Bayesian Regime Averaging for Time Series subject to Structural Breaks

Hashem Pesaran
University of Cambridge

Davide Pettenuzzo
Bocconi University

Allan Timmermann
University of California, San Diego

March 17, 2004

Abstract

This paper provides a general framework to forecasting time series subject to discrete structural breaks. We develop the concept of parameter meta-distribution to deal with the possibility of a new break between the end of the sample and the forecast horizon. We build a Bayesian estimation procedure, drawing on the earlier work of Chib (1998) on Bayesian multiple break point detection but generalizing the method to account for the estimation of parameter meta distributions. We also implement a model selection criterion to select how many break points the data supports, and use the results to forecast future realizations conditional on new break and no new break occurring. We apply the methodology to forecast of US Treasury Bill rates.

1 Introduction

Structural change is a common phenomenon among economic and financial time series, c.f. Stock and Watson (1996), Pesaran and Timmermann (2003a, 2003b, 2003c). Although the form of the instability is generally unknown a widely used class of models assumes that it can be well approximated by a sequence of discrete structural breaks. The main problem that arises when forecasting such time series is how to select the optimal window of data needed to estimate the parameters of the forecasting model. The standard solution is to use only observations from the last post-break period ($t = \tau_K + 1, ..., T$), where τ_K is the time of the last break point in the series. In the case where the post-break window is sufficiently long, this will lead to unbiased estimates of the forecasting model. However in many other circumstances this might not be the best strategy to follow. For example, when the sample of observations after the last break is very small or, if dealing with multiple step ahead forecasts, when a new break may happen between the end of the sample and the forecasting time, this strategy can lead to very inaccurate results.

This paper provides a general framework to forecasting a time series under such discrete structural breaks, able to handle the different scenarios that might arise once new breaks in the series are allowed to occur. To account both for the timing and the nature of the breaks, we develop the concept of parameter meta-distribution.

To set up the problem suppose we have a univariate r^{th} order autoregressive model subject to K break points

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

This specification is quite general and allows for intercept and slope shifts, as well as changes in the variances. Each regime j , $j = 1, ...K + 1$, is characterized by a vector of regression coefficient, $\boldsymbol{\beta}_j = \left(\beta_{j,0}, \beta_{j,1}, ...\beta_{j,r}\right)'$, and an error term variance, σ_j^2, for $t = \tau_{j-1} + 1, ..., \tau_j$[1]. In order to forecast y_{T+h} we may distinguish two possibilities. Under the first case, there is no new break between the end of the sample, T, and the forecasting time, $T + h$, and, if the last regime sample size is not too small it will be optimal to forecast y_{T+h} by using $\boldsymbol{\beta}_{K+1}$ and σ_{K+1}^2. Under the second scenario, a new break occurs between y_T and y_{T+h} and using the last regime observation may lead to very inaccurate results. To deal with this possibility, we introduce the notion of parameter meta distribution and we will forecast y_{T+h} by means of

[1] througout the paper we assume that $\tau_0 = 0$ and $\tau_{K+1} = T$.

regression parameters and error term variance drawn from this distribution. This will provide a flexible way of using all the sample information (thus including also those regarding the size and the nature of the breaks) to draw inference on how the new regime parameters will be instead of discarding the observation before the last break point.

In this paper, we build a Bayesian estimation procedure, drawing from the earlier work of Chib (1998) on Bayesian multiple break point detection but generalizing the method to account for the estimation of the parameter meta distributions. We also implement a model selection criteria to select how many break points the data is supporting, and then we use the results to forecast future realizations under the two above mentioned possible scenarios. We apply the methodology to the monthly US Treasury Bill time series from July 1947 to December 2002.

The rest of the paper is organized as follows. Section 2 deals with the model, introducing the hierarchical structure into the Hidden Markov Chain model of Chib (1998) to account for the parameter meta distributions estimation. Section 3 explains how to forecast future realizations under the two scenarios of no new break or new break occurrence after the end of the sample, and section 4 illustrates the data used in the empirical example and the results. Section 5 concludes.

2 The model

Our model is built upon the Hidden Markov Chain formulation of Chib (1998) for the multiple change point problem. This formulation is based on the introduction of a state variable s_t in each time period referred as the state of the system at time t, that takes values on the integers $(1, 2, ..., K+1)$ and indicates the regime from which a particular observation y_t has been drawn. Specifically, $s_t = l$ indicates that y_t has been drawn from $f\left(y_t \middle| X_t, \boldsymbol{\theta}_l\right)$ where $\boldsymbol{\theta}_l = [\boldsymbol{\beta}_l, \sigma_l^2]$ stands for the parameters in regime l.

The variable s_t is modeled as a discrete time, discrete state Markov process with the transition probability matrix constrained so that the model is equivalent to a multiple change point model. In particular, at each point in time, the state variable s_t can either remain in the current state or jump to the next higher state. The one step ahead transition probability is represented as

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

where $p_{ij} = Pr\left(s_t = j | s_{t-1} = i\right)$ is the probability of moving to regime j at time t given that a time $t-1$ the state is i. We assume that the non-zero elements of (2), p_{ii}, are independent from p_{jj}, $j \neq i$, and come from a beta distribution, i.e.

$$p_{ii} \sim Beta\left(\underline{a}, \underline{b}\right) \tag{3}$$

The joint density of P is then given by

$$\pi\left(P\right) = c \prod_{i=1}^{K} p_{ii}^{(\underline{a}-1)} \left(1 - p_{ii}\right)^{(\underline{b}-1)} \tag{4}$$

where $c = \left\{\Gamma\left(\underline{a} + \underline{b}\right) / \Gamma\left(\underline{a}\right) \Gamma\left(\underline{b}\right)\right\}^K$. The parameters $\underline{a}$ and $\underline{b}$[2] may be specified so that they agree with the prior beliefs about the mean duration of each regime[3] or they can be set in order to reflect no information about the regime length.

For the scope of our analysis, we want to extend the original Chib setting and include the notion of parameter meta-distributions, to allow each regime coefficient vector, $\boldsymbol{\beta}_j$ and each regime error term precision, σ_j^{-2}, to come from common distributions, $\boldsymbol{\beta}_j \sim \left(\mathbf{b}_0, \mathbf{B}_0\right)$ and $\sigma_j^{-2} \sim \left(v_0, d_0\right)$ respectively, where $\mathbf{b}_0$ and v_0 are the locations and $\mathbf{B}_0$ and d_0 the scales of the two distributions. The inclusion of a parameter meta-distribution in this setting allows a higher degree of generality, where the pooled scenario (all regime parameters are equal) and the individual one (each regime has different parameters) can be seen as special cases, and in particular will be inferred from the estimates of $\mathbf{B}_0$ and d_0. To allow for this variant in the original model of Chib, we include a hierarchical prior for the regime coefficient $\boldsymbol{\beta}_j$'s and σ_j^{-2}'s. Such a model is often referred in the Bayesian literature as the random coefficient

[2] throughout the paper we use bars under parameters (e.g. $\underline{a}$) to denote parameters of a prior density that has to be specified by the user.

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

model. The hierarchical prior places some structure on the differences between regime coefficients, but at the same time postulates that they all come from a common distribution.

We assume that each regime regression parameter vector, $\boldsymbol{\beta}_j$, is of dimension $(r+1 \times 1)$, where r is the specified lag length, and is a independent draw from a normal distribution, $\boldsymbol{\beta}_j \sim N(\mathbf{b}_0, \mathbf{B}_0)$. The regime error term precisions $h_j = \sigma_j^{-2}$ (we prefer to model the precision parameter because it is easy to deal with its distribution in the hierarchical step) instead are i.i.d. draws from a gamma distribution, i.e. $h_j \sim G(v_0, d_0)$.

We assume that at the lower level, the betas prior mean and precision follow a normal and Wishard distribution (a matrix generalization of the gamma distribution) respectively:

$$\mathbf{b}_0 \sim N\left(\underline{\boldsymbol{\mu}_\beta}, \underline{\boldsymbol{\Sigma}_\beta}\right) \tag{5}$$

$$\mathbf{B}_0^{-1} \sim W\left(\underline{v_\beta}, \underline{\mathbf{V}_\beta^{-1}}\right) \tag{6}$$

where $\underline{\boldsymbol{\mu}_\beta}$, $\underline{\boldsymbol{\Sigma}_\beta}$, $\underline{v_\beta}$ and $\underline{\mathbf{V}_\beta^{-1}}$ are hyperparameters that need to be specified. Finally the regime error term precision hyperparameters v_0 and d_0 follow an exponential and Gamma distribution respectively:

$$v_0 \sim Exp\left(\underline{\rho_0}\right) \tag{7}$$

$$d_0 \sim Gamma\left(\underline{c_0}, \underline{d_0}\right) \tag{8}$$

and $\underline{\rho}$,$\underline{c_0}$ and $\underline{d_0}$ are their hyperparameters.

