

# Bayesian simultaneous determination of structural breaks and lag lengths

Brigitta Hultblad\*  
Statistics Sweden

Sune Karlsson†  
Swedish Business School  
Örebro University

SSE/EFI Working Paper Series in Economics and Finance  
No 630  
June 2006, Revised April 2008

## Abstract

The detection of structural change and determination of lag lengths are long-standing issues in time series analysis. This paper demonstrates how these can be successfully married in a Bayesian analysis. By taking account of the inherent uncertainty about the lag length when deciding on the number of structural breaks and vice versa we avoid some common pitfalls and are able to draw more robust conclusions. The approach is illustrated using both real and simulated data.

**Keywords:** Regime shifts; Model uncertainty; Markov chain Monte Carlo; Model averaging; Real interest rate

**JEL codes:** C11, C15, C22, C51

Stockholm School of Economics, P.O. Box 6501, SE-113 83 Stockholm, Sweden

---

\*Brigitta.Hultblad@scb.se

†Corresponding author. Address: Örebro University, Swedish Business School, 701 82 Örebro, Sweden.  
E-mail: sune.karlsson@oru.se

# 1 Introduction

The detection of structural change and determination of lag lengths are long-standing issues in time series analysis. Much of the interest in structural change has been spurred by the sensitivity of unit root tests to the presence of breaks in trends or levels of time series, Perron (1989), Rappaport and Reichlin (1989), Banerjee, Lumsdaine and Stock (1992), Zivot and Andrews (1992) are early examples of this strand of the literature. More recently, Bai, Lumsdaine and Stock (1998) and Bai and Perron (1998) resolve many of the inferential issues in a frequentist analysis of multiple structural breaks while Bai and Perron (2003) provide efficient computational algorithms. While much of the interest has focused on breaks in trends and levels there has also been a considerable amount of work on changing variances. E.g. Albert and Chib (1993), McCulloch and Tsay (1993), Inclán and Tiao (1994) and Inclán (1994), much of it based on the Markov switching models popularized by Hamilton (1990). A recent strand of the literature exemplified by Pesaran, Petenuzzo and Timmermann (2006) and Koop and Potter (2007) focus on forecasting under the presence of structural breaks and extends the break process out of the sample for this purpose. In the present paper we focus on the in-sample properties and the case where all parameters are allowed to change between regimes. The approach is, however, easily adopted to the case when only a subset of the parameters are allowed to change.

In a time series context where autoregressive models are almost uniformly applied the determination of lag lengths become an important practical issue. In particular so since tests for structural breaks can be quite sensitive to the choice of lag length, too short a lag length smooths the data and a too large lag length overfits and can "hide" structural breaks.

Adopting a Bayesian framework, this paper demonstrates how the detection of structural breaks and lag order determination can be successfully married. By explicitly taking account of the inherent uncertainty about the lag length when deciding on the number of structural breaks and vice versa we avoid some common pitfalls and are able to draw more robust conclusions.

Instead of proceeding sequentially, first determining lag order and then testing for structural breaks conditional on the chosen lag order as well as testing for the number of structural breaks sequentially, we propose to focus on the joint posterior distribution of the number of structural breaks and lag length. The joint posterior summarizes all the relevant information about both structural breaks and lag lengths and leads directly to a joint determination by choosing the combination of structural breaks and lag length with highest posterior probability. In addition the joint posterior provides direct evidence of the uncertainty about these quantities. If there isn't a clear winner it is unwise to conduct any subsequent analysis conditional on a single model. Instead one should acknowledge the model uncertainty and robustify against misspecification by performing Bayesian model averaging (see Hoeting, Madigan, Raftery and Volinsky (1999) for a review of the Bayesian model averaging literature).

If, as we do, the lag length is allowed to differ between regimes the joint posterior of the number of breaks and lag lengths is not well defined. The problem arises because

the regimes are different in a two regime and a three regime model and the lag length of the third regime in a two regime model is undefined. The property "number of lags in the different regimes" is only meaningful conditional on the number of regimes. The property "number of breaks", on the other hand, has meaning independent of the lag lengths and the tool of choice is the marginal posterior distribution of the number of breaks. By marginalizing with respect to the lag lengths in a model with a given number of regimes we robustify against misspecification of the lag length and obtain a valid basis for determining the number of breaks. Having selected the number of breaks from the marginal posterior for the breaks we can choose the lag lengths conditional on the number of breaks. Again, a dispersed posterior for the number of breaks or a dispersed posterior for the lag lengths conditional on the chosen number of breaks cautions against proceeding conditional on a single model.

