Forecasting Time Series Subject to Multiple Structural Breaks

M. Hashem Pesaran, Davide Pettenuzzo
and Allan Timmermann

June 2004

CWPE 0433

Forecasting Time Series Subject to Multiple Structural Breaks[*]

M. Hashem Pesaran
University of Cambridge and USC

Davide Pettenuzzo
Bocconi University

Allan Timmermann
University of California, San Diego

June 2004

Abstract

This paper provides a novel approach to forecasting time series subject to discrete structural breaks. We propose a Bayesian estimation and prediction procedure that allows for the possibility of new breaks over the forecast horizon, taking account of the size and duration of past breaks (if any) by means of a hierarchical hidden Markov chain model. Predictions are formed by integrating over the hyper parameters from the meta distributions that characterize the stochastic break point process. In an application to US Treasury bill rates, we find that the method leads to better out-of-sample forecasts than alternative methods that ignore breaks, particularly at long horizons.

Keywords: Structural Breaks, Forecasting, Hierarchical hidden Markov Chain Model, Bayesian Model Averaging.
JEL Classifications: C110, C150, C530.

1 Introduction

Structural changes or "breaks" have been observed in many economic and financial time series. In a study of a large set of macroeconomic time series, Stock and Watson (1996) reported that the majority of the series displayed evidence of instability.[1] Such structural breaks pose a formidable challenge to economic forecasting and have led authors such as Clements and Hendry (1998, 1999) to view it as the main source of forecast failure.[2]

A key question that arises in the context of forecasting is how future values of the time-series of interest might be affected by breaks. If breaks occurred in the past, surely they could also happen in the future. For forecasting purposes a model of the stochastic process underlying the breaks is therefore required to address questions such as how many breaks are likely to occur over the forecasting sample, how large such breaks will be and at which dates they occur. Approaches that view breaks as being generated deterministically are not applicable when forecasting future events unless, of course, future break dates as well as the size of such breaks are known in advance. In most applications this is not a plausible assumption and so a need arises to model the stochastic process underlying the breaks.

In this paper we provide a general framework to forecasting time series under structural breaks that is capable of handling the different scenarios that arise once new breaks can occur over the forecast horizon. Allowing for breaks complicates the forecasting problem considerably. To illustrate this, consider the problem of forecasting some variable, y, h periods ahead using a historical data sample $\{y_1, ..., y_T\}$ in which the conditional distribution of y has been subject to a certain number of breaks. First suppose that it is either known or assumed that no new break occurs between the end of the sample, T, and the end of the forecast horizon, $T+h$. In this case y_{T+h} can be forecast based on the posterior parameter distribution from the last break segment. Next, suppose that we allow for a single break which could occur in any one

---

[1] A small subset of the many papers that have reported evidence of breaks in economic and financial time series includes Alogouskofis and Smith (1991), Ang and Bekaert (2002), Garcia and Perron (1996), Koop and Potter (2001), Pastor and Stambaugh (2001), Pesaran and Timmermann (2002) and Siliverstovs and van Dijk (2002).

[2] Clements and Hendry (1999) introduce their book as follows: "Economies evolve and are subject to sudden shifts precipitated by legislative changes, economic policy, major discoveries and political turmoil. Macroeconometric models are an imperfect tool for forecasting this highly complicated and changing process. Ignoring these factors leads to a wide discrepancy between theory and practice."

of the h different locations. Each scenario has a different probability assigned to it that must be computed under the assumed breakpoint model. As the number of potential breaks grows, the number of possible break locations grows more than proportionally, complicating the problem even further.

Although breaks are found in most economic time-series, the likely paucity of breaks in a given data sample means that it is important to see how much can be learned about future breaks from the breaks that occurred in the past. This is related to how similar the parameters are across various break segments. A narrow dispersion of the distribution of parameters across breaks suggests that parameters from previous break segments contain considerable information on the parameters after a subsequent break while a wider spread suggests less commonality and more uncertainty.

To address this question we propose a hierarchical hidden Markov chain (HMC) approach which assumes that the parameters within each break segment are drawn from some common meta distribution. Our approach provides a flexible way of using all the sample information to compute forecasts that embody information on the size and frequency of past breaks instead of discarding observations prior to the most recent break point. As new regimes occur, the priors of the meta distribution are updated using Bayes' rule. Furthermore, uncertainty about the number of break points during the in-sample period can be integrated out by means of Bayesian model averaging techniques.

Our breakpoint detection, model selection and estimation procedures build on existing work in the Bayesian literature including Gelman et al (2002), Inclan (1994), Kim, Nelson and Piger (2004), Koop (2003), McCulloch and Tsay (1993) and−most notably−Chib (1998). However, to handle forecasting outside the data sample we extend existing papers by allowing for the occurrence of random breaks drawn from the meta distribution. We apply the proposed method in an empirical exercise that forecasts US Treasury Bill rates out-of-sample. The results show the success of the Bayesian hierarchical HMC method that accounts for the possibility of breaks over the forecast horizon vis-a-vis procedures that ignore future breaks, particularly at long forecast horizons.

The paper is organized as follows: Section 2 generalizes the hidden Markov chain model of Chib (1998) by extending it with a hierarchical structure to account for estimation of the parameters of the meta distribution. Section 3 explains how to forecast future realizations under different break point scenarios. Section 4 provides the empirical application and Section 5 concludes.

2

2 Modeling the Break Process

Our break point model builds on the Hidden Markov Chain (HMC) formulation of the multiple change point problem proposed by Chib (1998). Breaks are captured through a state variable, S_t, that takes integer values $(1, 2, ..., K+1)$ tracking the regime from which a particular observation, y_t, of some underlying stochastic process is drawn. Thus, $s_t = l$ indicates that y_t has been drawn from $f\left(y_t | \mathcal{Y}_{t-1}, \boldsymbol{\theta}_l\right)$, where $\mathcal{Y}_t = \{y_1, ..., y_t\}$ is the current information set, $\boldsymbol{\theta}_l = [\boldsymbol{\beta}_l, \sigma_l^2]$ represents the location and scale parameters in regime l, i.e. $\boldsymbol{\theta}_t = \boldsymbol{\theta}_l$ if $\tau_{l-1} \leq t \leq \tau_l$ and $\Upsilon_K = \{\tau_0, ..., \tau_{K+1}\}$ is the collection of break points.[3]

The state variable S_t is modeled as a discrete state first order Markov process with the transition probability matrix constrained to reflect a multiple change point model. At each point in time, S_t can either remain in the current state or jump to the next state. Conditional on the presence of K breaks in the in-sample period, the one-step-ahead transition probability matrix takes the form

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & 0 & \ldots & 0 \\ 0 & p_{22} & p_{23} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & p_{KK} & p_{K,K+1} \\ 0 & 0 & \ldots & 0 & 1 \end{pmatrix}, \tag{1}$$

where $p_{j-1,j} = Pr\left(s_t = j | s_{t-1} = j-1\right)$ is the probability of moving to regime j at time t given that the state at time $t-1$ is $j-1$. Note that $p_{ii} + p_{i,i+1} = 1$ and $p_{K+1,K+1} = 1$ due to the assumption of K breaks which means that - conditional on K breaks occurring in the data sample - the process terminates in state $K+1$.[4] Once we turn to out-of-sample forecasting, we show how to relax this condition and integrate out uncertainty about the number of breaks.

The regime switching model proposed by Hamilton (1988) arises as a special case of this setup when the parameters after a break are drawn from a discrete distribution with a finite number of states. If identical states are

---

[3]Throughout the paper we assume that $\tau_0 = 0$.

[4]Strictly speaking the transition probability matrix—and the other model parameters— should be indexed by the number of breaks, K, i.e. P_K. However, to keep the notation as simple as possible we do not use this notation.

3

known to recur, imposing this structure on the transition probability matrix can lead to efficiency gains as it can lower the number of parameters that need to be estimated. Conversely, wrongly imposing the assumption of recurring states will lead to inconsistent parameter estimates.

We assume that the non-zero elements of (1), p_{ii}, are independent of p_{jj}, $j \neq i$, and are drawn from a beta distribution,[5]

$$p_{ii} \sim Beta\left(\underline{a}, \underline{b}\right), \text{ for } i = 1, 2..., K. \tag{2}$$

The joint density of $\mathbf{p} = (p_{11}, ..., p_{KK})'$ is then given by

$$\pi\left(\mathbf{p}\right) = c_K \prod_{i=1}^{K} p_{ii}^{(\underline{a}-1)} \left(1 - p_{ii}\right)^{(\underline{b}-1)}, \tag{3}$$

where $c_K = \{\Gamma\left(\underline{a} + \underline{b}\right) / \Gamma\left(\underline{a}\right) \Gamma\left(\underline{b}\right)\}^K$. The parameters $\underline{a}$ and $\underline{b}$ can be specified to reflect any prior beliefs about the mean duration of each regime.[6]

Since we are interested in forecasting values of the time-series outside the estimation sample, we extend this set up to a hierarchical break point formulation (see Carlin et al. (1992)) by making use of meta distributions for the unknown parameters. We do so by assuming that the coefficient vector, $\boldsymbol{\beta}_j$, and error term precision, σ_j^{-2}, in each regime are drawn from common distributions, $\boldsymbol{\beta}_j \sim (\mathbf{b}_0, \mathbf{B}_0)$ and $\sigma_j^{-2} \sim (v_0, d_0)$, respectively, where $\mathbf{b}_0$ and v_0 are the location and $\mathbf{B}_0$ and d_0 the scale parameters of the two distributions.[7] The assumption that the parameters are drawn from a meta distribution is not very restrictive. For example, the pooled scenario (all parameters are identical across regimes) and the regime-specific scenario (each regime has different (own) parameters) can be seen as special cases. Which scenario most closely represents the data can be inferred from the estimates of $\mathbf{B}_0$ and d_0. To facilitate the analysis, we posit a hierarchical prior for the regime coefficients $\{\boldsymbol{\beta}_j, \sigma_j^{-2}\}$ using a random coefficient model. The hierarchical prior places structure on the differences between regime coefficients, but at the same time posits that they come from a common distribution.

