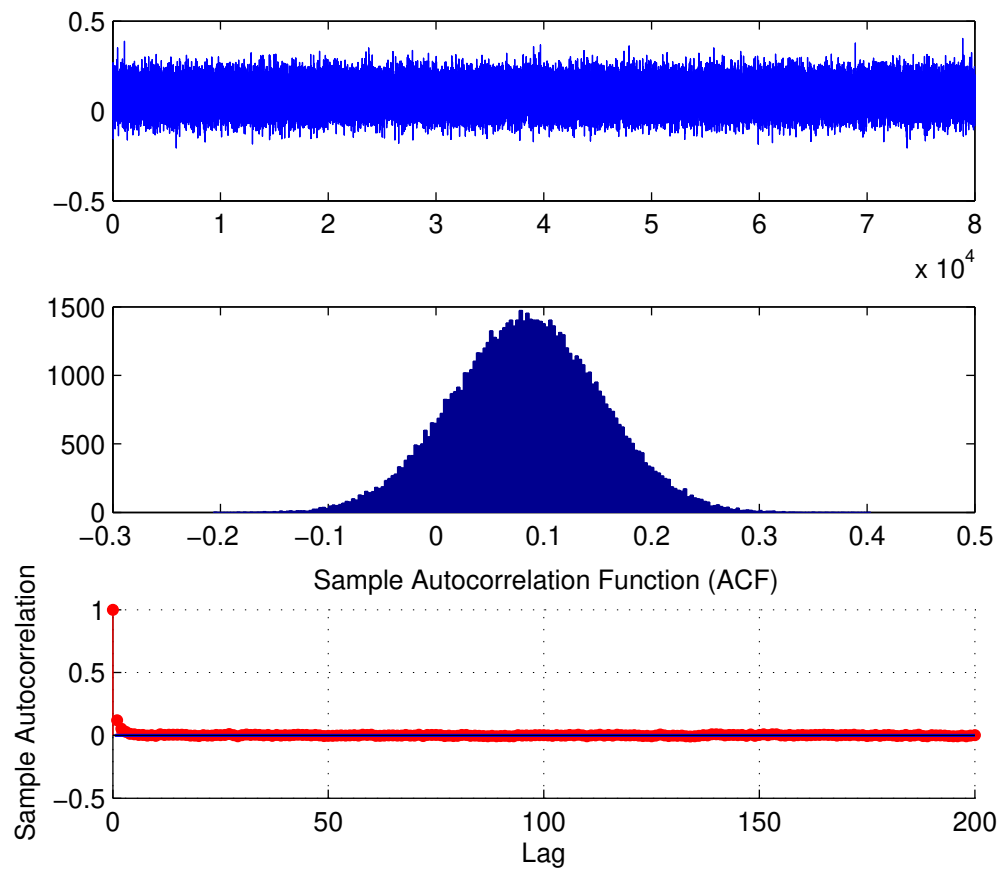Figure 1: The Monthly US Data: January 1964 to December 2006

Figure 2: Posterior Draws of the Predictive Coefficient of Term Spread in Regime 1