Note how the selection of the number of breaks differs from frequentist sequential testing. A typical sequential test procedure would first determine the lag length and then conditionally on one single lag length determine the number of breaks. In contrast, the Bayesian procedure considers all lag lengths and the information about the number of breaks provided by models with different lag lengths.

In relation to previous work, the present paper is on a technical level closely related to Inclán (1994) and Wang and Zivot (2000) on determination of structural breaks and Troughton and Godsill (1998) on determination of lag lengths. The value added stems not so much from the merging of the techniques as from their application in obtaining the joint posterior distributions and the use that can be made of this information. Marriot and Newbold (2000) is related in spirit and shows how unit root inference can be robustified by explicitly allowing for structural breaks.

The remainder of the paper proceeds as follows: Section two outlines the modelling approach and provides the necessary analytical results and computational tools. In section three we assess the performance in a small simulation study while section four provides an application to the U.S. ex-post real interest rate previously analyzed by Garcia and Perron (1996) and Bai and Perron (2003). Section 5 concludes.

## 2 Autoregressive models with structural breaks

The autoregressive model subject to  $r$  structural breaks can be written as

$$y_t = \beta_{i0} + \sum_{j=1}^{p_i} \beta_{ij} y_{t-j} + \sigma_i u_t, \quad b_{i-1} < t \leq b_i, \quad i = 1, \dots, r+1$$

where  $b_i$ ,  $i = 1, \dots, r$ , denotes the actual break dates with  $b_0$  and  $b_{r+1}$  fixed at 0 and  $T$  respectively. We allow the lag lengths,  $p_i$ , as well as all the parameters to vary over the regimes. Each of the  $r+1$  regimes is thus characterized by a parameter vector  $\theta_i = (\beta_{i0}, \beta_{i1}, \dots, \beta_{i,p_i}, \sigma_i^2)'$ . Given the number of break points,  $r$ , the vector of break dates,  $\mathbf{b}$ , and the vector of lag lengths,  $\mathbf{p}$ , the analysis is straightforward. Difficulties do arise when, as is the case in practice,  $r$ ,  $\mathbf{b}$  and  $\mathbf{p}$  are unknown and must be determined from the data. The aim of this paper is to propose practical methods for the simultaneous determination of  $r$ ,  $\mathbf{b}$  and  $\mathbf{p}$ . This is in contrast to the previous literature which, in general, proceeds sequentially. Decide on a lag length and conditional on this test for a

structural break and estimate the break date. If a structural break is found the analysis then proceeds by testing for additional breaks conditional on the previously estimated break dates.

Instead we propose focusing on the joint or marginal posterior distribution of these quantities. We will primarily focus on the marginal posterior distribution of the number of break points. By marginalizing out the other unknown quantities,  $\mathbf{b}$ ,  $\mathbf{p}$  and  $\theta$ , and thus accounting for the uncertainty about them, we obtain a sound basis for determining  $r$ . Having determined  $r$  from the marginal posterior we can proceed with the other unknown quantities. Since the dimension of  $\mathbf{b}$ ,  $\mathbf{p}$  and  $\theta$  depends on  $r$  it is in general only meaningful to consider their posterior distribution conditional on the number of breaks and we will proceed in this way to determine  $\mathbf{b}$  and  $\mathbf{p}$ . An exception to this is when the interpretation of a property of interest, e.g. the presence of a unit root, does not depend on the number of breaks, we can then marginalize over  $r$  to obtain inference that is robust to uncertainty about the number of breaks. Another important special case arises when the lag length is assumed to be common across regimes. We can then obtain the joint marginal posterior for  $r$  and  $p$  and jointly determine these quantities.