---

[5]Throughout the paper we use underscore bars (e.g. $\underline{a}$) to denote parameters of a prior density.

[6]Because the prior mean of p_{ii} equals $\underline{p} = \underline{a}/(\underline{a} + \underline{b})$, the prior density of the regime duration, d, is approximately $\pi(d) = \underline{p}^{d-1}\left(1 - \underline{p}\right)$ with a mean of $(\underline{a} + \underline{b})/\underline{b}$.

[7]We model the precision parameter because it is easier to deal with its distribution in the hierarchical step.

We assume that the $(r+1) \times 1$ vectors of regime-specific coefficients, $\boldsymbol{\beta}_j$, $j = 1, ..., K+1$ are independent draws from a normal distribution, $\boldsymbol{\beta}_j \sim N(\mathbf{b}_0, \mathbf{B}_0)$, while the regime error term precisions σ_j^{-2} are IID draws from a Gamma distribution, i.e. $\sigma_j^{-2} \sim G(v_0, d_0)$. At the next level of the hierarchy we assume that

$$
\begin{align}
\mathbf{b}_0 &\sim N\left(\underline{\boldsymbol{\mu}}_\beta, \underline{\boldsymbol{\Sigma}}_\beta\right) \tag{4}\\
\mathbf{B}_0^{-1} &\sim W\left(\underline{v}_\beta, \underline{\mathbf{V}}_\beta^{-1}\right) \tag{5}
\end{align}
$$

where $W(.)$ represents a Wishart distribution and $\underline{\boldsymbol{\mu}}_\beta$, $\underline{\boldsymbol{\Sigma}}_\beta$, $\underline{v}_\beta$ and $\underline{\mathbf{V}}_\beta^{-1}$ are hyperparameters that need to be specified *a priori*. Finally, following George et al. (1993), the error term precision hyperparameters v_0 and d_0 are assumed to follow an exponential and Gamma distribution, respectively:

$$
\begin{align}
v_0 &\sim Exp\left(\underline{\rho}_0\right) \tag{6}\\
d_0 &\sim Gamma\left(\underline{c}_0, \underline{d}_0\right) \tag{7}
\end{align}
$$

and $\underline{\rho}_0$, $\underline{c}_0$ and $\underline{d}_0$ are their hyperparameters. Appendix A contains details of the Gibbs sampler used to simulate our hierarchical HMC model.[8]

2.1 Model Comparisons Under Different Numbers of Breaks

To assess how many break points (K) the data supports, we estimate separate models for a range of sensible numbers of break points and then compare the results across models. A variety of classical and Bayesian approaches are available to select the appropriate number of breaks in regression models. A classical approach that treats the parameters of the different regimes as given and unrelated has been advanced by Bai and Perron (1998, 2003). This approach is not, however, suitable for forecasting as it does not account for new regimes occurring after the end of the estimation sample.

Here we adopt the Bayesian approach developed by Chib (1995, 1996) that is well suited for model comparisons under high dimensional parameter

---

[8]Maheu and Gordon (2004) also use a Bayesian method to forecasting under structural breaks but assume that the post-break distribution is given by a subjective prior and do not apply a hierarchical hidden Markov chain approach to update the prior distribution after a break.

spaces. Let the model with i breaks be denoted by M_i. The method obtains an estimate for the marginal likelihood of each model $f(y|M_i)$ and ranks the different models by means of their Bayes factors:

$$B_{ij} = \frac{f(y|M_i)}{f(y|M_j)},$$

where

$$f(y|M_i) = \frac{f(y|M_i, \Theta, \mathbf{p}) \pi(\Theta, \mathbf{p}|M_i)}{\pi(\Theta, \mathbf{p}|M_i, y)}, \tag{8}$$

$$\Theta = \left(\boldsymbol{\beta}_1, \sigma_1^2, ..., \boldsymbol{\beta}_{K+1}, \sigma_{K+1}^2, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0\right).$$

The unknown parameters, Θ and $\mathbf{p}$ can be replaced by maximum likelihood estimates or their posterior means or modes. Large values of B_{ij} indicate that the data supports M_i over M_j (Jeffreys, 1961). Appendix B gives details of how the three components of (8) are computed.

3   Posterior Predictive Distributions

In this section we show how to generate h step ahead out-of-sample forecasts from the model proposed in Section 2. Our information set is given by data up to the point of the prediction, T, i.e. $\mathcal{Y}_T = \{y_1, ..., y_T\}$. Armed with estimates of the break points and parameters in the different regimes, we update the meta distribution and use this information to forecast future values of y occurring after the end of our sample, T.

Conditional on $\mathcal{Y}_T$, density or point forecasts of the y process h steps ahead, y_{T+h}, can be made under a range of scenarios depending on what is assumed about the possibility of breaks over the period $[T, T+h]$. For illustration we compare the 'no break' and 'break' scenarios. Under the 'no break' scenario y_{T+h} can be forecast using only the posterior distribution of the parameters from the last regime, $\{\boldsymbol{\beta}_{K+1}, \sigma_{K+1}^2\}$. Under the 'break' scenario we allow for the possibility of multiple new breaks between T and $T+h$. In the event of such breaks, forecasts of y_{T+h} based solely on the posterior distribution of $\boldsymbol{\beta}_{K+1}$ and σ_{K+1}^2 will be biased and information about the break process is required. To compute the probabilities of all possible break dates, an estimate of the probability of staying in the last regime, $p_{K+1,K+1}$, is also required. The posited meta distribution for the regression parameters, $\boldsymbol{\beta}_j \sim (\mathbf{b}_0, \mathbf{B}_0)$, and the error term precisions, $h_j \sim (v_0, d_0)$,

assumed in the hierarchical structure provide the distribution from which the parameters of any new regimes over the forecast horizon are drawn.

Using the Markov chain property and conditioning on being in regime $K+1$ at time T, the probability of a break at time $T+j$ is

$$Pr\left(s_{T+h} = K+2 \middle| \tau_{K+1} = T+j, s_T = K+1\right) = \left(1 - p_{K+1,K+1}\right) p_{K+1,K+1}^{j}, \tag{9}$$

where $\tau_{K+1} \in [T+1; T+h]$ tracks when break number $K+1$ happens and $p_{K+1,K+1}$ is the probability of remaining in state $K+1$.[9] Notice that, for forecasting purposes, an estimate of the transition probability in the last regime, $p_{K+1,K+1}$, is required in order to compute the probability of a new break occurring in the out-of-sample period. Hence, while the transition probability matrix (1) conditional on K breaks in the in-sample period assumes that $p_{K+1,K+1} = 1$, the second-stage out-of-sample forecast replaces (1) by the following transition probability matrix

$$\tilde{\mathbf{P}} = \begin{pmatrix} p_{11} & p_{12} & 0 & \dots & & 0 & \\ 0 & p_{22} & p_{23} & \dots & & 0 & \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \\ 0 & \dots & 0 & p_{KK} & p_{K,K+1} & & \\ 0 & 0 & \dots & 0 & p_{K+1,K+1} & p_{K+1,K+2} & \\ 0 & 0 & \dots & & 0 & p_{K+2,K+2} & \\ & & & & & & \ddots \end{pmatrix}, \tag{10}$$

where the $(K+1) \times (K+1)$ sub-matrix in the upper left corner pertains to the observed data sample, $\{y_1, ..., y_T\}$ and is identical to (1) except for the final element, $p_{K+1,K+1}$ which in general is different from unity. The remaining part of $\tilde{\mathbf{P}}$ describes the breakpoint dynamics in the out-of-sample period.

We next show how forecasts are computed under different out-of-sample scenarios before computing a composite forecast as a probability-weighted average of the forecasts under each out-of-sample breakpoint scenario and showing how to integrate out the uncertainty surrounding the number of in-sample breaks.

---

[9]To simplify notation in the following, we do not explicitly condition posterior distributions on the fixed prior hyperparameters in (4)-(7).

3.0.1 No Break in $[T, T+h]$

Suppose there is no break between T and $T+h$, so the new data is generated from the last regime $(K+1)$ in the observed sample. Then $p\left(y_{T+h}|\, s_{T+h} = K+1, y_T\right)$ is drawn from

$$\int \int p\left(y_{T+h}|\, \boldsymbol{\beta}_{K+1}, \sigma^2_{K+1}, s_{T+h} = K+1, \mathcal{Y}_T\right)$$
$$\times \pi\left(\boldsymbol{\beta}_{K+1}, \sigma^2_{K+1}\Big|\, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T\right) d\boldsymbol{\beta}_{K+1} d\sigma^2_{K+1},$$

where $\mathcal{S}_T = (s_1, ..., s_T)$ is the collection of values of the latent state variable up to period T. We thus proceed as follows:

Obtain a draw from $\pi\left(\boldsymbol{\beta}_{K+1}, \sigma^2_{K+1}\Big|\, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T\right)$,

Draw y_{T+h} from the posterior predictive density,

$$y_{T+h} \sim p\left(y_{T+h}|\, \boldsymbol{\beta}_{K+1}, \sigma^2_{K+1}, s_{T+h} = K+1, \mathcal{Y}_T\right). \qquad (11)$$