2.1 Prior elicitation and Posterior inference

We have already specified in the last section the distributions of the quantities of interest of the model. We specified a beta distribution for the diagonal elements of the transition probability matrix in (2), a Normal-Wishard distribution for the meta-distribution parameters of the regression coefficients and a Gamma-Exponential for the error term precision ones. Here we need to assign some values to their hyperparameters. Regarding the p_{ii}'s, we specify $\underline{a} = \underline{b} = 0.5$, i.e. we assign a non informative prior equal for all the diagonal elements of (2). Concerning the Normal-Wishard distribution, we specify $\underline{\boldsymbol{\mu}_\beta} = \mathbf{0}$, $\underline{\boldsymbol{\Sigma}_\beta} = 1000 \times I_{r+1}$, $\underline{v_\beta} = 2$ and $\underline{\mathbf{V}_\beta^{-1}} = I_{r+1}$. $\mathbf{0}$ is a $(r+1 \times 1)$ 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 quite diffuse over sensible ranges of values for both the Normal and the Wishard distribution. In the same line, we also set $\underline{\rho_0} = 0.01$, $\underline{c_0} = 1$ and $\underline{d_0} = 0.01$. Appendix A contains the details of the Gibbs sampler used to simulate our hierarchical Hidden Markov Chain model.

2.2 Model Comparison

In order to assess how many break points the data is supporting, we need to estimate a different model for each sensible number of break points and then compare the results. We index each model by the number of break points estimated, thus we will have M_0,M_1,M_2, and so on. Different approach are available for model comparison. In this section we describe one approach, developed by Chib (1996), particularly suited for model comparisons when the dimensionality of the parameter space is high. This method consent to obtain an estimate for the marginal likelihood of each model $f\left(y|\, M_i\right)$ and permit to rank the different model according to it.

Chib's key point is that the marginal likelihood of model i

$$f\left(y|\, M_i\right) = \int f\left(y|\, M_i, \Theta, P\right) \pi\left(\Theta, P|\, M_i\right) d\Theta^* dP^*$$

can be rewritten as

$$f\left(y|\, M_i\right) = \frac{f\left(y|\, M_i, \Theta^*, P^*\right) \pi\left(\Theta^*, P^*|\, M_i\right)}{\pi\left(\Theta^*, P^*|\, M_i, y\right)} \tag{9}$$

where Θ^*, P^* can be any point in the parameter space but for convenience it is better to use a point whose posterior density is significantly different from zero, as the maximum likelihood estimates or the parameter posterior means or modes. Appendix B shows in details how the three components of (9) can be computed.

3 Posterior predictive distributions

In this section we deal with forecasting from the estimated model of section 2. After we estimated break points and parameters under the different regimes from the original sample data, we can use them to forecast a new realization after time T. Two possible scenarios arise. Under the first scenario (NNB hereafter), there is no new structural break between the last observation in the sample, y_T, and the new realization y_{T+h}, where h is the forecast horizon. Thus we can forecast y_{T+h} using the estimates from the last regime, $\boldsymbol{\beta}_{K+1}$ and σ^2_{K+1}. Under the second scenario (NB hereafter) instead a new break occurs between the end of the original sample and the new realization. We are not able to use the last regime estimates, $\boldsymbol{\beta}_{K+1}$ and σ^2_{K+1}, so we need to rely on something else. The hierarchical structure of the estimated model allows us to deal with this problem in a straightforward way. One of the assumptions of section 2 was the existence of a meta-distribution both for the regression parameters, $\boldsymbol{\beta}_j \sim (\mathbf{b}_0, \mathbf{B}_0)$, and for the error term precisions, $\sigma_j^{-2} \sim (v_0, d_0)$, and their hyperparameters were estimated from the data. Under this scenario, $\boldsymbol{\beta}_{K+2}$ and σ^2_{K+2} will be drawn from those two meta distributions.

As last step, we need to combine the forecasts under the two previous scenarios to obtain an unique predictive density. The combination weights will be given through an estimate of the probability of staying in the last regime, $p_{K+1,K+1}$[4]. In particular, following the Markov Chain property and conditioning on being in the last regime $K+1$ until time T,

$$Pr\left(s_{T+h} = K+2 \middle| s_T = K+1\right) = \left(1 - p_{K+1,K+1}\right)^h \tag{10}$$

where $n_{K+1,K+1}$ is the number of observations ascribed to regime $K+1$ and can be derived from the model original estimates.

Let's first consider how to forecast a new realization, y_{T+h}, where h is the forecast horizon, under the two mentioned scenarios.

1. There is no change of regime between y_T (the last observation in the sample) and the future data y_{T+h}. This means that the new data comes from the last regime $K+1$ and we can draw $p\left(y_{T+h}\middle| s_{T+h} = K+1, y_T\right)$ from

$$\begin{aligned}&\int\int p\left(y_{T+h}\middle| \boldsymbol{\beta}_{K+1}, \sigma^2_{K+1}, s_{T+h} = K+1, y_T\right)\\ &\times \pi\left(\boldsymbol{\beta}_{K+1}, \sigma^2_{K+1}\middle| y_T\right) d\boldsymbol{\beta}_{K+1} d\sigma^2_{K+1}.\end{aligned}$$

[4]Recall from (2) that in the original model we needed to impose that $p_{K+1,K+1} = 1$, to ensure that once s_t reached regime $K+1$ there was no possibility for the state variable to move further forward. Here we need to adapt the model and remove this assumption to provide an estimate of this probability which will allow us to combine the forecasts under the two scenarios.

Proceed as follows:

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

(b) draw y_{T+h} from the posterior predictive density,

$$y_{T+h} \sim p\left(y_{T+h}\middle| \boldsymbol{\beta}_{K+1}, \sigma^2_{K+1}, s_{T+h} = K+1, y_T\right) \tag{11}$$

2. A new break occurs between y_T and the future realization y_{T+h}. This means that the new data comes from a new regime $K+2$ and we need to draw $p\left(y_{T+h}\middle| s_{T+h} = K+2, y_T\right)$ from

$$\begin{aligned}&\int \cdots \int p\left(y_{T+h}\middle| \boldsymbol{\beta}_{K+2}, \sigma^2_{K+2}, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, s_{T+h} = K+2, y_T\right)\\ &\times\pi\left(\boldsymbol{\beta}_{K+2}, \mathbf{b}_0, \mathbf{B}_0\middle| y_T\right)\\ &\times\pi\left(\sigma^2_{K+2}, v_0, d_0\middle| y_T\right) d\boldsymbol{\beta}_{K+2} d\mathbf{b}_0 d\mathbf{B}_0 d\sigma^2_{K+2} dv_0 dd_0.\end{aligned}$$

Proceed as follows:

(a) draw $\mathbf{b}_0$ from

$$\mathbf{b}_0 \sim \pi\left(\mathbf{b}_0\middle| \boldsymbol{\beta}, \sigma^2, \mathbf{B}_0, v_0, d_0, P, \mathcal{S}_T, y\right)$$

and $\mathbf{B}_0$ from

$$\mathbf{B}_0 \sim \pi\left(\mathbf{B}_0\middle| \boldsymbol{\beta}, \sigma^2, \mathbf{b}_0, v_0, d_0, P, \mathcal{S}_T, y\right),$$

draw v_0 from

$$v_0 \sim \pi\left(v_0\middle| \boldsymbol{\beta}, \sigma^{-2}, \mathbf{b}_0, \mathbf{B}_0, d_0, P, \mathcal{S}_T, y\right)$$

and d_0 from

$$d_0 \sim \pi\left(d_0\middle| \boldsymbol{\beta}, \sigma^{-2}, \mathbf{b}_0, \mathbf{B}_0, v_0, P, \mathcal{S}_T, y\right).$$

(b) draw $\boldsymbol{\beta}_{K+2}$ from $\pi\left(\boldsymbol{\beta}_{K+2}\middle| \mathbf{b}_0, \mathbf{B}_0\right)$ and σ^{-2}_{K+2} from $\pi\left(\sigma^{-2}_{K+2}\middle| v_0, d_0\right)$, which are the prior distribution for $\boldsymbol{\beta}_{K+2}$ and σ^{-2}_{K+2} given the drawn hyperparameters

(c) draw y_{T+1} from the posterior predictive density,

$$y_{T+h} \sim p\left(y_{T+h}\middle| \boldsymbol{\beta}_{K+2}, \sigma^2_{K+2}, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, s_{T+h} = K+2, y_T\right) \tag{12}$$

Second, to get an estimate for $p_{K+1,K+1}$ needed in (10), we combine the information from the last regime with the prior information. As in section 2, we use the prior $p_{K+1,K+1} \sim Beta(\underline{a}, \underline{b})$, thus

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

where $n_{K+1,K+1}$ is the number of observation ascribed to regime $K + 1$.

Combining (11), (12) and (10), we are able to integrate out the $T + h$ state probability s_{T+h} from the predictive densities and then finally compute the posterior predictive density $p\left(y_{T+h} | \, y_T \right)$ as

$$\begin{aligned} p\left(y_{T+h} | \, y_T \right) &= p\left(s_{T+h} = K+1 | \, y_T \right) \times p\left(y_{T+h} | \, y_T, s_{T+h} = K+1 \right) + \\ &\quad p\left(s_{T+h} = K+2 | \, y_T \right) \times p\left(y_{T+h} | \, y_T, s_{T+h} = K+2 \right). \end{aligned} \tag{14}$$

4 Empirical Illustration

To illustrate our approach, we apply the methodology to U.S. Treasury Bill monthly data over the period December 1925, December 2002. The next two paragraphs describe the data used and our main findings.

4.1 Data

The data comes from the Monthly CRSP US Treasury Database, developed by the Center for Research in Security Prices at the Graduate School of Business, University of Chicago. The treasury Bill yields are computed as the average of the ask and bid prices and they are continuously compounded 365 day rates. We use nominal three month risk free rates from July 1947 through December 2002. Data is monthly. We split the data in two groups. We use the observation from the beginning of the sample through December 1997 as the estimation sample and the data from January 1998 through December 2002 as the forecasting sample. Figure 1 plot the data series, dividing the sample from the forecasting period.