For convenience we will assume that  $u_{it} \sim NID(0, 1)$ . That is, the likelihood is,

$$\begin{aligned} L(\mathbf{y}|\theta, \mathbf{p}, \mathbf{b}, r) &= \prod_{i=1}^{r+1} (\sigma_i^2)^{-(b_i-b_{i-1})/2} \exp \left[ -\frac{1}{2\sigma_i^2} \sum_{t=b_{i-1}+1}^{b_i} \left( y_t - \beta_{i0} - \sum_{j=1}^{p_i} \beta_{ij} y_{t-j} \right)^2 \right] \\ &= \prod_{i=1}^{r+1} L(\mathbf{y}_i|\theta_i, p_i), \end{aligned}$$

the product of  $r + 1$  normal likelihoods. The normality assumption is not essential for the main thrust of our analysis which relies on MCMC methods to obtain the posterior distributions and can easily be adopted to other distributional assumptions. The normality assumption allow us to obtain analytical expressions for the posterior distributions of interest. As the analytical expressions depend on enumerating all possible combinations of break dates and/or lag lengths they are impractical for longer time series and MCMC methods are preferable in general. They do serve one important purpose in this paper as we can check the output of the Markov chains against the analytical solutions for problems of moderate size.

## 2.1 Prior specification

Unless we have prior information specific to the application at hand (in which case the model should probably be set up differently) a few reasonable independence assumptions leads to a simple form for the prior distribution. First, the prior for the break dates naturally depends on the number of breaks but we do not expect the break dates to carry information about the lag lengths or the regression parameters or vice versa, consequently we have  $\mathbf{b} \sim \pi(\mathbf{b}|r)$ . Second, an exchangeability argument leads to an iid prior on  $(\theta_i, p_i)$  across regimes which we decompose as  $\pi(\theta_i, p_i) = \pi(\theta_i|p_i) \pi(p_i)$ . In the case of a common lag length the exchangeability argument only applies to  $\theta_i$  and we let  $\theta_i$  be iid  $\pi(\theta_i|p)$ . Finally we specify a marginal prior on the number of breaks,  $\pi(r)$ . This leads to the prior

structure

$$\pi(r) \pi(\mathbf{b}|r) \prod_{i=1}^{r+1} \pi(p_i) \pi(\theta_i|p_i)$$

or  $\pi(r) \pi(\mathbf{b}|r) \pi(p) \prod_{i=1}^{r+1} \pi(\theta_i|p)$  in the case of a common lag length.

Alternatively we can frame the exercise in terms of model selection or model averaging and think of a combination of number of breaks, break dates and lag lengths as specifying a model  $\mathcal{M}_{r,b,p} = (r, \mathbf{b}, \mathbf{p})$  with  $\pi(\theta|\mathcal{M}_{r,b,p}) = \prod_{i=1}^{r+1} \pi(\theta_i|p)$  the prior on the parameters for the model and  $\pi(\mathcal{M}_{r,b,p}) = \pi(r) \pi(\mathbf{b}|r) \prod_{i=1}^{r+1} \pi(p_i)$  the prior model probability. This can of course be done at different levels with  $\mathcal{M}_{r,p} = (r, \mathbf{p})$  or  $\mathcal{M}_r = r$  constituting models with obvious changes in the prior model probabilities and the prior on the parameters. We will frequently make use of these different levels of conditioning below.

Turning to the specifics of the prior distributions we opt for simplicity and specify a natural conjugate prior for  $\theta_i = (\beta_i, \sigma_i^2)$  conditional on  $p_i$ . That is,

$$\begin{aligned} \beta_i | \sigma_i^2, p_i &\sim N(\beta_0, \sigma_i^2 \mathbf{M}_0^{-1}) \\ \sigma_i^{-2} &\sim G(v_0/2, 2/S_0) \end{aligned}$$

where, at a minimum, the dimension of  $\beta_0$  and  $\mathbf{M}_0$  depend on  $p_i$ . For the lag lengths we specify a maximum allowed lag length,  $p_m$ , and a uniform prior over the range  $p_i = 0, \dots, p_m$ . That is,

$$\pi(p_i) = (p_m + 1)^{-1} \quad \text{and} \quad \pi(\mathbf{p}|r) = (p_m + 1)^{-(r+1)}$$

or  $\pi(p) = 1/(p_m + 1)$  in the case of common lag lengths.