3.0.2 Single Break in $[T, T+h]$

After a new break the process is generated under the parameters from regime number $K+2$. For a given break date, $T+j$ $(1 \le j \le h)$, we obtain $p\left(y_{T+h}|\, s_{T+h} = K+2, \tau_{K+1} = T+j, \mathcal{Y}_T\right)$ from

$$\int \cdots \int p\left(y_{T+h}|\, \boldsymbol{\beta}_{K+2}, \sigma^2_{K+2}, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, s_{T+h} = K+2, \tau_{K+1} = T+j, \mathcal{Y}_T\right)$$
$$\times \pi\left(\boldsymbol{\beta}_{K+2}, \mathbf{b}_0, \mathbf{B}_0\Big|\, \mathcal{Y}_T\right) \times \pi\left(\sigma^2_{K+2}, v_0, d_0\Big|\, \mathcal{Y}_T\right) d\boldsymbol{\beta}_{K+2} d\mathbf{b}_0 d\mathbf{B}_0 d\sigma^2_{K+2} dv_0 dd_0.$$

To see how we update the posterior distributions of $\mathbf{b}_0$, $\mathbf{B}_0$, v_0 and d_0 in this case, define $\boldsymbol{\beta}_{1:K+1} = \left(\boldsymbol{\beta}'_1, ..., \boldsymbol{\beta}'_{K+1}\right)'$ and $\sigma^2_{1:K+1} = \left(\sigma^2_1, ..., \sigma^2_{K+1}\right)'$. We then proceed as follows:

Draw $\mathbf{b}_0$ from

$$\mathbf{b}_0 \sim \pi\left(\mathbf{b}_0|\, \boldsymbol{\beta}_{1:K+1}, \sigma^2_{1:K+1}, \mathbf{B}_0, v_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T\right),$$

and $\mathbf{B}_0$ from

$$\mathbf{B}_0 \sim \pi\left(\mathbf{B}_0|\, \boldsymbol{\beta}_{1:K+1}, \sigma^2_{1:K+1}, \mathbf{b}_0, v_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T\right).$$

8

Draw v_0 from

$$v_0 \sim \pi\left(v_0\middle|\boldsymbol{\beta}_{1:K+1}, \sigma_{1:K+1}^{-2}, \mathbf{b}_0, \mathbf{B}_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T\right),$$

and d_0 from

$$d_0 \sim \pi\left(d_0\middle|\boldsymbol{\beta}_{1:K+1}, \sigma_{1:K+1}^{-2}, \mathbf{b}_0, \mathbf{B}_0, v_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T\right).$$

Draw $\boldsymbol{\beta}_{K+2}$ and h_{K+2} from their respective priors given by $\pi\left(\boldsymbol{\beta}_{K+2}\middle|\mathbf{b}_0, \mathbf{B}_0\right)$ and $\pi\left(h_{K+2}\middle|v_0, d_0\right)$, respectively, for a fixed set of hyperparameters.

Draw y_{T+h} from the posterior predictive density,

$$y_{T+h} \sim p\left(y_{T+h}\middle|\boldsymbol{\beta}_{K+2}, \sigma_{K+2}^2, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, s_{T+h} = K+2, \tau_{K+1} = T+j, \mathcal{Y}_T\right). \tag{12}$$

To obtain the estimate of $p_{K+1,K+1}$ needed in (9), we combine information from the last regime with prior information, assuming the prior $p_{K+1,K+1} \sim Beta(\underline{a}, \underline{b})$, so

$$p_{K+1,K+1}\middle|\mathcal{Y}_T \sim Beta(\underline{a} + n_{K+1,K+1}, \underline{b} + 1) \tag{13}$$

where $n_{K+1,K+1}$ is the number of observations from regime $K+1$.

3.0.3 Multiple Breaks in $[T, T+h]$

Assuming $h \geq 2$, we can readily extend the previous discussion to multiple out-of-sample break points. When considering the possibility of two or more breaks, we need an estimate of the probability of staying in regime $K+j$, $p_{K+j,K+j}$, $j \geq 2$. This is not needed for the single break case since $-$ by assumption $-$ $p_{K+2,K+2}$ is set equal to one. To this end we modify the hierarchical HMC by adding a prior distribution for the hyperparameters a and b of the transition probability,[10]

$$\begin{aligned} a &\sim Gamma\left(\underline{a_0}, \underline{b_0}\right), \tag{14} \\ b &\sim Gamma\left(\underline{a_0}, \underline{b_0}\right). \end{aligned}$$

---

[10]Following earlier notations, these parameters appear here without the underscore bar since they will be estimated from the data.

$p_{ii}-$values are now drawn from the conditional beta posterior

$$p_{ii}|\,\Theta, \boldsymbol{\tau}, y \sim Beta(a + l_i, b + 1),$$

where $l_i = \tau_i - \tau_{i-1} - 1$ is the duration of regime i. The distribution for the hyperparameters a and b is not conjugate so sampling is accomplished using a Metropolis-Hasting step. The conditional posterior distribution for a is

$$\pi\left(a|\,\Theta, \boldsymbol{\tau}, \mathbf{p}, b, \mathcal{Y}_T\right) \propto \prod_{i=1}^{K} Beta\left(p_{ii}|\,a, b\right) Gamma\left(a|\,\underline{a_0}, \underline{b_0}\right).$$

To draw candidate values, we use a Gamma proposal distribution with shape parameter ς, mean equal to the previous draw a^g

$$q\left(a^*|\,a^g\right) \sim G\left(\varsigma, \varsigma/a^g\right),$$

and acceptance probability

$$\xi\left(a^*|\,a^g\right) = \min\left[\frac{\pi\left(a^*|\,\Theta, \boldsymbol{\tau}, P, b, y\right)/q\left(a^*|\,a^g\right)}{\pi\left(a^g|\,\Theta, \boldsymbol{\tau}, P, b, y\right)/q\left(a^g|\,a^*\right)}, 1\right]. \tag{15}$$

Using these new posterior distributions, we generate draws for $p_{K+2,K+2}$ using the convolution of the prior distribution for the p_{ii}'s and the resulting posterior densities for a and b,[11]

$$p_{K+2,K+2}|\,a, b \sim Beta(a, b).$$

Allowing for two break points in the out-of-sample period, we get $\sum_{i=1}^{h-1}(h-i)$ $= h(h-1)/2$ possible break point locations and the weight assigned to such scenarios is (assuming $j \le l \le h$)[12]

$$\Pr\left(s_{T+h} = K + 3|\,\tau_{K+2} = T + l, \tau_{K+1} = T + j, s_T = K + 1\right)$$
$$= p_{K+1,K+1}^{j}\left(1 - p_{K+1,K+1}\right) p_{K+2,K+2}^{l-j-1}\left(1 - p_{K+2,K+2}\right).$$

---

[11]In contrast to the case for $p_{K+1,K+1}$, we do not have any information about the length of regime $K+2$ from the estimation sample and rely on prior information to get an estimate for $p_{K+2,K+2}$.

[12]To be more precise, only $\sum_{i=1}^{h-1}(h-i)$ combinations of two break points are considered, while the remaining $h + 1$ cases correspond to the occurrence of zero or one new break. The general weight for this case is $p_{K+1,K+1}^{j}\left(1 - p_{K+1,K+1}\right)^{I(h>j)} p_{K+2,K+2}^{(h-j-1)I(h-1>j)}$, where $j = 0, ..., h - 1$ is the break date and $I(\cdot)$ is the indicator function taking a value of 1 if the inner argument is valid and zero otherwise.

The extension to $m > 2$ break points is straightforward and the transition probability for this case takes the form

$$\Pr\left(s_{T+h} = K + m + 1 \middle| \tau_{K+m} = T + z, ..., \tau_{K+2} = T + l, \tau_{K+1} = T + j, s_T = K + 1\right)$$
$$= \prod_{i=1}^{m} p_{K+i,K+i}^{d_i} \left(1 - p_{K+i,K+i}\right), \tag{16}$$

where $d_i \geq 1$ is the duration of regime $K + i$.

3.1 Combining the Forecasts using Bayesian Model Averaging

So far we have shown how to integrate out uncertainty about the number of out-of-sample (future) breaks. However, we have not dealt with the fact that we do not know the true number of in-sample breaks. To integrate out the uncertainty about the number of in-sample breaks, we compute the predictive density as a weighted average of the predictive densities under the meta distributions, each of which conditions on a given value of K, using the model posteriors as (relative) weights. To this end let $p_K(y_{T+h}|\mathcal{Y}_T) \equiv p\left(y_{T+h}| s_T = K + 1, \mathcal{Y}_T\right)$ be the posterior density conditional on K breaks having occurred at time t. Combining (9)-(11), we can integrate out the state probability S_{T+h} from the predictive densities and compute the posterior predictive density that considers the forecasts under the no-break and break scenarios as follows.

$$
\begin{aligned}
p_K\left(y_{T+h}|\mathcal{Y}_T\right) =\ & \Pr\left(s_{T+h} = K + 1 | \mathcal{Y}_T\right) \times p\left(y_{T+h}| s_T = K + 1, s_{T+h} = K + 1, \mathcal{Y}_T\right) + \\
& \sum_{j=1}^{h} \Pr\left(s_{T+h} = K + 2 | \tau_{K+1} = T + j, s_T = K + 1\right) \\
& \times p\left(y_{T+h}| \tau_{K+1} = T + j, s_T = K + 1, s_{T+h} = K + 2, \mathcal{Y}_T\right) + \\
& \sum_{j=1}^{h-m+1} \sum_{l=j+1}^{h-m+2} \cdots \sum_{m=l+1}^{h} \\
& \Pr\left(s_{T+h} = K + m + 1 | \tau_{K+m} = T + l, ..., \tau_{K+1} = T + j, s_T = K + 1\right) \\
& \times p\left(y_{T+h}| \tau_{K+m} = T + l, ..., \tau_{K+1} = T + j, \mathcal{Y}_T\right).
\end{aligned}
\tag{17}
$$

This forecast conditions on the existence of K breaks in the in-sample period that terminates at time T. However, in some applications there may be considerable uncertainty surrounding the number of in-sample breaks and

so it is reasonable to integrate out uncertainty about the right number of break points in the data. We do this by means of Bayesian model averaging techniques. This requires computing a weighted average of the composite distributions based on the models that assume different values of s_T. Let M_k be the model that assumes $k-1$ breaks at time T (i.e., $s_T = k$). The predictive density under the Bayesian model average is

$$p(y_{T+h}|\mathcal{Y}_T) = \sum_{k=1}^{\bar{K}} p_k(y_{T+h}|\mathcal{Y}_T)p(M_k|\mathcal{Y}_T), \qquad (18)$$

where $\bar{K}$ is some upper limit on the largest number of breaks that is entertained. The weights used in the average are given by the posterior model probabilities:

$$p\left(\left.M_k\right|y\right) \propto f\left(\left.y\right|M_k\right)p\left(M_k\right) \qquad (19)$$

where $f\left(\left.y\right|M_k\right)$ is the marginal likelihood from (8) and $p(M_k)$ is the prior for model M_k.