4.2 Results

We estimate the Hierarchical Hidden Markov Chain of section 2 with the T-Bill series from 1947:7 to 1997:12 using a AR(1) and AR(2) specifications and then rank the different models by looking at the marginal likelihood computed with the Chib method described in section 2.2. Table 1 reports log likelihoods, marginal likelihoods computed with the Chib method and break dates for different number of break point models. In both panels, the marginal likelihood is maximized for $M = 6$, with a value of -310.87 and -395.46 respectively. We can see for example how the models with $M = 7$ break points, which compared to the previous ones find an additional break around 1968-69, obtain basically the same marginal likelihood, suggesting that the additional break point is not supported by the data. This evidence is both for the AR(1) and for the AR(2) specifications. Figures 2 and 3 plot the posterior probability for the estimated 6 break points for the AR(1) and AR(2) models. All break points are precisely estimated as it can be inferred from the unimodality of their posterior distributions. The models with 7 break points, whose graphs are not showed here, were instead displaying a very wide posterior density for the 1969 break, another evidence against the significance of this structural change. Tables 2, 3, 4 and 5 report the parameter estimates for the 6 break points model both for the AR(1) and AR(2) specifications. Tables 2 and 4 show for each regime the regression parameters, the variance, the probability and the average number of months of staying in that regime. All regimes are close to unit root processes. Error term variance is particularly high for regime 5 (from Oct 1982 to Jul 1989), and quite low for regimes 1 (from Sep 1947 to Nov 1957), 3 (from Jul 1960 to Sep 1966) and 7 (from Jul 1989 to Dec 1997). This last result will be of particular importance to explain the differences when comparing the predictive densities coming from the scenarios illustrated in section 3. Tables 3 and 5 report prior parameter estimates, i.e. the two meta distribution parameters, respectively under the AR(1) and AR(2) models. For example, for the AR(1) specification, the two meta distributions are

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

and

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

From the Gamma distribution properties, the mean of the precision meta distribution will be almost 18 and the standard error around 20.

Conditional on the 6 break point model, we are able to compute the predictive distributions for future realizations of the T-Bill series. We start from the end of the in-sample observation T (December 1997), and compute the predictive distributions for period $T+h$ under the three scenarios described in section 3 (NNB, NB and the combination of the two, C hereafter). h, the forecasting horizon, ranges from 1 month to 5 year, December 2002, the end of the forecasting sample. In the following, we restrict our attention to the AR(1) specification, since predictive distributions under the AR(2) model are very similar.

A plausible concern here is the near unit root in some of the regimes and its possible influence in the multi step ahead forecasts, especially when dealing with the meta distribution forecasts which average across all the regimes. To deal with this problem, we propose an alternative constrained parametrization of the model we have discussed so far. Let's rewrite the AR(1) specification for the T-Bill in each regime as

$$\triangle\, y_{t+1} = \alpha_j \phi_j - \phi_j y_t + \epsilon_{t+1}. \tag{15}$$

where $\epsilon_{t+1} \sim N(0, \sigma_j^2)$ and again the j subscript referes to regime specific parameters. If $\phi_j = 0$, the process follows an unit root while if $0 < \phi_j < 2$, it is a stationary AR(1) model. Furthermore in the case of unit root, there is no drift irrespectively of the value of α_j. The long run mean of the process, under the stationarity assumption, is $\frac{\alpha_j \phi_j}{\phi_j}$.

We run our Hierachical Hidden Markov Chain with this new parametrization, and to avoid explosive roots and negative unconditional mean, we constrain the lagged T-Bill coefficient in each regime and in the meta distriburion,ϕ_j, to lie between 0 and 1 (we rule out counter stationarity, i.e. $1 < \phi_j < 2$), and α_j to be strictly positive. Tables 6 and 7 report the HMC results for the six break points case (the break points found are the same as those from the AR(1) unrestricted model). To be consistent with the previous tables, the regime coefficients and the meta distribution results refers to $\alpha_j \phi_j$ and $1 - \phi_j$. They also shows, for each regime and for the meta distribution lagged T-Bill parameter, the probability of unit root, $\Pr\left(\phi_j = 0\right)$. Regimes 1 and 3 are the more likely non stationary ones, both with a probability slightly higher than one third. This is also reflected in the meta distribution results, $\Pr\left(b_0(2) = 0.3835\right)$. Figure 7 shows the posterior densities for the meta distribution parameters both for the unrestricted and restricted models and clearly points out how the new constraints imposed are binding.

Then we use these results to compute the multi step ahead forecasts. As explained in details in section 3, to obtain the predictive density under the

NNB scenario we use the information from the last estimated regime of the hierarchical HMC model, while under NB we use the meta distributions to draw a new vector of regression parameters and a new error term variance. Scenario C is a combination of the previous two densities using (13) and (14). Figures 4 and 5 plot the NNB (solid line) and NB (dashed dotted line) densities together with their combination C (dashed line). The forecast horizon ranges from one month to 5 years. As expected from the variance estimates in table 2, under NNB the predictive density is very concentrated around its mean, while there is much more uncertainty under the NB scenario, which need to balance for all the differences in the 7 regimes. We also inspect graphically the performance of the forecasts under the different scenarios. In each panel, we plot a straight line corresponding to the realized T Bill value for that period. We can see how NNB clearly outperforms NB untill three years after the end of the estimation sample. After year 2000, the Treasury Bill series is characterized by a big drop in values and possibly a new break point, as confirmed from a separate Hidden Markov Chain run restricted only to the period 1998:01-2002:12. Under this possibility, the NNB forecast would not be optimal anymore and we would expect the NB forecasts obtained from the meta distribution to outperform the NNB ones. This is indeed the case, as our graphical inspection clearly shows. The last three panels in figure 5 display how NNB forecasts are upward biased and unable to capture the big T-Bill drop while the NB forecasts clearly are more accurate.

We also investigate the performance of the NNB and C models with two more metrics. First,we use the posterior predictive p-value approach. The idea behind this method is that, if the model is a reasonable one, the actual observed data should be of the type which is commonly generated by the model. Finding out at what percentile the point y_{T+h} lies in the density $p\left(y_{T+h}\right| y_T, M)$, where M is the model indicator and stands here for NNB or C, is the formal metric used. Namely, after generating the forecasts under the two models, we evaluate, for each forecast horizon, how many time the predicted T-Bill lies below the realized one. If one model yields posterior predictive p-values which are much lower than another, this is evidence against the former model. A common rule of thumbs is to take a posterior predictive p-value of less (more) than 0.05 (0.95) as evidence against a model. Figure 8 and 9 report the posterior predictive p-values for both NNB (solid line) and C (dashed line) models, together with the 90% confidence interval, for the recursive one step ahead and the multi step ahead forecasts respectively. We can see from figure 9 how the NNB line always lies between the 90%

percent interval, while this is not the case for model C after the first half of the out of sample period. In particular, after 4 years from the end of the sample, model C constantly overestimates the T-Bill realizations, when the NNB model instead is far more reliable. Figure 8 reinforce this result. NNB forecasts perform badly in the second part of the out of sample, whereas the NNB forecasts in that period always lie close to 0.5, i.e. providing density forecasts correctly centered around the realized T-Bill. Second, we compute RMSFE (both for the unrestricted and restricted models now) under the NNB and C scenarios for the multi step ahead forecasts, using either the Posterior Mean and the Posterior mode of the density forecast as our guess for the T-Bill future realization. Table 8 again shows how the combined model C is outperforming the NNB one, both in the restricted and unrestricted case, and also shows, altough not very strongly, how the constrained model seems to outperform the unconstrained one.

Finally, figure 6 shows the weights assigned to the NNB model (one minus this quantity is the weight assigned to the meta distribution model) as a function of the forecast horizon. As expected, it is a monotonic decreasing function of the forecast horizon h, implying that higher weight is given to the meta distribution model the more distant we go from the end of the in sample.

One last issue that may be relevant when forecasting the T-Bill rates is the possible ARCH structure of the residuals. In our approach, we dealt with that by letting the innovation variance to vary across regimes so that it effectively follows a step function. We run some regressions on the residuals and on the residuals scaled by the estimated volatility parameter on their lagged values and we found that, even though there is still some ARCH structure in the scaled residuals, after scaling the residuals its R-squared reduces from 16% to 2%, a confirmation that the ARCH structure has been removed from them.

5 Conclusion

In this paper we address the problem of forecasting from a time series subject to multiple structural breaks. The simple approach of drawing inferences only from the last post break period can be the best way to go when the number of observations belonging to the last regime is high enough to get consistent parameter estimates and there is a low probability of a new break. However, in a lot of circumstances this solution may not be of any help. Consider for example the case when the last break point is very close to the end of the sample or the possibility that a new break occurs after the end of the estimation sample. Under these scenarios, this simple approach could deliver quite misleading results.