When it comes to the prior for the break dates we follow the tradition in the structural break literature and do not consider breaks at the very beginning or end of the sample. In addition we require a minimum distance between breaks, all in order to guard against overfitting and the detection of spurious breaks.<sup>1</sup> We do this in terms of one parameter,  $d$ , specifying the minimum number of observations in a regime. Lacking prior information about the location of the breaks we specify a uniform prior over the possible combinations of break dates,  $\mathbf{b}$ . In addition to  $d$ , the number of possible combinations depend on the number of observations,  $T$ , and the number of breaks,  $r$ . A simple combinatorial argument gives the prior for the break dates as the binomial expression

$$\pi(\mathbf{b}|r) = \binom{T - (r+1)d + r}{r}^{-1}.$$

The prior is completed by specifying a uniform prior over the number of breaks,  $r = 0, \dots, r_m$ , given the maximum allowed number of breaks,  $r_m$ ,

$$\pi(r) = (r_m + 1)^{-1}.$$

---

<sup>1</sup>It is certainly possible to allow for very few observations, even less than the number of parameters, in a regime in a Bayesian analysis. But this would simply return the prior distribution as the posterior and would not carry any sensible information about the number of breaks.

## 2.2 Posterior analysis

Conditional on the break dates,  $\mathbf{b}$ , and the lag lengths,  $\mathbf{p}$ , the analysis is completely standard and can be conducted separately for each regime. The posterior for  $\theta_i = (\beta_i, \sigma_i^2)$  has the familiar normal-gamma form,

$$\begin{aligned}\beta_i | \sigma_i^{-2}, \mathbf{y}_i, \mathbf{b}, \mathbf{p} &\sim N(\bar{\beta}_i, \sigma_i^2 \mathbf{M}_{1,i}^{-1}) \\ \sigma_i^{-2} | \mathbf{y}_i, \mathbf{b}, \mathbf{p} &\sim G(v_i^*/2, 2/S_i^*),\end{aligned}\quad (1)$$

with  $\bar{\beta}_i = \mathbf{M}_{1,i}^{-1} (\mathbf{M}_0 \beta_0 + \mathbf{X}'_i \mathbf{X}_i \hat{\beta}_i)$ ,  $\mathbf{M}_{1,i} = \mathbf{M}_0 + \mathbf{X}'_i \mathbf{X}_i$ ,  $v_i^* = v_0 + b_i - b_{i-1}$ ,  $\hat{\beta}_i = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \mathbf{y}_i$  and  $S_i^* = S_0 + \mathbf{y}'_i \mathbf{y}_i + \beta'_0 \mathbf{M}_0 \beta_0 - \bar{\beta}'_i \mathbf{M}_{1,i} \bar{\beta}_i$ . For future reference we also note that the full conditional posterior for  $b_i$ ,  $i = 1, \dots, r$  is given by

$$p(b_i | \mathbf{y}, b_{i-1}, b_{i+1}, \theta_i, \theta_{i+1}, \mathbf{p}, r) \propto L(\mathbf{y}_i | \theta_i, p_i) L(\mathbf{y}_{i+1} | \theta_{i+1}, p_{i+1}) \pi(\mathbf{b} | r). \quad (2)$$

The conditional marginal likelihood for the data in regime  $i$  has the student- $t$  form (see Bauwens, Lubrano and Richard (1999))

$$\begin{aligned}m(\mathbf{y}_i | \mathbf{b}, \mathbf{p}) &= t_{b_i - b_{i-1}}(\mathbf{y}_i | \mathbf{X}_i \beta_0, S_0, (\mathbf{I} + \mathbf{X}_i \mathbf{M}_{1,i}^{-1} \mathbf{X}'_i)^{-1}, v_0) \\ &= \frac{\Gamma(v_i^*/2) S_0^{v_0/2} |\mathbf{M}_0|^{1/2}}{\pi^{(b_i - b_{i-1})/2} \Gamma(v_0/2) |\mathbf{M}_{1,i}|^{1/2}} (S_i^*)^{-(v_0 + b_i - b_{i-1})/2}\end{aligned}\quad (3)$$

and across the regimes the marginal likelihood is simply

$$m(\mathbf{y} | r, \mathbf{b}, \mathbf{p}) = \prod_{i=1}^{r+1} m(\mathbf{y}_i | \mathbf{b}, \mathbf{p}).$$