4 Empirical Application

We apply the proposed methodology to model U.S. Treasury Bill rates, a key economic variable of general interest. This variable is ideally suited for our analysis since previous studies have documented structural instability and regime changes in the underlying process, c.f. Ang and Bekaert (2002), Garcia and Perron (1996) and Gray (1996).

4.1 Data

We analyze monthly data on the nominal three month US T-bill rate from July 1947 through December 2002. Prior to the beginning of this sample interest rates were fixed for a lengthy period so our data set is the longest available post-war sample with variable interest rates. The data source is the Center for Research in Security Prices at the Graduate School of Business, University of Chicago. T-bill yields are computed from the average of bid and ask prices and are continuously compounded 365 day rates. Figure 1 plots the monthly yields. We divide the data into two parts. Observations from the beginning of the sample through December 1997 are used as the estimation sample, while data from January 1998 through December 2002 (60 observations) are used as the forecasting (out-of-sample) period.

4.2 Prior Elicitation and Posterior Inference

In Section 2 we specified a beta distribution for the diagonal elements of the transition probability matrix, a Normal-Wishart distribution for the meta-distribution parameters of the regression coefficients and a Gamma-Exponential for the error term precision parameters. Implementation of our method requires assigning values to the associated hyperparameters. For the p_{ii}'s we assume a non-informative prior for all the diagonal elements of (1), i.e., $\underline{a} = \underline{b} = 0.5$. For the Normal-Wishart distribution, we specify $\boldsymbol{\mu}_\beta = \mathbf{0}$, $\boldsymbol{\Sigma}_\beta = 1000 \times I_{r+1}$, $v_\beta = 2$ and $\mathbf{V}_\beta = I_{r+1}$, where $\mathbf{0}$ is an $(r + 1 \times 1)$ vector of zeros, while I_{r+1} is the $(r + 1 \times r + 1)$ identity matrix. These values reflect no specific prior knowledge and are diffuse over sensible ranges of values for both the Normal and Wishart distribution. Similarly, we set $\underline{\rho_0} = 0.01$, $\underline{c_0} = 1$ and $\underline{d_0} = 0.01$, allowing the prior for v_0 and d_0 to be uninformative over the positive real line.

We also conducted a prior sensitivity analysis to ensure that the empirical results are robust to different prior beliefs. For the transition probability matrix, $\mathbf{P}$, we modified $\underline{a}$ and $\underline{b}$ to account for a wide set of regime durations; we also changed the beta prior mean hyperparameters $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\Sigma}_\beta$, and the regime error term precision hyperparameters $\underline{c_0}$, $\underline{d_0}$ and $\underline{\rho_0}$. In all cases we found the HMC estimates were insensitive to changes in the prior hyperparameters. More care is needed when dealing with the beta prior precision hyperparameter $\underline{\mathbf{V}_\beta}$. For small values of its diagonal elements (less than 0.5), the meta distribution for the regression coefficients will not allow enough variation across regimes, and as a consequence the regime regression coefficients are clustered around the mean of the meta-distribution, $\mathbf{b}_0$. Empirical results were found to be robust for values of the diagonal elements of $\underline{\mathbf{V}_\beta}$ greater than or equal to 1.

4.3 Model Estimates

In view of their empirical success and extensive use in forecasting[13] we model the process underlying T-bill rates $\{y_t\}$ as an rth order autoregressive (AR)

---

[13]See Pesaran and Timmermann (2004) for further references to the literature on forecasts from AR models subject to breaks.

model allowing for up to K breaks over the observed sample $\{y_1, ..., y_T\}$:[14]

$$
y_t = \begin{cases}
\beta_{1,0} + \beta_{1,1}y_{t-1} + ... + \beta_{1,r}y_{t-r} + \sigma_1\epsilon_t, & t = 1, ..., \tau_1 \\
\beta_{2,0} + \beta_{2,1}y_{t-1} + ... + \beta_{2,r}y_{t-r} + \sigma_2\epsilon_t, & t = \tau_1 + 1, ..., \tau_2 \\
\vdots & \\
\beta_{K+1,0} + \beta_{K+1,1}y_{t-1} + ... + \beta_{K+1,r}y_{t-r} + \sigma_{K+1}\epsilon_t, & t = \tau_K + 1, ..., T.
\end{cases}
\tag{20}
$$

Within the class of AR processes, this specification is quite general and allows for intercept and slope shifts as well as changes in the error variances. Each regime j, $j = 1, ...K+1$, is characterized by a vector of regression coefficients, $\boldsymbol{\beta}_j = \left(\beta_{j,0}, \beta_{j,1}, ...\beta_{j,r}\right)'$, and an error term variance, σ_j^2, for $t = \tau_{j-1}+1, ..., \tau_j$.

Following previous studies we consider first and second order autoregressive specifications for the T-bill rate. The AR(1) model is our base specification and results for the AR(2) model are included to demonstrate the robustness of our empirical findings. For each AR specification we obtain a different model by varying the number of breaks, K, and we rank these models by means of their marginal likelihoods computed using the Chib method from Section 2.1. Table 1 reports maximized log-likelihood values, marginal log-likelihoods and break dates for values of K ranging from zero to seven. For both autoregressive specifications, the marginal log-likelihood is maximized at $K = 6$. The models with $K = 7$ break points obtain basically the same marginal log-likelihood, suggesting that the additional break is not supported by the data.

Figure 2 plots the posterior probability for the six break points under the AR(1) model. Results are nearly identical for the AR(2) model. The local unimodality of the posterior distributions shows that the break points are generally precisely estimated.[15] For the autoregressive models with six break points, Tables 2, 3, 4 and 5 report the autoregressive parameters, variance, transition probability and the average number of months spent in each regime. In all regimes the interest rate is highly persistent and close to a unit root process. The error term variance is particularly high for regime 5 (lasting from October 1982 to July 1989), and quite low for regimes 1 (Sep

---

[14]Although structural break tests often do not reveal the form of the instability, a widely used class of models assumes that it can be well approximated by a sequence of discrete structural breaks.

[15]In contrast, the plot for the model with seven break points (not shown here) had a very wide posterior density for the 1969 break, providing further evidence against the inclusion of an additional break point.

1947 - Nov 1957), 3 (Jul 1960 - Sep 1966) and 7 (July 1989 - Dec 1997). Tables 3 and 5 report prior parameter estimates, i.e. the meta distribution parameters under the AR(1) and AR(2) models. For example, in the case of the AR(1) specification, the two meta distributions are

$$\boldsymbol{\beta}_j \sim N \left(\begin{array}{c} 0.1908 \\ 0.9438 \end{array}, \left[\begin{array}{cc} 0.2731 & -0.0088 \\ & 0.1981 \end{array} \right] \right),$$

$$\sigma_j^{-2} \sim Gamma\,(0.7748, 0.0431)\,.$$

>From the properties of the Gamma distribution, the mean of the precision of the meta distribution is almost 18 and the standard error is around 20. These values are consistent with the values of the inverse of the variance estimates in Table 2.

4.4 Unit Root Dynamics

The persistent dynamics observed in some of the regimes may be a cause for concern when calculating multi-step forecasts. Unit roots or even explosive roots could affect the meta distribution which averages parameter values across the regimes. To deal with this problem, we propose the following alternative constrained parameterization of the AR(1) model:

$$\Delta y_{t+1} = \alpha_j \phi_j - \phi_j y_t + \epsilon_{t+1}, \quad j = 1, ..., K + 1, \tag{21}$$

where $\epsilon_{t+1} \sim N(0, \sigma_j^2)$. If $\phi_j = 0$, the process has a unit root while if $0 < \phi_j < 2$, it is a stationary AR(1) model. Notably, in the case with a unit root there is no drift irrespective of the value of α_j. Assuming that the process is stationary, its long run mean is simply α_j.

We estimate our hierarchical HMC model under this new parameterization. To avoid explosive roots and negative unconditional mean, we constrain ϕ_j to lie in $[0, 1]$ and α_j to be strictly positive. We accomplish this by assuming that the priors for the regime regression parameters and error term precisions are

$$\begin{aligned} \boldsymbol{\beta}_j &\sim N\,(\mathbf{b}_0, \mathbf{B}_0)\, I\left(\boldsymbol{\beta}_j \in A\right), \\ h_j &\sim G\,(v_0, d_0), \end{aligned} \tag{22}$$

15

with $\boldsymbol{\beta}_j = \left(\alpha_j, \phi_j \right)'$, while the priors for the meta-distribution hyperparameters in this case become

$$
\begin{aligned}
\mathbf{b}_0 &\sim N\left(\underline{\boldsymbol{\mu}}_\beta, \underline{\boldsymbol{\Sigma}}_\beta\right) I\left(\mathbf{b}_0 \in A\right) && (23) \\
\mathbf{B}_0^{-1} &\sim W\left(\underline{v}_\beta, \underline{\mathbf{V}}_\beta^{-1}\right).
\end{aligned}
$$

$I\left(x \in A\right)$ in (22) and (23) is an indicator function that equals 1 if x belongs to A and is zero otherwise. A is the set $[0, \infty) \times (0, 1]$. No changes are needed in the priors for the meta-distribution hyperparameters of the error term precision, v_0 and d_0. We obtain the same posterior densities as under the unrestricted model although these distributions are now truncated due to the inequality constraints.