We propose a new and more general approach that is able to handle these possible complications, drawing from the Bayesian Hidden Markov Chain literature. Namely, we extend the original method of Chib (1998) and develop a hierarchical alternative of it, able to provide parameter meta distribution estimates and forecasts under the above mentioned scenarios. We apply this method to an AR(1) specification for the U.S. monthly T-Bill series from 1947:7 to 1997:12. Using Chib (1995) method for comparing models, we find evidence of 6 structural breaks, the more recent one around the end of the eighties. We then compute density forecasts for the series over the period 1998:01-2002:12 under three different scenarios. First under the case that no new break occurs between the end of the sample and the forecasting period, thus using only the information from the last estimated regime, i.e. circa 1990-1997:12. Then we consider the possibility of a new break after the end of the sample and thus use the meta distribution estimates to infer how the new regime regression coefficients and variance may look like. This corresponds to using the all sample information, but still accounting for the fact that the forecasted value will belong to a new unknown regime. Finally we consider the combination of these two forecasts, where the weights are proportional to the duration of the last estimated regime.

We find evidence of a break in the T-Bill series around the beginning of year 2001 and thus we expect the forecasting density coming from the meta distribution estimates to outperform the classical forecasts under no new break occurrence from this period onward. This is indeed the case, as confirmed from a graphical inspection and from the Bayesian posterior predictive p-values of the density forecasts under the two scenarios.

References

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

[2] Chib, S., 1993, "Bayes regression with autocorrelated errors: a Gibbs sampling approach", *Journal of Econometrics*, 58, 275-294.

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

[4] Chib, S., 1996, "Calculating posterior distribution and modal estimates in Markov mixture models", *Journal of Econometrics*, 75, 79-97.

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

[6] Diebold, F.X., T. Gunther and A. Tay, 1998, "Evaluating density forecasts, with applications to financial risk management", *International Economic Review*, 39, 863-883.

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

[8] George, E. I., U. E. Makov and A. F. M. Smith, 1993, "Conjugate likelihood distributions", *Scandinavian Journal of Statistics,* 20, no. 2, 147-156.

[9] Koop, G., 2003, "Bayesian Econometrics", John *Wiley & Sons Editor,* New York.

[10] Pesaran, M.H. and A. Timmermann (2003a) ”Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks”. *Unpublished manuscript, University of Cambridge and UCSD.*

[11] Pesaran, M.H. and A. Timmermann (2003b) ”Optimal Window Selection under Breaks”. *Unpublished manuscript, University of Cambridge, USC and UCSD.*

[12] Pesaran, M.H. and A. Timmermann (2003c) ”Model Instability and Choice of Observation Window in Autoregressive Models”. *Unpublished manuscript, University of Cambridge and UCSD.*

[13] 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.

[14] Tay, A.S. and K.F. Wallis, 2000, "Density Forecasting: A survey", *Journal of Forecasting*, 19, 235-254.

Appendix A. The Gibbs Sampler

The posterior distribution of interest is $\pi\left(\Theta, P, \mathcal{S}_T \middle| y\right)$, 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 and P is the transition probability matrix in (2). The sampling method works recursively. First the states are simulated conditioned on the data y and the other parameters, and second, the parameter are simulated conditioned on the data and $\mathcal{S}_T$. Specifically, the Gibbs sampling is implemented by simulating the following full conditional distributions:

1. $\pi\left(\mathcal{S}_T \middle| \Theta, P, y\right)$
2. $\pi\left(\Theta, \middle| y, P, \mathcal{S}_T\right)$
3. $\pi\left(P \middle| \mathcal{S}_T\right)$

where we used $\pi\left(\Theta, P \middle| \mathcal{S}_T, y\right) = \pi\left(\Theta \middle| y, P, \mathcal{S}_T\right) \pi\left(P \middle| \mathcal{S}_T\right)$ and $\pi\left(P \middle| \Theta, \mathcal{S}_T, y\right) = \pi\left(P \middle| \mathcal{S}_T\right)$.

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. We can partition the states' joint density as

$$p(s_{T-1}| y, s_T, \Theta, P) \times \cdots \times p(s_t| y, \mathcal{S}^{t+1}, \Theta, P) \times \cdots \times p(s_1| y, \mathcal{S}^2, \Theta, P) \quad \text{(A1)}$$

Chib (1996) shows how the generic element of (A1) can be decomposed as

$$p(s_t| y, \mathcal{S}^{t+1}, \Theta, P) \propto p(s_t| y_t, \Theta, P) p(s_t| s_{t-1}, \Theta, P) \quad \text{(A2)}$$

where the normalizing constant is easily obtained since s_t takes only 2 values conditional on the value taken by s_{t+1}. The second term of (A2) is simply the transition probability from the Markov chain, while the first term can be obtain by a recursive calculation (the forward pass through the data) where given $p(s_{t-1}|\, y_{t-1}, \Theta, P)$, we can obtain $p(s_t|\, y_t, \Theta, P)$ and $p(s_{t+1}|\, y_{t+1}, \Theta, P)$, and so on until $p(s_T|\, y_T, \Theta, P)$. Suppose $p(s_{t-1}|\, y_{t-1}, \Theta, P)$ is available, then

$$p(s_t = k|\, y_t, \Theta, P) = \frac{p(s_t = k|\, y_{t-1}, \Theta, P) \times f(y_t|\, y_{t-1}, \boldsymbol{\theta}_k)}{\sum_{l=k-1}^{k} p(s_t = l|\, y_{t-1}, \Theta, P) \times f(y_t|\, y_{t-1}, \boldsymbol{\theta}_l)}$$

where

$$p(s_t = k|\, y_{t-1}, \Theta, P) = \sum_{l=k-1}^{k} p_{lk} \times p(s_{t-1} = l|\, y_{t-1}, \Theta, P)$$

for $k = 1, 2, ..., K+1$ and p_{lk} is the Markov transition probability.

Next , according to the simulated states $\mathcal{S}_T$ the data splits into $K+1$ groups and the conditional distributions for the regression parameters, prior locations and scales are obtained as follows. The conditional distribution of the $\boldsymbol{\beta}_j$'s are mutually independent, with

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

where

$$\begin{aligned} \overline{\mathbf{V}}_j &= \left(\sigma^{-2} X_j' X_j + \mathbf{B}_0^{-1}\right)^{-1} \\ \overline{\boldsymbol{\beta}}_j &= \overline{\mathbf{V}}_j \left(\sigma^{-2} X_j' y_j + \mathbf{B}_0^{-1} \mathbf{b}_0\right). \end{aligned}$$

For the location and scale parameters of the regression parameter meta-distribution, $\mathbf{b}_0$ and $\mathbf{B}_0$, the relevant densities are

$$\mathbf{b}_0|\, \boldsymbol{\beta}, \sigma^2, \mathbf{B}_0, v_0, d_0, P, \mathcal{S}_T, y \sim N\left(\overline{\boldsymbol{\mu}}_\beta, \overline{\boldsymbol{\Sigma}}_\beta\right)$$

and

$$\mathbf{B}_0^{-1} \big|\, \boldsymbol{\beta}, \sigma^2, \mathbf{b}_0, v_0, d_0, P, \mathcal{S}_T, y \sim W\left(\overline{v}_\beta, \overline{\mathbf{V}}_\beta^{-1}\right)$$

where

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

and

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

Moving to the posterior for each regime precision, we have that

$$\sigma_j^{-2}\middle|\, \boldsymbol{\beta}_j, \mathbf{b}_0, \mathbf{B}_0, v_0, d_0, P, \mathcal{S}_T, y \sim G\left(\frac{v_0 + \sum_{i=\tau_{j-1}+1}^{\tau_j}\left(y_i - X_i\boldsymbol{\beta}_j\right)^2}{2}, \frac{d_0 + n_j}{2}\right).$$

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

$$v_0|\, \boldsymbol{\beta}, \sigma^{-2}, \mathbf{b}_0, \mathbf{B}_0, d_0, P, \mathcal{S}_T, y \propto \sum_{j=1}^{K+1} G\left(\sigma_j^{-2}\middle|\, \underline{v_0}, \underline{d_0}\right) Exp\left(v_0|\, \underline{\rho_0}\right) \tag{A3}$$

$$d_0|\, \boldsymbol{\beta}, \sigma^{-2}, \mathbf{b}_0, \mathbf{B}_0, v_0, P, \mathcal{S}_T, y \sim G\left(\underline{v_0}(K+1) + \underline{c_0}, \sum_{j=1}^{K+1}\sigma_j^{-2} + \underline{d_0}\right)$$

Drawing v_0 from (A3), is slightly more complicated as we cannot make use of any standard distributions. In particular, we proceed introducing a Metropolis-Hastings (M-H) step into the Gibbs sampling algorithm. At each loop of the Gibbs Sampling, we propose a candidate value v_0^* drawn from a gamma distributed candidate generating density of the form

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

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

$$\xi\left(v_0^*\middle|\, v_0^{g-1}\right) = \min\left[\frac{\pi\left(v_0^*\middle|\, \boldsymbol{\beta}, \sigma^{-2}, \mathbf{b}_0, \mathbf{B}_0, d_0, P, \mathcal{S}_T, y\right)/q\left(v_0^*\middle|\, v_0^{g-1}\right)}{\pi\left(v_0^{g-1}\middle|\, \boldsymbol{\beta}, \sigma^{-2}, \mathbf{b}_0, \mathbf{B}_0, d_0, P, \mathcal{S}_T, y\right)/q\left(v_0^{g-1}\middle|\, v_0^*\right)}, 1\right]. \tag{A4}$$

With probability $\xi\left(v_0^* \middle| v_0^{g-1}\right)$ the candidate value v_0^* will be accepted as the next value in the chain, while with probability $\left(1-\xi\left(v_0^* \middle| v_0^{g-1}\right)\right)$ the chain remains at v_0^{g-1}. The acceptance ratio is a fundamental quantity to monitor during the iterations; a value around 50 percent normally implies a well behaved M-H. Its importance is crucial in this step. The candidate generating density is only an approximation to the real posterior distribution of v_0 and its role is to wander freely across the posterior space. The role of the acceptance ratio is instead to penalize and reject values of v_0 which belong to low posterior density areas.