The marginal likelihood for the model  $\mathcal{M}_{r,b,p}$  provides the basic component in the posterior distribution of the conditioning quantities,  $r$ ,  $\mathbf{b}$ , and  $\mathbf{p}$ . Exact posterior results can thus be obtained if we are willing to enumerate all the possible models and calculate their marginal likelihoods. More precisely, the posterior model probabilities are given by

$$p(\mathcal{M}_{r^*, b^*, p^*}) = \frac{m(\mathbf{y} | r^*, \mathbf{b}^*, \mathbf{p}^*) \pi(r^*) \pi(\mathbf{b}^* | r^*) \pi(\mathbf{p}^* | r^*)}{\sum_{r,b,p} m(\mathbf{y} | r, \mathbf{b}, \mathbf{p}) \pi(r) \pi(\mathbf{b} | r) \pi(\mathbf{p} | r)}.$$

The other quantities of interest can be obtained in a similar fashion, the joint posterior for  $\mathbf{b}$  and  $\mathbf{p}$  conditional on the number of breaks is given by

$$p(\mathbf{b}^*, \mathbf{p}^* | \mathbf{y}, r) = \frac{m(\mathbf{y} | r, \mathbf{b}^*, \mathbf{p}^*) \pi(\mathbf{b}^* | r) \pi(\mathbf{p}^* | r)}{\sum_{b,p} m(\mathbf{y} | r, \mathbf{b}, \mathbf{p}) \pi(\mathbf{b} | r) \pi(\mathbf{p} | r)}.$$

The marginal posteriors are obtained by marginalizing out  $\mathbf{b}$  or  $\mathbf{p}$ ,

$$\begin{aligned}p(\mathbf{b}^* | \mathbf{y}, r) &= \sum_p p(\mathbf{b}^*, \mathbf{p} | \mathbf{y}, r) = \frac{m(\mathbf{y} | r, \mathbf{b}^*) \pi(\mathbf{b}^*)}{\sum_b m(\mathbf{y} | r, \mathbf{b}) \pi(\mathbf{b})} \\ p(\mathbf{p}^* | \mathbf{y}, r) &= \sum_b p(\mathbf{b}, \mathbf{p}^* | \mathbf{y}, r) = \frac{m(\mathbf{y} | r, \mathbf{p}^*) \pi(\mathbf{p}^*)}{\sum_p m(\mathbf{y} | r, \mathbf{p}) \pi(\mathbf{p})}\end{aligned}$$

where  $m(\mathbf{y}|r^*, \mathbf{b}^*) = \sum_p m(\mathbf{y}|r^*, \mathbf{b}^*, \mathbf{p}) \pi(\mathbf{p}|r^*)$  is the marginal likelihoods for the model  $\mathcal{M}_{r^*, b^*}$  and  $m(\mathbf{y}|r^*, \mathbf{p}^*) = \sum_b m(\mathbf{y}|r^*, \mathbf{b}, \mathbf{p}^*) \pi(\mathbf{b}|r^*)$  is the marginal likelihood for the model  $\mathcal{M}_{r^*, p^*}$ . Finally, the marginal posterior distribution of the number of breaks is obtained by marginalizing out  $\mathbf{b}$  and  $\mathbf{p}$ ,

$$p(r^*|\mathbf{y}) = \sum_{b,p} p(r^*, \mathbf{b}, \mathbf{p}|\mathbf{y}) = \frac{m(\mathbf{y}|r^*)}{\sum_r m(\mathbf{y}|r)}$$

for  $m(\mathbf{y}|r^*) = \sum_b m(\mathbf{y}|r^*, \mathbf{b}) \pi(\mathbf{b}) = \sum_p m(\mathbf{y}|r^*, \mathbf{p}) \pi(\mathbf{p})$  the marginal likelihood for the model  $\mathcal{M}_{r^*}$ .