Tables 6 and 7 report estimates assuming six breaks. The detected break points are the same as those found for the unrestricted AR(1) model. To be comparable to the earlier tables, the regime coefficients and meta distribution results refer to $\alpha_j \phi_j$ and $1 - \phi_j$. For each regime and for the meta distribution, Table 6 shows the probability of a unit root, calculated as $\Pr\left(\phi_j = 0\right)$. Regimes 1 and 3 are more likely to be non-stationary, both have a unit root probability slightly higher than one third. This is also reflected in the meta distribution results, $\Pr\left(b_0(2) = 1\right) = 0.38$. Figure 3 shows the posterior densities for the meta distribution parameters for the unrestricted and restricted models.

4.5 Forecasts

We finally turn to the calculation of out-of-sample predictive distributions. We base our results on the model with six breaks occurring over the estimation sample, but later also present results based on Bayesian Model Averaging across different numbers of breaks using (18) and (19). Attention is restricted to the AR(1) specification, since predictive distributions under the AR(2) model are very similar. We start from the end of the in-sample period T (December 1997) and compute predictive distributions for period $T + h$ $(h = 1, .., 60)$ under the scenarios described in section 3. To obtain the predictive density under the no break scenario we use the information from the last regime of the hierarchical HMC model. To gain intuition we also show results under the meta distribution scenario which represents the opposite extreme case and assumes that a single break occurs at the beginning of the out-of-sample period, $T + 1$. This case draws a new vector of

regression parameters and a new error term variance from the meta distribution. Finally, the hierarchical HMC forecast that allows for out-of-sample breaks is based on the scenarios in (17). We refer to this as the 'composite' forecast. This forecast assumes uninformative priors for both a and b, with $\underline{a_0} = 1$ and $\underline{b_0} = 0.1$. Different values for $\underline{a_0}$ and $\underline{b_0}$ were tried and the results were found to be robust to changes in $\underline{a_0}, \underline{b_0}$.[16]

Figure 4 shows the weights on the 'no break' model as a function of the forecast horizon. One minus this quantity is the weight assigned to draws from the meta distribution. As expected, the weight decreases monotonically in the forecast horizon, h, implying that higher weight is given to parameters drawn from the meta distribution the longer the forecast horizon and the higher therefore the chance of experiencing a break prior to $T + h$.

Figure 5 plots the associated forecast densities. The forecast horizon ranges from one to sixty months. As expected from the variance estimates in Table 2, under no break the predictive density is concentrated around its mean, while there is much more uncertainty under the meta distribution which balances differences in parameter estimates across the seven break segments. We also inspect graphically the performance of these forecasts. In each panel the actual (realized) value of the T-bill rate for that period is indicated by a vertical line. After year 2000, the Treasury bill rate declines significantly. In fact, a separate Hidden Markov Chain run restricted only to the period 1998:01-2002:12 confirmed that a break may have occurred at this time. Consequently, after this period the forecasts from the no break model are very poor while the composite model forecasts perform considerably better. The last three panels in Figure 5 show that the no break forecasts are upward biased and unable to capture the decline in the T-Bill rate while the composite forecasts are more accurate.

4.6   Forecast Evaluation

We next investigate the performance of the no break and hierarchical HMC models using the posterior predictive p-value approach. If the model fits the data reasonably well, realized values of the T-bill rate in the out-of-sample period should not be too far out in the tails of the predictive density. To see

---

[16]The posterior means of the hyperparameters a and b are 28.54 and 0.83 while the standard deviations are 17.26 and 0.44. The prior for the transition probabilities is, at the posterior mean of a and b, a beta distribution with a mean of 0.9717 and a variance of 0.0311.

if this holds, for each of the models, M, we compute the percentile of the realized value, $\int_{-\infty}^{y_{T+h}} p\left(z \mid \mathcal{Y}_T, M\right) dz$ and report how many times the model yields posterior predictive p-values below 0.05 or above 0.95, viewing an over-representation of such occurrences as evidence against a model.

Table 8 reports empirical results both for the full out-of-sample period, 1998:01-2002:12 and for the subperiod 2001:01-2002:12. We consider these samples separately due to the evidence of a break around the end of 2000. Forecasting performance is computed as averages across different horizons, h, based on forecasts originating from period T (i.e. based on $\mathcal{Y}_T$). In both cases the forecast density based on the composite model outperforms forecasts from the model that assumes no new breaks. Results are particularly striking for the second subsample where the no break density leads to poor predictions two thirds of the time, while the composite model is never rejected.

We also computed root mean squared forecast errors under the no break and composite models - using both the restricted and unrestricted versions - based either on the posterior mean or the posterior mode of the density forecast as our point forecast of the future T-bill rate. This is a popular measure of forecasting performance. Table 9 shows that the composite model produces more precise point forecasts than the no break model, both in the restricted and unrestricted case. The constrained model seems to marginally outperform the unconstrained one.

When compared to the performance of forecasts from classical approaches, the root mean squared forecast error value of 1.36% from the composite model is substantially below the value generated by forecasts based on estimates from the full sample (2.06%), a five-year rolling window (2.00%) and discounted least squares (2.04%), assuming a discount factor of 0.95.

One last issue that may be relevant when forecasting T-Bill rates is the possible presence of autoregressive conditional heteroskedasticity (ARCH) in the residuals. To some extent our approach deals with this by letting the innovation variance vary across regimes so that the volatility parameter effectively follows a step function. To investigate if the normalized residuals (scaled by the estimated standard deviation) still are heteroskedastic, we ran Lagrange Multiplier (LM) tests, regressing the squared normalized residuals on their lagged values. Even though some ARCH effects remain in the scaled residuals, after scaling the residuals we found that the R^2 of the LM regression was reduced from 16% to 2%.

4.7 Uncertainty about the Number of in-sample Breaks

The empirical results presented thus far were computed from the hierarchical HMC model under the assumption of six breaks in the in-sample period. One can, however, argue that it is not reasonable to condition on the number of breaks. We therefore explicitly consider models with different numbers of break points, integrating out uncertainty about the number of break points in the data by means of the Bayesian model averaging formulas (18)-(19). Specifically, we consider between one and seven breaks in the data and assign equal prior probabilities to each of these models.

Table 10 shows the posterior mass assigned to each of the models. A probability mass close to zero is assigned to the models with five or fewer break points while 82 and 18 percent of the posterior mass is assigned to the models with six and seven break points, respectively.

Figure 6 plots the combined predictive density under Bayesian model averaging along with the predictive density that conditions on six breaks. The graphs reveal that the two densities are very similar. It is therefore not surprising that the out-of-sample root mean squared forecast error values obtained under the two approaches are basically the same and differ only after the fourth decimal. This means that the forecasting results reported so far are very robust to uncertainty about the number of in-sample breaks.

5   Conclusion

The key contribution of this paper has been to introduce a hierarchical hidden Markov chain approach to model the meta distribution for the parameters of the stochastic process underlying structural breaks. This allowed us to forecast economic time series that are subject to structural breaks. Our approach is quite general and can be implemented in different ways from that assumed in the current paper. For example, the state transitions could be allowed to depend on time-varying regressors tracking factors related to uncertainty about institutional shifts or the likelihood of macroeconomic or oil price shocks. When applied to autoregressive specifications for U.S. T-Bill rates, an out-of-sample forecasting exercise found that our approach produces better forecasts than methods that ignore the possibility of future breaks.

The simple 'no new break' approach that forecasts using parameter estimates solely from the last post-break period can be expected to perform well

when the number of observations from the last regime is sufficiently large to deliver precise parameter estimates, and the possibility of new breaks occurring over the forecast horizon is very small, c.f. Pesaran and Timmermann (2004). However, when forecasting many periods ahead or when breaks occur relatively frequently−so the last break point is close to the end of the sample and a new break is likely to occur shortly after the end of the estimation sample− this approach is unlikely to produce satisfactory forecasts.

Intuition for why our approach appears to work quite well in forecasting interest rates is that it effectively shrinks the new parameters drawn after a break towards the meta distribution. Shrinkage has widely been found to be a useful device for improving forecasting performance in the presence of parameter estimation and model uncertainty, c.f. Diebold and Pauly (1990), Stock and Watson (2003), Garratt, Lee, Pesaran and Shin (2003), and Aiolfi and Timmermann (2004). Here it appears to work because the number of breaks that can be identified empirically tends to be small and the parameters of the meta distribution from which such breaks are drawn is reasonably precisely estimated.

References

[1] Aiolfi, M. and A. Timmermann, 2004, Persistence in Forecasting Performance and Conditional Combination Strategies, Unpublished Manuscript, UCSD.

[2] Alogoskoufis, G.S. and R. Smith, 1991, The Phillips Curve, the Persistence of Inflation, and the Lucas Critique: Evidence from Exchange Rate Regimes. American Economic Review 81, 1254-1275.

[3] Ang, A., and G., Bekaert, 2002, Regime Switches in Interest Rates, Journal of Business and Economic Statistics, 20, 163-182.

[4] Bai, J. and P. Perron, 1998, Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica 66, 47-78.