Finally the simulation of P from $\pi\left(P \middle| \mathcal{S}_T\right)$ is easily obtained for all the elements in the main diagonal, p_{ii}. Under the beta prior in (3) 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. Model comparison with Chib (1996) method

In this appendix we want to show the way we implement Chib (1996) method for comparing models with different number of break points. We describe in details the way the different parts of (9) are computed.

The likelihood function at Θ^*, P^* is available from the proposed parametrization of the change point and can be obtained (for notation clarity, hereafter we suppress the model indicator) as

$$\log f\left(y \middle| \Theta^*, P^*\right) = \sum_{t=1}^{n} \log f\left(y_t \middle| Y_{t-1}, \Theta^*, P^*\right),$$

where

$$f\left(y_t \middle| Y_{t-1}, \Theta^*, P^*\right) = \sum_{k=1}^{m} f\left(y_t \middle| Y_{t-1}, \Theta^*, P^*, s_t = k\right) p\left(s_t = k \middle| Y_{t-1}, \Theta^*, P\right)$$

is the one step ahead prediction density. The prior density evaluation at the posterior means or modes is easily computed as the prior density are known in advance while the denominator of (9) needs some explanations. We can decompose the posterior density as

$$\pi\left(\Theta^*, P^* \middle| y\right) = \pi\left(\Theta^* \middle| y\right) \pi\left(P^* \middle| \Theta^*, y\right)$$

where

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

and

$$\pi\left(P^{*}\right|\Theta^{*},y)=\int\pi\left(P^{*}\right|\mathcal{S}_{T})\,\pi\left(\mathcal{S}_{T}\right|\Theta^{*},y)\,d\mathcal{S}_{T}$$

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

The idea outlined above is easily extended to the case where the Gibbs sampler involves blocking Θ into B segments, i.e. $\Theta=(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,...,\boldsymbol{\theta}_B)$. That is we can use the fact that

$$\pi\left(\Theta^{*}\right|y)=\pi\left(\boldsymbol{\theta}_1^{*}\right|y)\,\pi\left(\boldsymbol{\theta}_2^{*}\right|\boldsymbol{\theta}_1^{*},y)\cdots\pi\left(\boldsymbol{\theta}_B^{*}\right|\boldsymbol{\theta}_1^{*},...,\boldsymbol{\theta}_{B-1}^{*},y\big)$$

and use different Gibbs sampler to calculate the all posterior $\pi\left(\Theta^{*}\right|y)$. In our example, we have $\boldsymbol{\theta}_1=\boldsymbol{\beta}_j$, $\boldsymbol{\theta}_2=\sigma_j^{-2}$ $(j=1,...,K+1)$, $\boldsymbol{\theta}_3=\mathbf{b}_0$, $\boldsymbol{\theta}_4=\mathbf{B}_0$, $\boldsymbol{\theta}_5=v_0$ and $\boldsymbol{\theta}_6=d_0$. The Chib method can then become computationally demanding, but the various samplers have all the same structure.

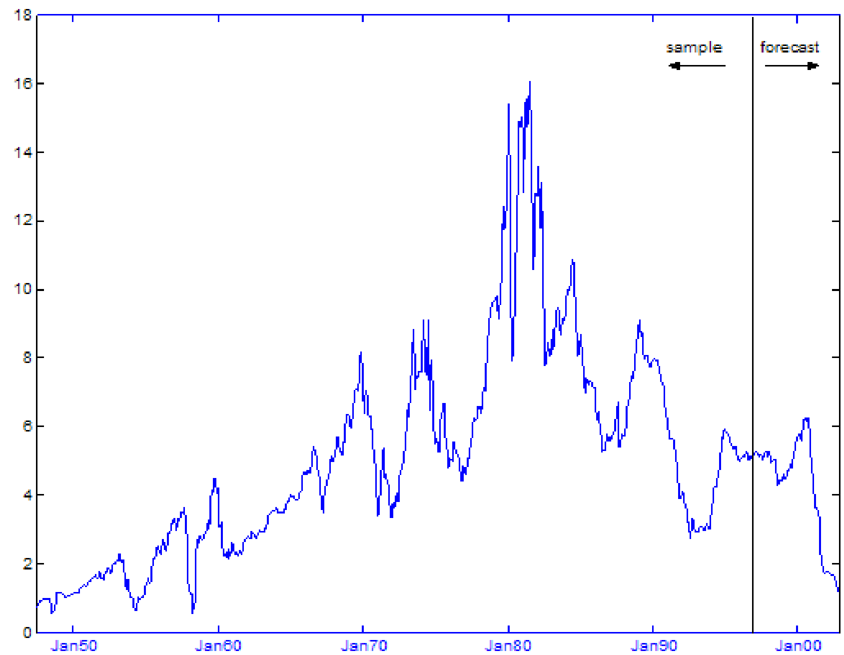

Figure 1: T-Bill series. The graph shows the monthy T-Bill series from 1947:7 to 2002:12. The vertical line delimites the sample observations from the forecast ones. The forecast sample is 1998:01-2002:12.

1 lag model						
# of breaks	Log lik.	Marginal lik.	Break dates			
0	-455.05	-455.40				
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	
2 lag model						
# of breaks	Log lik.	Marginal lik.	Break dates			
0	-449.57	-449.92				
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: Bayesian Model comparison. This table shows the log likelihood and the marginal likelihood estimate with Chib (1995) method for different number of breaks along with the time of the break points for the different models. The top panel displays the results for the AR(1) model while the bottom one shows the AR(2) model results. Estimation period is 1947:7-1997:12.

Parameters estimates							
Regimes							
	1	2	3	4	5	6	7
Constant							
P. Mean	0.0204	0.2668	0.0201	0.2241	0.3958	0.247	0.1208
P. S.e.	0.0322	0.221	0.0716	0.1493	0.4978	0.2091	0.0722
Lag 1							
P. Mean	1.002	0.8862	1.0052	0.9679	0.9581	0.9675	0.9708
P. S.e.	0.0184	0.0747	0.0216	0.0244	0.0431	0.0272	0.0134
Variances							
P. Mean	0.0226	0.2686	0.0158	0.2596	2.5252	0.16	0.039
P. S.e.	0.003	0.0767	0.0032	0.0292	0.687	0.027	0.0057
Transition Probability matrix							
P. Mean	0.988	0.9578	0.9798	0.9904	0.9653	0.9815	1
P. S.e.	0.0101	0.0337	0.0161	0.0075	0.0277	0.015	0
M dur.	120	37	72	156	37	84	99

Table 2: Parameter estimates for the unconstrained AR(1) hierachical Hidden Markov Chain model with 6 break points. Estimation period is 1947:7-1997:12.

Means				
	Post. Mean	Post. Std. err.	95% conf interval	
$b_0(1)$	0.1908	0.2107	-0.2093	0.6328
$b_0(2)$	0.9438	0.1606	0.6195	1.2199
Variances				
	Post. Mean	Post. Std. err.		
$B_0(1,1)$	0.2731	0.3125		
$B_0(2,2)$	0.1981	0.1467		
Error term precision				
	Post. Mean	Post. Std. err.	95% conf interval	
v_0	0.7748	0.344	0.26	1.5002
d_0	0.0431	0.0239	0.0098	0.1014

Table 3: Prior parameter estimates for the unconstrained AR(1) hierachical Hidden Markov Chain model with 6 break points. Estimation period is 1947:7-1997:12.

Parameters estimates							
Regimes							
	1	2	3	4	5	6	7
Constant							
P. Mean	0.0228	0.2989	0.0092	0.2084	0.5724	0.2894	0.108
P. S.e.	0.0344	0.2248	0.0695	0.1557	0.5931	0.2119	0.0686
Lag 1							
P. Mean	1.0444	1.04	0.9498	0.9344	1.1526	1.098	1.2675
P. S.e.	0.0913	0.1611	0.1125	0.0797	0.1534	0.108	0.092
Lag 2							
P. Mean	-0.0447	-0.1623	0.0595	0.0364	-0.2085	-0.1364	-0.2922
P. S.e.	0.093	0.1602	0.1155	0.0809	0.1561	0.1087	0.0901
Variances							
P. Mean	0.0232	0.2591	0.0155	0.2603	2.4676	0.1599	0.0351
P. S.e.	0.0031	0.0715	0.0033	0.0307	0.646	0.0266	0.0051
Transition Probability matrix							
P. Mean	0.988	0.9588	0.9799	0.9904	0.961	0.9824	1
P. S.e.	0.0097	0.0333	0.0163	0.0077	0.031	0.0143	0

Table 4: Parameter estimates for the unconstrained AR(2) hierachical Hidden Markov Chain model with 6 break points. Estimation period is 1947:7-1997:12.