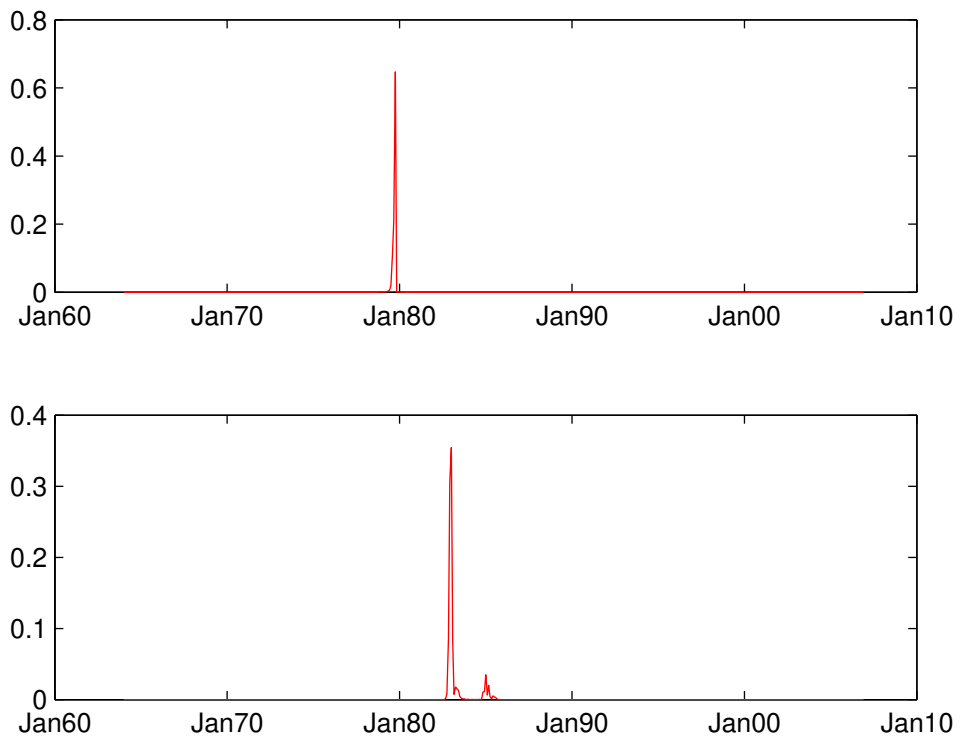
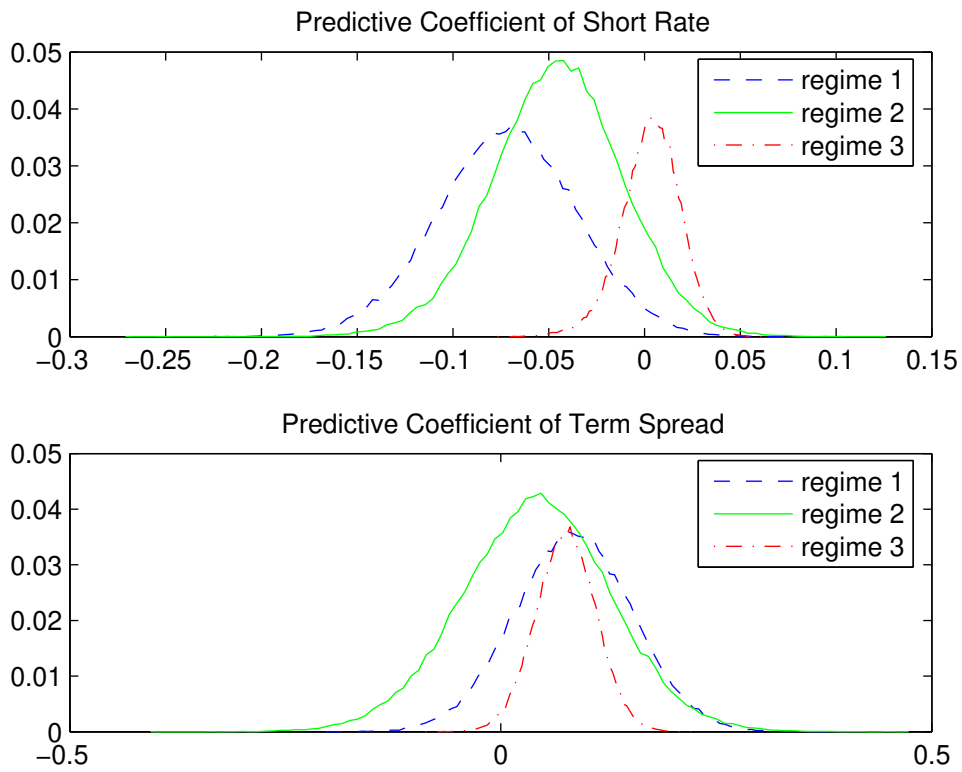Figure 3: Posterior Distributions of the Break Dates

Figure 4: Posterior Distributions of the Predictive Coefficients of Short Rate and Term Spread