[5] Bai, J. and P. Perron, 2003, Computation and Analysis of Multiple Structural Change Models, Journal of Applied Econometrics, 18, 1-22.

[6] Carlin, B., A.E. Gelfand and A.F.M. Smith, 1992, Hierarchical Bayesian analysis of changepoint problems, Applied Statistics, 41, 389-405.

[7] Chib, S., 1995, Marginal Likelihood from the Gibbs output, Journal of the American Statistical Association, 90, 1313-1321.

[8] Chib, S., 1996, Calculating Posterior Distribution and Modal Estimates in Markov Mixture Models, Journal of Econometrics, 75, 79-97.

[9] Chib, S., 1998, Estimation and Comparison of Multiple Change Point Models, Journal of Econometrics, 86, 221-241.

[10] Chib, S. and I. Jeliazkov, 2001, Marginal Likelihood from the Metropolis-Hastings Output, Journal of the American Statistical Association, 96, 270-281.

[11] Clements, M.P. and D.F. Hendry, 1998, Forecasting Economic Time Series, Cambridge University Press.

[12] Clements, M.P. and D.F. Hendry, 1999, Forecasting Non-stationary Economic Time Series, The MIT Press.

[13] Diebold, F.X and P. Pauly, 1990, The Use of Prior Information in Forecast Combination, International Journal of Forecasting, 6, 503-508.

[14] Garratt, A, K. Lee, M. H. Pesaran and Y. Shin, 2003, Forecast Uncertainties in Macroeconometric Modelling: An Application to the UK Economy. Journal of the American Statistical Association, 98, 829-838.

[15] Garcia, R. and P. Perron, 1996, An Analysis of the Real Interest Rate under Regime Shifts. Review of Economics and Statistics, 78, 111-125.

[16] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D., 2002, Bayesian Data Analysis, Second Edition, Chapman & Hall Editors.

[17] George, E. I., U. E. Makov and A. F. M. Smith, 1993, Conjugate Likelihood Distributions, Scandinavian Journal of Statistics, 20, 147-156.

[18] Gray, S., 1996, "Modeling the Conditional Distribution of Interest Rates as Regime-Switching Process", Journal of Financial Economics, 42, 27-62.

[19] 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.

[20] Inclan, C., 1994, Detection of Multiple Changes of Variance Using Posterior Odds. Journal of Business and Economic Statistics, 11, 289-300.

[21] Jeffreys, H., 1961, Theory of Probability, Oxford University Press, Oxford.

[22] Kim, C.J., C.R. Nelson and J. Piger, 2004, The Less-Volatile US Economy. A Bayesian Investigation of Timing, Breadth, and Potential Explanations. Journal of Business and Economic Statistics 22, 80-93.

[23] Koop, G., 2003, Bayesian Econometrics, John Wiley & Sons, New York.

[24] Koop, G. and S. Potter, 2001, Are Apparent Findings of Nonlinearity Due to Structural Instability in Economic Time Series? Econometrics Journal, 4, 37-55.

[25] Maheu, J.M. and S. Gordon, 2004, Learning, Forecasting and Structural Breaks. Manuscript University of Toronto.

[26] McCulloch, R.E. and R. Tsay, 1993, Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series. Journal of the American Statistical Association 88, 965-978.

[27] Pastor, L. and R.F. Stambaugh, 2001, The Equity Premium and Structural Breaks. Journal of Finance, 56, 1207-1239.

[28] Pesaran, M.H. and A. Timmermann, 2002, Market Timing and Return Prediction under Model Instability. Journal of Empirical Finance, 9, 495-510.

[29] Pesaran, M.H. and A. Timmermann, 2004, Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks. Forthcoming in Journal of Econometrics.

[30] Siliverstovs, B. and D. van Dijk, 2002, Forecasting Industrial Production with Linear, Non-linear and Structural Breaks Models. Manuscript DIW Berlin.

[31] Stock, J.H. and M.W. Watson, 1996, Evidence on Structural Instability in Macroeconomic Time Series Relations, Journal of Business and Economic Statistics, 14, 11-30.

[32] Stock, J.H. and M. Watson, 2003, Combination Forecasts of Output Growth in a Seven-Country Data Set. Forthcoming, Journal of Forecasting.

Appendix A. Gibbs Sampler for the Multiple Regime Model

The posterior distribution of interest is $\pi\left(\Theta, \mathbf{p}, \mathcal{S}_T | \mathcal{Y}_T\right)$, where

$$\Theta = \left(\boldsymbol{\beta}_1, \sigma_1^2, ..., \boldsymbol{\beta}_{K+1}, \sigma_{K+1}^2, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0\right)$$

includes the $K + 1$ regime coefficients and the prior locations and scales, $\mathcal{S}_T = (s_1, ..., s_T)$ is the collection of values of the latent state variable, $\mathcal{Y}_T = (y_1, ..., y_T)'$, and $\mathbf{p} = (p_{11}, p_{22}, ..., p_{KK})'$ summarizes the unknown parameters of the transition probability matrix in (1). The Gibbs sampler applied to our set up works as follows: First the states are simulated conditional on the data, $\mathcal{Y}_T$, and the parameters and, second, the parameters are simulated conditional on the data and $\mathcal{S}_T$. Specifically, the Gibbs sampling is implemented by simulating the following set of conditional distributions:

1. $\pi\left(\mathcal{S}_T | \Theta, \mathbf{p}, \mathcal{Y}_T\right)$

2. $\pi\left(\Theta, | \mathcal{Y}_T, \mathbf{p}, \mathcal{S}_T\right)$

3. $\pi\left(\mathbf{p} | \mathcal{S}_T\right)$

where we have used the identity $\pi\left(\Theta, \mathbf{p} | \mathcal{S}_T, \mathcal{Y}_T\right) = \pi\left(\Theta | \mathcal{Y}_T, \mathbf{p}, \mathcal{S}_T\right) \pi\left(\mathbf{p} | \mathcal{S}_T\right)$, noting that under our assumptions $\pi\left(\mathbf{p} | \Theta, \mathcal{S}_T, \mathcal{Y}_T\right) = \pi\left(\mathbf{p} | \mathcal{S}_T\right)$.

The simulation of the states $\mathcal{S}_T$ requires 'forward' and 'backward' passes through the data. Define $\mathcal{S}_t = (s_1, ..., s_t)$ and $\mathcal{S}^{t+1} = (s_{t+1}, ..., s_T)$ as the state history up to time t and from time t to T, respectively. We partition the states' joint density as follows:

$$p(s_{T-1} | \mathcal{Y}_T, s_T, \Theta, \mathbf{p}) \times \cdots \times p(s_t | \mathcal{Y}_T, \mathcal{S}^{t+1}, \Theta, \mathbf{p}) \times \cdots \times p(s_1 | \mathcal{Y}_T, \mathcal{S}^2, \Theta, \mathbf{p}).$$
$$(A1)$$

Chib (1995) shows that the generic element of (24) can be decomposed as

$$p(s_t | \mathcal{Y}_T, \mathcal{S}^{t+1}, \Theta, \mathbf{p}) \propto p(s_t | \mathcal{Y}_T, \Theta, \mathbf{p}) p(s_t | s_{t-1}, \Theta, \mathbf{p}), \qquad (A2)$$

where the normalizing constant is easily obtained since s_t takes only two values conditional on the value taken by s_{t+1}. The second term in (A2) is simply the transition probability from the Markov chain. The first term can be computed by a recursive calculation (the forward pass through the data) where,

for given $p(s_{t-1}|\mathcal{Y}_{t-1},\Theta,\mathbf{p})$, we obtain $p(s_t|\mathcal{Y}_t,\Theta,\mathbf{p})$ and $p(s_{t+1}|\mathcal{Y}_{t+1},\Theta,\mathbf{p})$, ..., $p(s_T|\mathcal{Y}_t,\Theta,\mathbf{p})$. Suppose $p(s_{t-1}|\mathcal{Y}_{t-1},\Theta,\mathbf{p})$ is available, then

$$p(s_t = k|\mathcal{Y}_t,\Theta,\mathbf{p}) = \frac{p(s_t = k|\mathcal{Y}_{t-1},\Theta,\mathbf{p}) \times f(y_t|\mathcal{Y}_{t-1},\boldsymbol{\theta}_k)}{\sum\limits_{l=k-1}^{k} p(s_t = l|\mathcal{Y}_{t-1},\Theta,\mathbf{p}) \times f(y_t|\mathcal{Y}_{t-1},\boldsymbol{\theta}_l)},$$

where, for $k = 1, 2, ..., K+1$,

$$p(s_t = k|\mathcal{Y}_{t-1},\Theta,\mathbf{p}) = \sum_{l=k-1}^{k} p_{lk} \times p(s_{t-1} = l|\mathcal{Y}_{t-1},\Theta,\mathbf{p}),$$

and p_{lk} is the Markov transition probability.

For a given set of simulated states, $\mathcal{S}_T$, the data is partitioned into $K+1$ groups. To obtain the conditional distributions for the regression parameters, prior locations and scales, note that the conditional distribution of the $\boldsymbol{\beta}_j$'s are mutually independent with

$$\boldsymbol{\beta}_j\Big|\, \sigma_j^2, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T \sim N\left(\overline{\boldsymbol{\beta}}_j, \overline{V}_j\right),$$

where

$$\overline{\mathbf{V}}_j = \left(\sigma^{-2}\mathbf{X}_j'\mathbf{X}_j + \mathbf{B}_0^{-1}\right)^{-1}, \quad \overline{\boldsymbol{\beta}}_j = \overline{\mathbf{V}}_j\left(\sigma^{-2}\mathbf{X}_j'\mathbf{y}_j + \mathbf{B}_0^{-1}\mathbf{b}_0\right),$$

$\mathbf{X}_j$ is the matrix of observations on the regressors in regime j, and y_j is the vector of observations on the dependent variable in regime j.