Means				
	Post. Mean	Post. Std. err.	95% conf interval	
$b_0(1)$	0.2154	0.2587	-0.2717	0.7606
$b_0(2)$	1.0718	0.2127	0.6483	1.4903
$b_0(3)$	-0.1094	0.2096	-0.5222	0.3076
Variances				
	Post. Mean	Post. Std. err.		
$B_0(1,1)$	0.4054	0.443		
$B_0(2,2)$	0.2892	0.2609		
$B_0(3,3)$	0.2955	0.2921		
Error term precision				
	Post. Mean	Post. Std. err.	95% conf interval	
v_0	0.8725	0.3587	0.3399	1.6744
d_0	0.0466	0.0254	0.0108	0.1094

Table 5: Prior parameter estimates for the unconstrained AR(2) hierachical Hidden Markov Chain model with 6 break points. Estimation period is 1947:7-1997:12.

Parameters estimates							
	reg1	reg2	reg3	reg4	reg5	reg6	reg7
			$\alpha_j \phi_j$				
P. Mean	0.0286	0.2929	0.0547	0.2413	0.534	0.2808	0.1265
P. S. e.	0.0284	0.1757	0.053	0.1377	0.4245	0.1673	0.0619
			$1 - \phi_j$				
P. Mean	0.9911	0.8819	0.9904	0.9649	0.947	0.9632	0.9699
P. S. e.	0.0112	0.0602	0.0122	0.0217	0.0365	0.0217	0.0118
$\Pr\left(\phi_j = 0\right)$	0.3745	0.0100	0.3615	0.0370	0.0560	0.0160	0.0005
Variances							
P. Mean	0.0229	0.2557	0.0157	0.2575	2.5259	0.1586	0.0393
P. S. e.	0.003	0.0709	0.0035	0.0299	0.6534	0.0261	0.006
Transition Probability matrix							
P. Mean	0.9879	0.959	0.9797	0.9905	0.9623	0.9819	1
P. S. e.	0.0098	0.0336	0.0168	0.0077	0.031	0.0143	0
M. dur.	120	37	72	156	37	84	99

Table 6: Parameter estimates for the AR(1) hierachical Hidden Markov Chain model with 6 break points under the constrained parametrization in (15). Estimation period is 1947:7-1997:12.

Means				
	Post. Mean	Post. Std. err.	95% conf interval	
$b_0(1)$	0.173	0.1977	0	0.6504
$b_0(2)$	0.9138	0.112	0.6268	1
$\Pr\left(b_0(2) = 1\right)$	0.3835			
Variances				
	Post. Mean	Post. Std. err.		
$B_0(1,1)$	0.282	0.2536		
$B_0(2,2)$	0.1891	0.1403		
Error term precision				
	Post. Mean	Post. Std. err.	95% conf interval	
v_0	0.8585	0.3588	0.3348	1.6768
d_0	0.0475	0.0264	0.0117	0.1139

Table 7: Prior parameter estimates for the AR(1) hierachical Hidden Markov Chain model with 6 break points under the constrained parametrization in (15). Estimation period is 1947:7-1997:12.

	Unconstrained	Constrained
	Posterior Mean	
NNB	1.5701	1.5742
C	1.3734	1.3728
	Posterior Mode	
NNB	1.6157	1.6082
C	1.3314	1.3028

Table 8: RMSFE for the posterior means and modes of the predictive densities under scenario NNB (no new break point after the end of the sample) and C (Combined case) both for the unconstrained and constrained model for the T-Bill

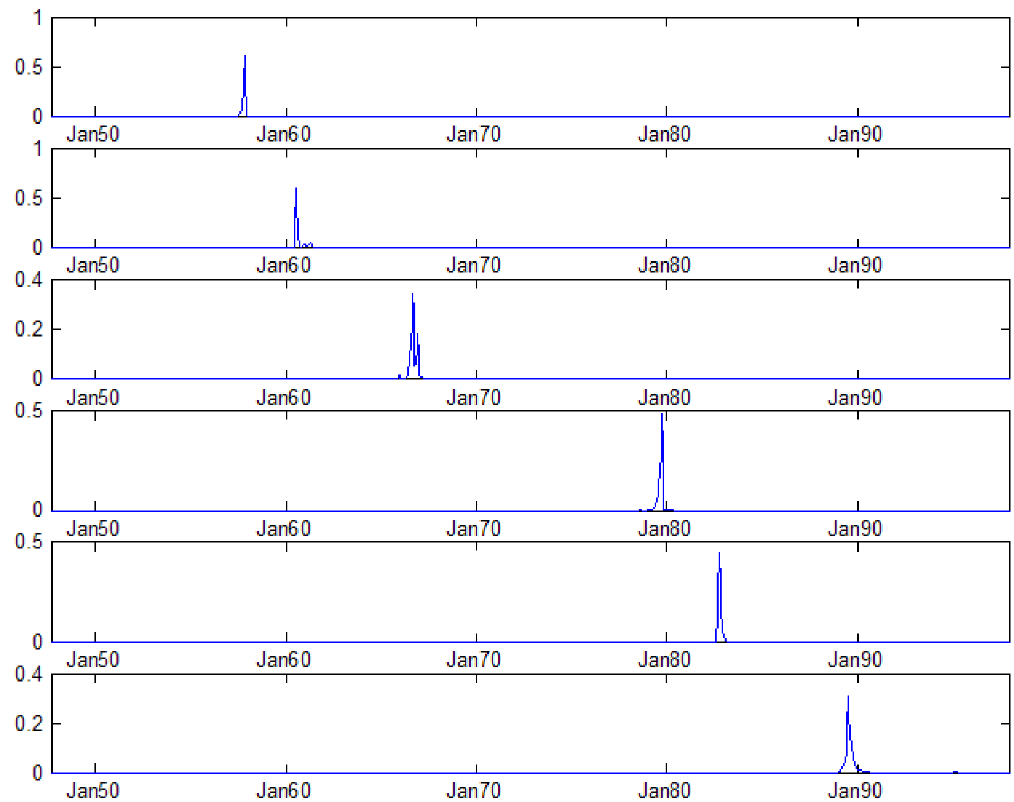

Figure 2: Posterior probability of break occurance for the AR(1) model. Each subgraph show the posterior probability of break occurance as computed from the output of the hierachical hidden markov chain model of section 2. The number of breaks is set equal to 6.

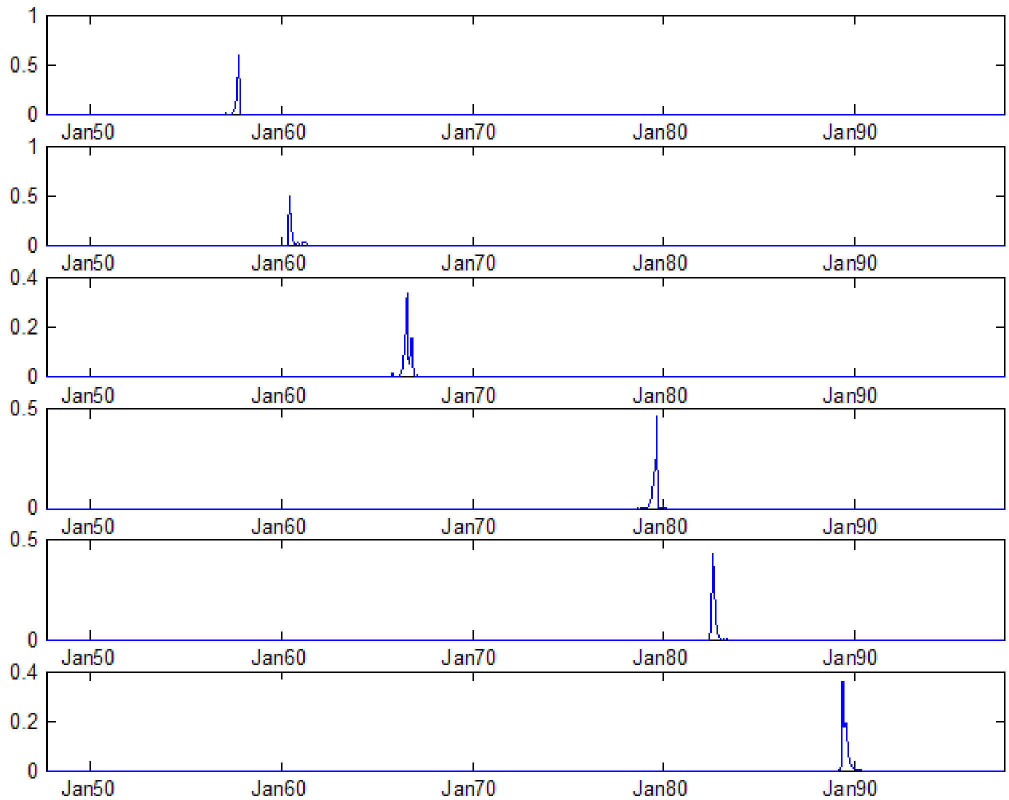

Figure 3: Posterior probability of break occurance for the AR(2) model. Each subgraph show the posterior probability of break occurance as computed from the output of the hierachical hidden markov chain model of section 2. The number of breaks is set equal to 6.