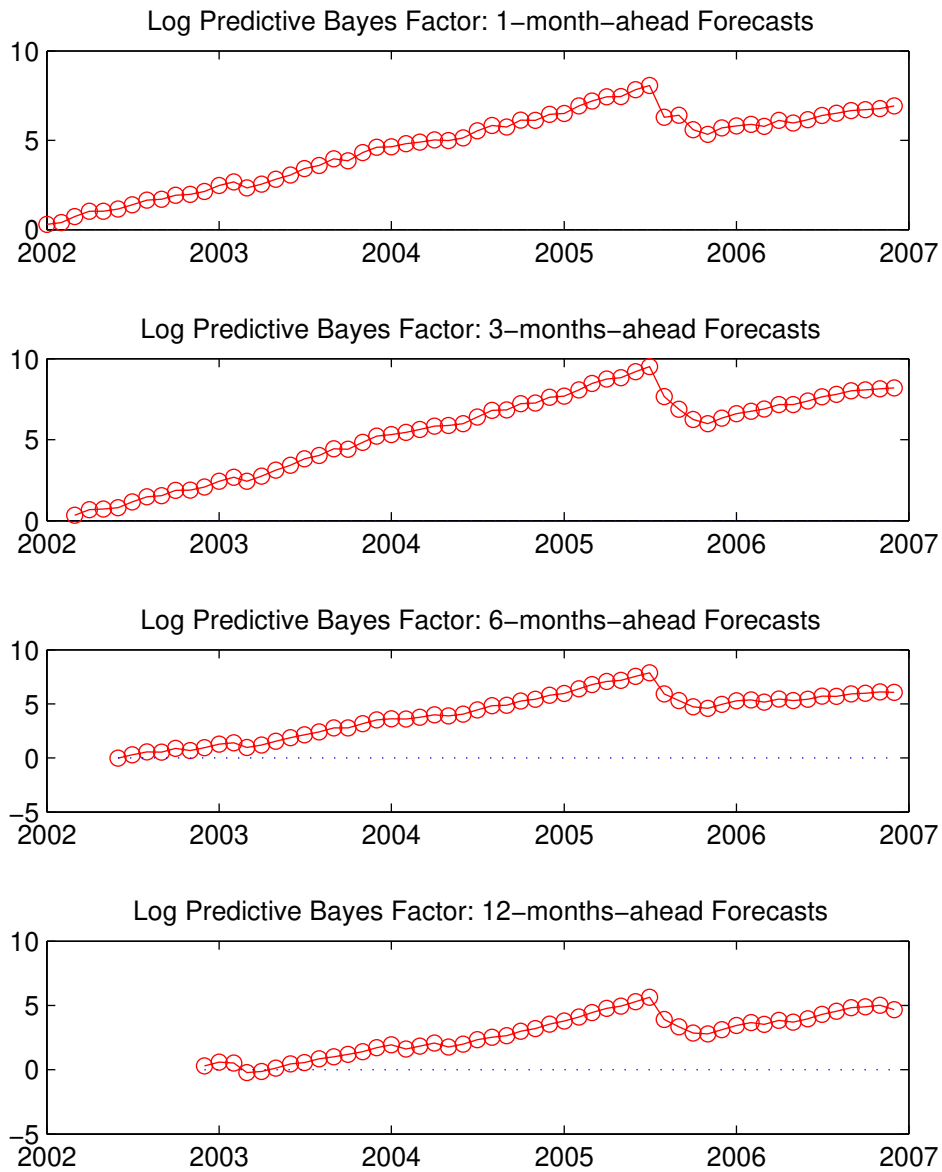
Figure 5: Log Predictive Bayes Factors: Break vs. No Break (Note: A positive value of log predictive Bayes factor favors the full-break VAR model.)