Defining $\boldsymbol{\beta}_{1:k+1} = \left(\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_{K+1}'\right)'$ and $\boldsymbol{\sigma}_{1:k+1}^2 = \left(\sigma_1^2, ..., \sigma_{K+1}^2\right)'$, the densities of the location and scale parameters of the regression parameter meta-distribution, $\mathbf{b}_0$ and $\mathbf{B}_0$, can be written

$$\mathbf{b}_0|\,\boldsymbol{\beta}_{1:k+1}, \boldsymbol{\sigma}_{1:k+1}^2, \mathbf{B}_0, v_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T \sim N\left(\overline{\boldsymbol{\mu}}_\beta, \overline{\boldsymbol{\Sigma}}_\beta\right)$$

$$\mathbf{B}_0^{-1}\Big|\,\boldsymbol{\beta}_{1:k+1}, \boldsymbol{\sigma}_{1:k+1}^2, \mathbf{b}_0, v_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T \sim W\left(\overline{v}_\beta, \overline{\mathbf{V}}_\beta^{-1}\right),$$

where

$$\overline{\boldsymbol{\Sigma}}_\beta = \left(\underline{\boldsymbol{\Sigma}_\beta^{-1}} + (K+1)\mathbf{B}_0^{-1}\right)^{-1}$$

$$\overline{\boldsymbol{\mu}}_\beta = \overline{\boldsymbol{\Sigma}}_\beta\left(\mathbf{B}_0^{-1}\sum_{j=1}^{J}\boldsymbol{\beta}_j + \underline{\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta}\right),$$

and

$$\overline{v}_\beta = \underline{v}_\beta + (K+1)$$

$$\overline{\mathbf{V}}_\beta = \sum_{j=1}^{J} \left(\boldsymbol{\beta}_j - \mathbf{b}_0\right)\left(\boldsymbol{\beta}_j - \mathbf{b}_0\right)' + \underline{\mathbf{V}}_\beta.$$

Moving to the posterior for the precision parameters within each regime, note that

$$\sigma_j^{-2}\big|\,\boldsymbol{\beta}_j, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T \sim G\left(\frac{v_0 + \sum\limits_{i=\tau_{j-1}+1}^{\tau_j} \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i\right)'\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i\right)}{2}, \frac{d_0 + n_j}{2}\right),$$

where n_j is the number of observations assigned to regime j. The location and scale parameters for the error term precision are then updated as follows:

$$v_0|\,\boldsymbol{\beta}_{1:k+1}, \sigma_{1:k+1}^{-2}, \mathbf{b}_0, \mathbf{B}_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T \propto \prod_{j=1}^{K+1} G\left(\sigma_j^{-2}\big|\,\underline{v}_0, \underline{d}_0\right) Exp\left(v_0|\,\underline{\rho}_0\right)$$

$$\text{(A3)}$$

$$d_0|\,\boldsymbol{\beta}_{1:k+1}, \sigma_{1:k+1}^{-2}, \mathbf{b}_0, \mathbf{B}_0, v_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T \sim G\left(\underline{v}_0\left(K+1\right) + \underline{c}_0, \sum_{j=1}^{K+1} \sigma_j^{-2} + \underline{d}_0\right)$$

Drawing v_0 from (A2) is slightly more complicated since we cannot make use of any standard distributions. We therefore introduce a Metropolis-Hastings (M-H) step in the Gibbs sampling algorithm. At each loop of the Gibbs sampling we draw a value v_0^* from a Gamma distributed candidate generating density of the form

$$q\left(v_0^*|\,v_0^{g-1}\right) \sim G\left(\varsigma, \varsigma/v_0^{g-1}\right).$$

This candidate generating density is centered on the last accepted value of v_0 in the chain, v_0^{g-1}, while the parameter ς defines the variance of the density and, as a by-product, the rejection in the M-H step. Higher values of ς mean a smaller variance for the candidate generating density and thus a smaller rejection rate. The acceptance probability is given by

$$\xi\left(v_0^*|\,v_0^{g-1}\right) = \min\left[\frac{\pi\left(v_0^*|\,\boldsymbol{\beta}, \sigma^{-2}, \mathbf{b}_0, \mathbf{B}_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T\right)/q\left(v_0^*|\,v_0^{g-1}\right)}{\pi\left(v_0^{g-1}\big|\,\boldsymbol{\beta}, \sigma^{-2}, \mathbf{b}_0, \mathbf{B}_0, d_0, \mathbf{p}, \mathcal{S}_T, \mathcal{Y}_T\right)/q\left(v_0^{g-1}\big|\,v_0^*\right)}, 1\right].$$

$$\text{(A4)}$$

With probability $\xi\left(v_0^*|\, v_0^{g-1}\right)$ the candidate value v_0^* is accepted as the next value in the chain, while with probability $\left(1-\xi\left(v_0^*|\, v_0^{g-1}\right)\right)$ the chain remains at v_0^{g-1}. The acceptance ratio penalizes and rejects values of v_0 drawn from low posterior density areas.

Finally, $\mathbf{p}$ is easily simulated from $\pi\left(\mathbf{p}|\, \mathcal{S}_T\right)$ since, under the beta prior in (2) and given the simulated states, $\mathcal{S}_T$, the posterior distribution of p_{ii} is $Beta(\underline{a}+n_{ii}, \underline{b}+1)$ where n_{ii} is the number of one step transitions from state i to state i in the sequence $\mathcal{S}_n$.

Appendix B. Estimation of Break Point Model

This appendix provides details of how we implement the Chib (1995) method for comparing models with different numbers of break points and how we compute the different components of (8).

Consider the points $(\Theta^*, \mathbf{p}^*)$ in $(\Theta, \mathbf{p})$, which could be maximum likelihood estimates or posterior means or modes. The likelihood function evaluated at Θ^* and $\mathbf{p}^*$ is available from the proposed parameterization of the change point model and can be obtained as

$$\log f\left(\mathcal{Y}_T|\, \Theta^*, \mathbf{p}^*\right) = \sum_{t=1}^{T} \log f\left(y_t|\, \mathcal{Y}_{t-1}, \Theta^*, \mathbf{p}^*\right),$$

where the one-step-ahead predictive density is

$$f\left(y_t|\, \mathcal{Y}_{t-1}, \Theta^*, \mathbf{p}^*\right) = \sum_{k=1}^{K+1} f\left(y_t|\, \mathcal{Y}_{t-1}, \Theta^*, \mathbf{p}^*, s_t = k\right) p\left(s_t = k|\, \mathcal{Y}_{t-1}, \Theta^*, \mathbf{p}^*\right).$$

For simplicity we suppressed the model indicator. The prior density evaluated at the posterior means or modes is easily computed since it is known in advance. The denominator of (8) needs some explanation, however. We can decompose the posterior density as

$$\pi\left(\Theta^*, P^*|\, \mathcal{Y}_T\right) = \pi\left(\Theta^*|\, \mathcal{Y}_T\right)\pi\left(\mathbf{p}^*|\, \Theta^*, \mathcal{Y}_T\right),$$

where

$$\pi\left(\Theta^*|\, \mathcal{Y}_T\right) = \int \pi\left(\Theta^*|\, \mathcal{Y}_T, \mathcal{S}_T\right) p\left(\mathcal{S}_T|\, \mathcal{Y}_T\right) d\mathcal{S}_T,$$

and

$$\pi\left(\mathbf{p}^*|\, \Theta^*, \mathcal{Y}_T\right) = \int \pi\left(\mathbf{p}^*|\, \mathcal{S}_T\right) \pi\left(\mathcal{S}_T|\, \Theta^*, \mathcal{Y}_T\right) d\mathcal{S}_T.$$

The first part can be estimated as $\widehat{\pi}\left(\Theta^{*}|\,\mathcal{Y}_{T}\right) = G^{-1}\sum_{g=1}^{G}\pi\left(\Theta^{*}|\,\mathcal{Y}_{T},\mathcal{S}_{T,g}\right)$ using G draws from the run of the Markov Chain Monte Carlo algorithm. The second part $\pi\left(\mathbf{p}^{*}|\,\Theta^{*},\mathcal{Y}_{T}\right)$ requires an additional simulation of the Gibbs sampler for $[\mathcal{S}_{T,g}]_{g=1}^{G}$ from $\pi\left(\mathcal{S}_{T}|\,\Theta^{*},\mathcal{Y}_{T}\right)$. These draws are obtained by adding steps at the end of the original Gibbs sampling in order to simulate $\mathcal{S}_{T}$ conditional on $(\mathcal{Y}_{T},\Theta^{*},\mathbf{p}^{*})$ and $\mathbf{p}^{*}$ conditional on $(\mathcal{Y}_{T},\Theta^{*},\mathcal{S}_{T})$.

The idea outlined above is easily extended to the case where the Gibbs sampler divides Θ into B blocks, i.e. $\Theta = (\Theta_{1},\Theta_{2},...,\Theta_{B})$. Since

$$\pi\left(\Theta^{*}|\,\mathcal{Y}_{T}\right) = \pi\left(\Theta_{1}^{*}|\,\mathcal{Y}_{T}\right)\pi\left(\Theta_{2}^{*}|\,\Theta_{1}^{*},\mathcal{Y}_{T}\right)\cdots\pi\left(\Theta_{B}^{*}|\,\Theta_{1}^{*},...,\Theta_{B-1}^{*},\mathcal{Y}_{T}\right),$$

we can use different Gibbs sampling steps to calculate the posterior $\pi\left(\Theta^{*}|\,\mathcal{Y}\right)$. In our example we have $\Theta_{1} = \boldsymbol{\beta}_{j}$, $\Theta_{2} = \sigma_{j}^{-2}$ $(j = 1,...,K+1)$, $\Theta_{3} = \mathbf{b}_{0}$, $\Theta_{4} = \mathbf{B}_{0}$, $\Theta_{5} = v_{0}$ and $\Theta_{6} = d_{0}$. The Chib method can become computationally demanding, but the various sampling steps all have the same structure. For some of the blocks in the hierarchical Hidden Markov Chain model, the full conditional densities are non-standard, and sampling requires the use of the Metropolis-Hastings algorithm (see for example the precision prior hyperparameter v_{0}). The original Chib 1995 algorithm is then modified following Chib and Jeliazkov (2001).