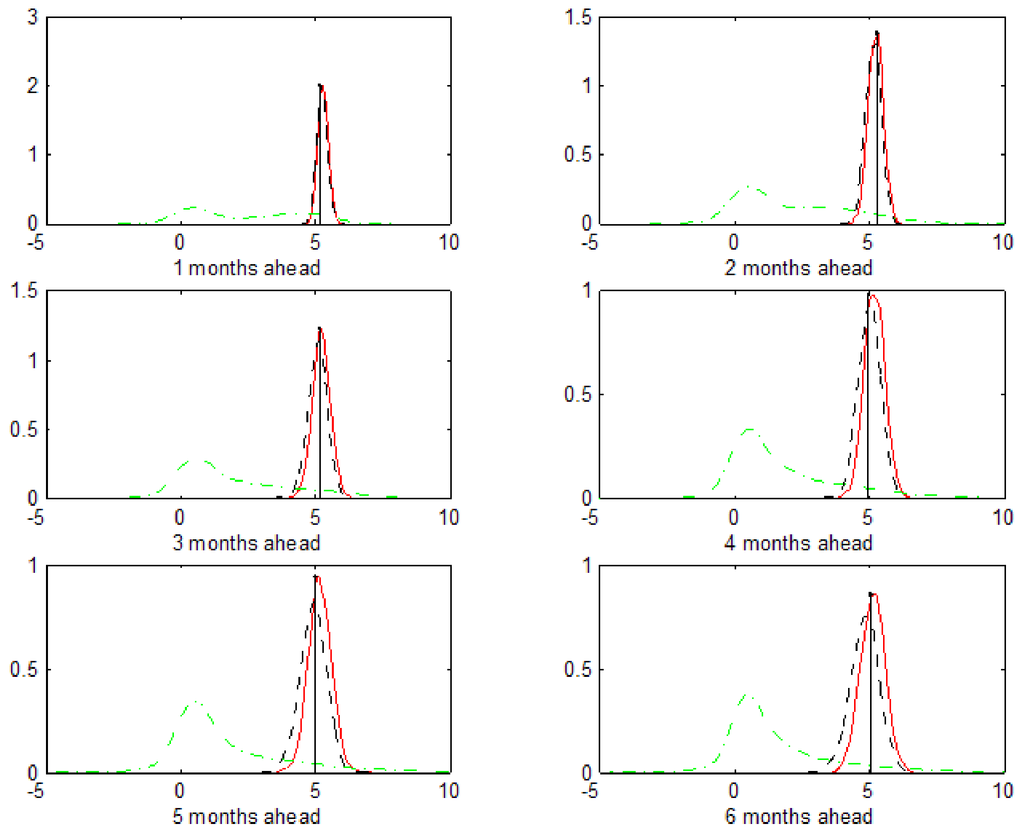

Figure 4: Predictive densities under three different models for the T-Bill series. This graphs shows the predictive distributions for the T-Bill series from 1 to 6 months ahead. The dotted line represents the forecast under the most supported model (M=6) and no break point in the new data by using only the information from the last regime (NNB), the solid line represents the forecast under the assumption of a new break after the end of the estimation sample and uses the meta-distribution information (NB) while the dash-dotted line is the combination of the two previous predictive densities (C) as in (14). Estimation period is 1947:1-1997:12.

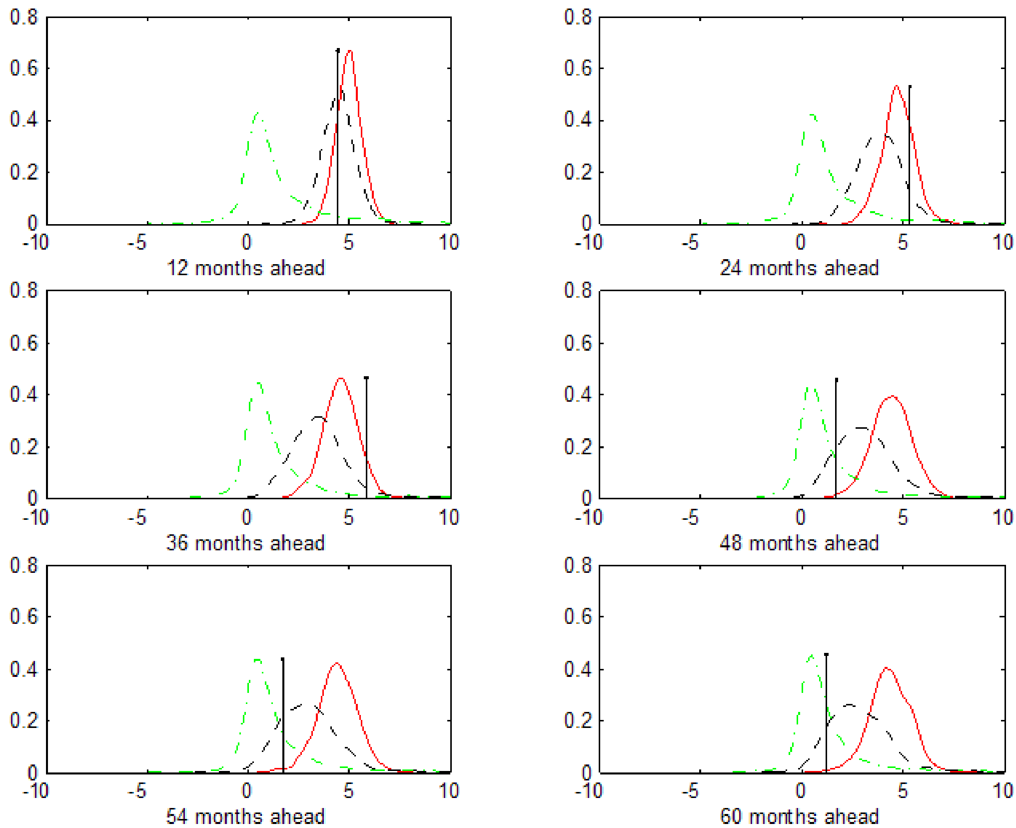

Figure 5: Predictive densities under three different models for the T-Bill series. This graphs shows the predictive distributions for the T-Bill series from 12 to 60 months ahead. The dotted line represents the forecast under the most supported model (M=6) and no break point in the new data by using only the information from the last regime (NNB), the solid line represents the forecast under the assumption of a new break after the end of the estimation sample and uses the meta-distribution information (NB) while the dash-dotted line is the combination of the two previous predictive densities (C) as in (14). Estimation period is 1947:1-1997:12.

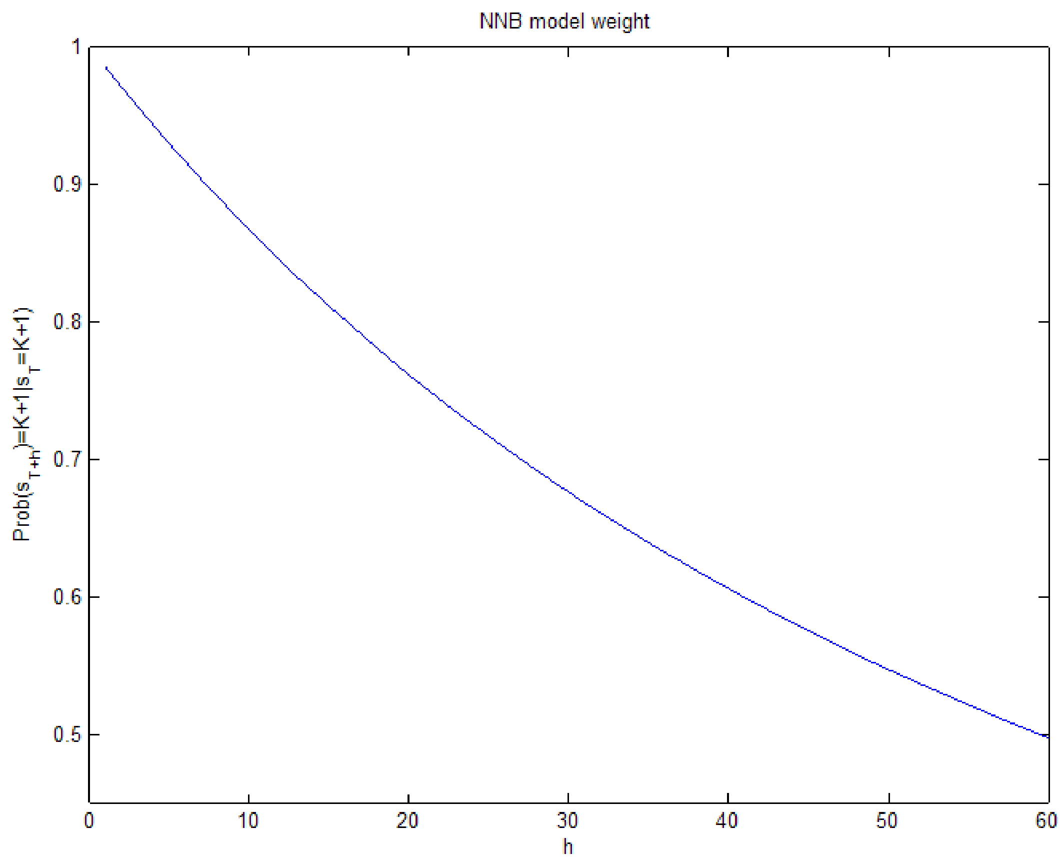

Figure 6: Posterior probability of staying in regime $K+1$ in time $T+h$. h is the forecast horizon and ranges from 1 to 60 months after the end of the sample. The graph shows the posterior probability of staying in the last regime after the end of the sample, $\Pr\left(s_{T+h} = K+1 \middle| s_T = K+1\right)$.