Figure 1: Monthly T-Bill rates, 1947:7 - 2002:12.

Figure 2: Posterior probability of break occurrence for the AR(1) model, assuming $K = 6$ breaks.

Figure 3: Posterior density of the meta distribution parameters b_0 under the unconstrained (dashed line) and constrained (solid line) models.

Figure 4: Posterior probability of staying in regime $K + 1$ at time $T + h$, $\Pr\left(s_{T+h} = K + 1 \mid s_T = K + 1\right)$. h is the forecast horizon and ranges from 1 to 60 months.

Figure 5: Predictive densities under three models for the T-Bill series. The graphs show the predictive distributions for the T-Bill series under various forecast horizons. The dotted line represents the forecast assuming no break point in the new data by using only the information from the last regime (assuming $K=6$), the solid line represents the 'meta-distribution' forecast under the assumption of a new break occuring immediately after the end of the estimation sample, while the dash-dotted line is the predictive density from the composite model.

Figure 6: Composite predictive densities under Bayesian model averaging (dashed dotted line) and under the model with six breaks (solid line).

			AR(1) model			
No. of breaks	Log lik.(LL)	Marginal LL		Break dates		
0	-455.05	-485.93				
1	-339.81	-433.86	Dec-69			
2	-271.49	-339.18	Dec-69	Jun-85		
3	-226.17	-318.50	Nov-69	Oct-79	Oct-82	
4	-190.13	-336.56	Jun-53	Nov-69	Oct-79	Oct-82
5	-170.84	-331.28	Jun-53	Nov-69	Oct-79	Oct-82
			Jul-89			
6	-135.61	-310.87	Nov-57	Jul-60	Sep-66	Oct-79
			Oct-82	Jul-89		
7	-128.46	-312.37	Nov-57	Jul-60	Dec-65	Nov-69
			Oct-79	Oct-82	Jul-89	

			AR(2) model			
# of breaks	Log lik.	Marg log lik.		Break dates		
0	-449.57	-514.97				
1	-514.79	-449.92	Nov-57			
2	-320.58	-498.99	Nov-53	Nov-69		
3	-213.18	-418.53	Nov-57	Oct-69	Sep-82	
4	-186.63	-412.14	Jun-53	Nov-69	Oct-79	Sep-82
5	-172.81	-420.75	Nov-57	Jul-60	Dec-68	Oct-82
			Jul-89			
6	-128.45	-395.46	Nov-57	Jul-60	Sep-66	Oct-79
			Sep-82	Jul-89		
7	-120.94	-398.77	Nov-57	Jul-60	Dec-65	Oct-68
			Oct-79	Sep-82	Jul-89	

Table 1: Model comparison. This table shows the log likelihood and the marginal log likelihood estimates for different numbers of breaks along with the time of the break points for the different models. The top and bottom panels display results for the AR(1) and AR(2) models, respectively.

Parameter estimates							
Regimes							
	1	2	3	4	5	6	7
Constant							
Mean	0.020	0.267	0.020	0.224	0.396	0.247	0.121
s.e.	0.032	0.221	0.071	0.149	0.498	0.209	0.072
AR(1) coefficient							
Mean	1.002	0.886	1.005	0.968	0.958	0.967	0.971
s.e.	0.018	0.075	0.022	0.024	0.043	0.027	0.013
Variances							
Mean	0.023	0.269	0.016	0.260	2.525	0.160	0.039
s.e.	0.003	0.077	0.003	0.029	0.687	0.027	0.006
Transition Probability matrix							
Mean	0.988	0.958	0.980	0.990	0.965	0.981	1
s.e.	0.010	0.034	0.016	0.007	0.028	0.015	0
Mean dur.	120	37	72	156	37	84	99

Table 2: Posterior parameter estimates for the unconstrained AR(1) hierarchical Hidden Markov Chain model with six break points.

Mean Parameters				
	Mean	s.e.	95% conf interval	
$b_0(1)$	0.191	0.211	-0.209	0.633
$b_0(2)$	0.944	0.161	0.619	1.220
Variance Parameters				
	Mean	s.e.		
$B_0(1,1)$	0.273	0.312		
$B_0(2,2)$	0.198	0.147		
Error term precision				
	Mean	s.e.	95% conf interval	
v_0	0.775	0.344	0.260	1.500
d_0	0.043	0.024	0.010	0.101

Table 3: Prior parameter estimates for the unconstrained AR(1) hierarchical Hidden Markov Chain model with six break points.

Parameter estimates						
Regimes						
1	2	3	4	5	6	7
Constant						
Mean	0.023	0.300	0.009	0.208	0.572	0.290	0.108
s.e.	0.034	0.225	0.069	0.156	0.593	0.212	0.069
AR(1) coefficient						
Mean	1.044	1.040	0.950	0.934	1.153	1.098	1.267
s.e.	0.091	0.161	0.112	0.080	0.153	0.108	0.092
AR(2) coefficient						
Mean	-0.045	-0.162	0.059	0.036	-0.208	-0.136	-0.292
s.e.	0.093	0.160	0.115	0.081	0.156	0.109	0.090
Variances						
Mean	0.023	0.2591	0.015	0.260	2.468	0.160	0.035
s.e.	0.003	0.071	0.003	0.031	0.646	0.027	0.005
Transition Probability matrix						
Mean	0.988	0.959	0.980	0.990	0.961	0.982	1
s.e.	0.010	0.033	0.016	0.008	0.031	0.014	0

Table 4: Posterior parameter estimates for the unconstrained AR(2) hierarchical Hidden Markov Chain model with six break points.

Mean Parameters				
	Mean	s.e.	95% conf interval	
$b_0(1)$	0.215	0.259	-0.272	0.761
$b_0(2)$	1.072	0.213	0.648	1.490
$b_0(3)$	-0.109	0.210	-0.522	0.308
Variance Parameters				
	Mean	S.e.		
$B_0(1,1)$	0.405	0.443		
$B_0(2,2)$	0.289	0.261		
$B_0(3,3)$	0.295	0.292		
Error term precision				
	Mean	S.e.	95% conf interval	
v_0	0.872	0.359	0.340	1.674
d_0	0.047	0.025	0.011	0.109

Table 5: Prior parameter estimates for the unconstrained AR(2) hierarchical Hidden Markov Chain model with six break points.

Parameter estimates							
Regimes							
	1	2	3	4	5	6	7
$\alpha_j \phi_j$							
Mean	0.029	0.293	0.055	0.241	0.534	0.281	0.126
s.e.	0.028	0.176	0.053	0.138	0.424	0.167	0.062
$1 - \phi_j$							
Mean	0.991	0.882	0.990	0.965	0.947	0.963	0.970
s.e.	0.011	0.060	0.012	0.022	0.036	0.022	0.012
$\Pr\left(\phi_j = 0\right)$	0.374	0.010	0.361	0.037	0.056	0.016	0.001
Variances							
Mean	0.023	0.256	0.016	0.257	2.526	0.159	0.039
s.e.	0.003	0.071	0.003	0.030	0.653	0.026	0.006
Transition Probability matrix							
Mean	0.988	0.960	0.980	0.990	0.962	0.982	1
s.e.	0.010	0.034	0.017	0.008	0.031	0.014	0
Mean dur.	120	37	72	156	37	84	99

Table 6: Posterior parameter estimates for the AR(1) hierarchical Hidden Markov Chain model with six break points under the constrained parameterization in (19).

Mean Parameters				
	Mean	s.e.	95% conf interval	
$b_0(1)$	0.173	0.198	0	0.650
$b_0(2)$	0.914	0.112	0.627	1
$\Pr\left(b_0(2) = 1\right)$	0.383			
Variance Parameters				
	Mean	s.e.		
$B_0(1,1)$	0.282	0.254		
$B_0(2,2)$	0.189	0.140		
Error term precision				
	Mean	s.e.	95% conf interval	
v_0	0.858	0.359	0.335	1.677
d_0	0.047	0.026	0.012	0.114

Table 7: Prior parameter estimates for the AR(1) hierarchical Hidden Markov Chain model with six break points under the constrained parameterization in (19).

	Violations %	
	1998:01-2002:12	2001:01-2002:12
No break	0.350	0.667
Combined	0.183	0

Table 8: Percentage of times the posterior predictive p-value lies outside the 0.05-0.95 interval under the model assuming no new break points after the end of the sample and the composite hierarchical Hidden Markov chain model.

	Unconstrained	Constrained
	Posterior Mean	
No break	1.575	1.579
Combined	1.366	1.363
	Posterior Mode	
No break	1.607	1.594
Combined	1.288	1.294

Table 9: Root mean squared forecast error for the posterior means and modes of the predictive densities under the model assuming no new break points after the end of the sample and the constrained and unconstrained composite hierarchical Hidden Markov chain models.

# of breaks	Posterior Prob
0	7.67E-77
1	3.15E-54
2	4.14E-13
3	3.97E-04
4	5.69E-12
5	1.12E-09
6	0.817
7	0.182

Table 10: Model Posterior Probabilities for different number of in sample break points. The posterior probabilities are computed by means of the margial log likelihood of table 1.