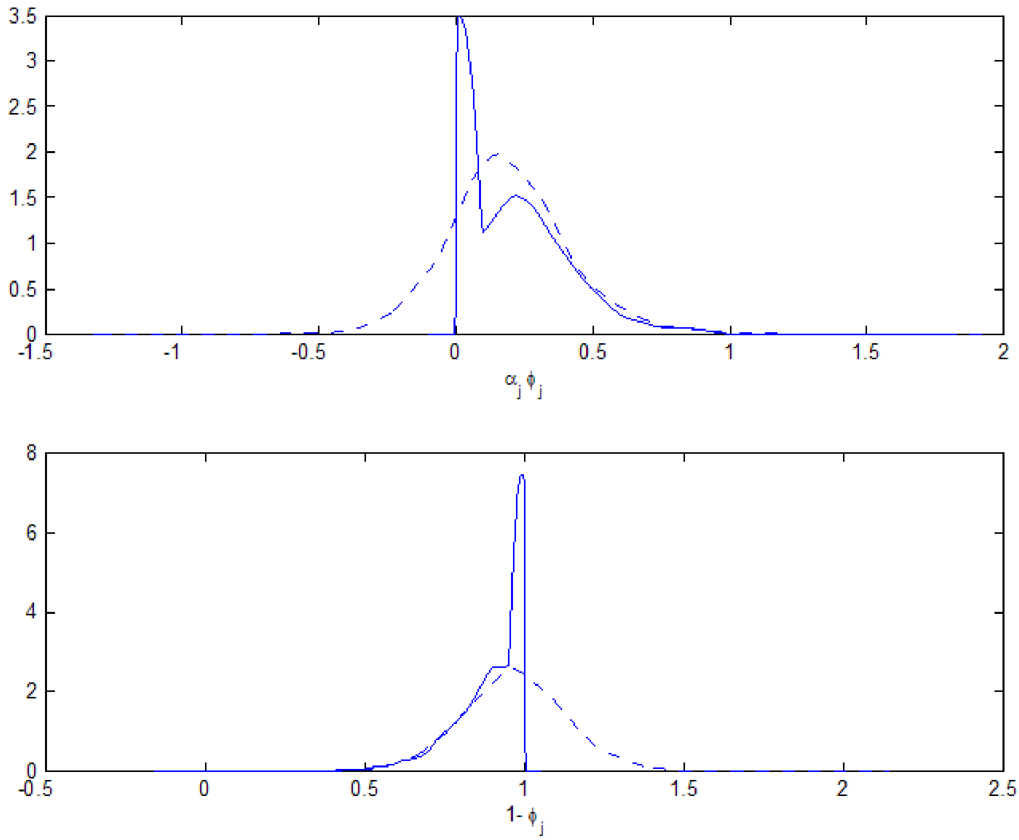

Figure 7: Posterior density of the meta distribution parameters b_0 under the unconstrained (dashed line) and constrained (solid line) models

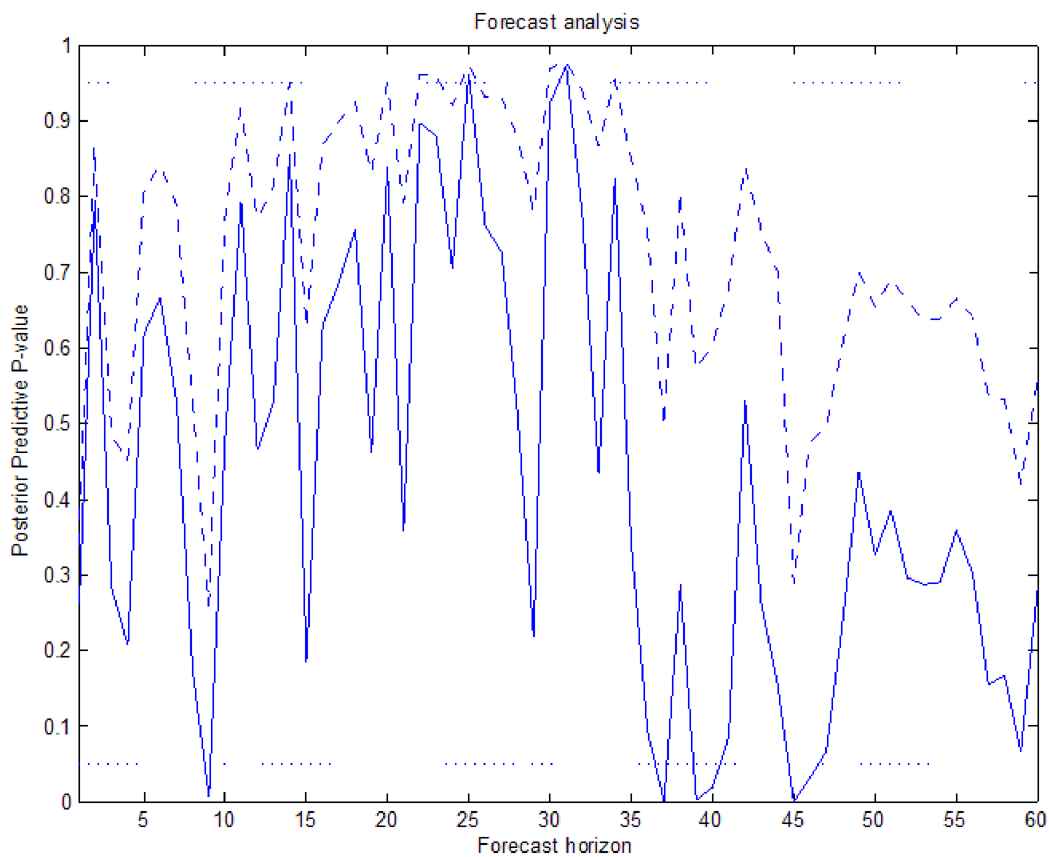

Figure 8: Posterior predictive p-values for Multi step ahead forecasts. This figure shows, for each model, in percentage, how many times the generated Multi step ahead forecasts lie below the realized T-Bill observation, for $t = T + 1, ..., T + 60$. The solid line refers to the NNB (No New Break) model while the dashed line shows the C (Combined) model. The dotted lines report the 90 percent confidence interval.

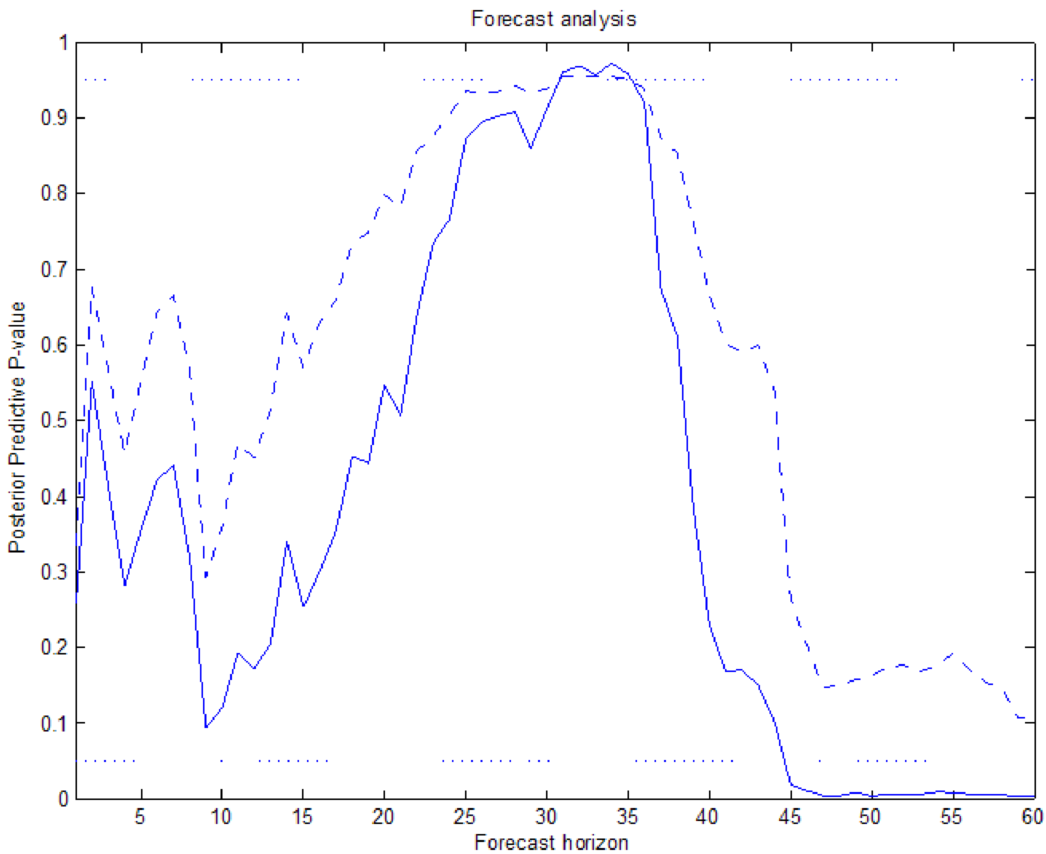

Figure 9: Posterior predictive p-values for Multi step ahead forecasts. This figure shows, for each model, in percentage, how many times the generated Multi step ahead forecasts lie below the realized T-Bill observation, for $t = T+1, ..., T+60$. The solid line refers to the NNB (No New Break) model while the dashed line shows the C (Combined) model. The dotted lines report the 90 percent confidence interval.