In the case of a common lag length in all regimes we can obtain the joint posterior for  $p$  and  $r$  as

$$p(\mathcal{M}_{r^*, p^*}) = p(p^*, r^*|\mathbf{y}) = \sum_b p(r^*, \mathbf{b}, p^*|\mathbf{y}) = \frac{m(\mathbf{y}|r^*, p^*) \pi(r^*) \pi(p^*)}{\sum_{r,p} m(\mathbf{y}|r, p) \pi(r) \pi(p)}.$$

As we illustrate in the simulation study in section 3, the posterior distribution over the number of lags and structural breaks is relatively insensitive to the priors for  $r$  and  $p$ . There is, however, substantial sensitivity to the overall scale of the prior variance for  $\beta$ ,  $\sigma_i^2 \mathbf{M}_0^{-1}$ , as can be seen from the form of the conditional marginal likelihood (3) for a single regime. Take for example  $\mathbf{M}_0 = c\mathbf{I}$  with  $c$  small for an uninformative prior and  $c$  large for an informative prior. Increasing the number of breaks by one essentially scales the full sample marginal likelihood with a factor  $c^{(p+1)/2}$ . A value of  $c$  substantially less than 1 (uninformative prior) penalizes structural breaks and long lags and a value larger than 1 (informative prior) favours structural breaks and long lags.

While maybe not immediately obvious, it should be clear that all posterior distributions are proper even if a model with too many breaks is estimated. Conditionally on the break dates, the posterior for  $\theta$  is proper and the marginal likelihood  $m(\mathbf{y}|r, \mathbf{b}, \mathbf{p})$  is well defined. Since the number of break date combinations and lag lengths is finite  $p(\mathbf{b}, \mathbf{p}|\mathbf{y}, r)$ ,  $p(\mathbf{b}|\mathbf{y}, r)$ ,  $p(\mathbf{p}|\mathbf{y}, r)$  and  $p(r|\mathbf{y})$  as well as the marginal posterior for  $\theta$  will all be well defined. One or more of the break dates will of course be badly determined with a relatively flat posterior distribution which leads to a small marginal likelihood for the model with too many breaks. Fitting models with too few breaks will typically lead to multimodal distributions for the break dates and small marginal likelihoods. To see this consider a one break model when there are two breaks in the data. Locating the single break at one of the true breaks will lead to a good fit for one part of the data and a bad fit for the rest. The marginal likelihood  $m(\mathbf{y}|r, \mathbf{b}, \mathbf{p})$  will thus be relatively small but still higher than the marginal likelihood when the single break is far away from both true breaks. Consequently the posterior probability for the one break model will also be small.

## 2.3 Marginal likelihood and model probabilities using MCMC

With long time series and/or many breaks and possible lag lengths the exact calculations outlined above can be quite time consuming. Monte Carlo methods can then provide a useful alternative.<sup>2</sup> Markov chain Monte Carlo methods are useful in this context in

<sup>2</sup>In our simulation study there are about 735 000 possible combinations of break dates with  $T = 270$  and 3 breaks. The exact calculations for all break dates with a common lag length of 4 takes approximately

two respects. The primary motivation is that they allow us to depart from the natural conjugate framework, the choice of likelihood and prior distributions can be based on other considerations than the availability of closed forms for the posterior distribution.<sup>3</sup> As alluded to above, MCMC can also be more computationally efficient since we, in many cases, can obtain useful approximations to the posterior distributions without visiting all possible model configurations. This works because the Markov chain avoids model configurations with low posterior probability and concentrates on the practically relevant subset of the model space.

We give two different MCMC algorithms in the Appendix (see Casella and George (1992) and Chib and Greenberg (1995) for accessible introductions to Markov chain Monte Carlo). One conditional on the lag lengths,  $\mathbf{p}$ , and the number of regimes,  $r$ , and one which allows the lag length to vary and only conditions on the number of regimes. Algorithm 1, is, except for the modification noted below, a Gibbs sampler similar to the one proposed by Wang and Zivot (2000). The algorithm simulates  $\theta$  and  $\mathbf{b}$  and will produce an estimate of  $p(\mathbf{b}, \theta | \mathbf{y}, r, \mathbf{p})$  as well as  $p(\theta | \mathbf{y}, r, \mathbf{p})$  and  $p(\mathbf{b} | \mathbf{y}, r, \mathbf{p})$  directly. In addition we can obtain an estimate of  $m(\mathbf{y} | r, \mathbf{p})$  and by simulating for different configurations of  $(r, \mathbf{p})$  we can obtain estimates of the posterior probabilities for  $r$  and  $\mathbf{p}$ .

Algorithm 2 simulates  $\mathbf{p}$  as well as  $\theta$  and  $\mathbf{b}$ . This complicates matters as the dimension of the parameter space changes when  $\mathbf{p}$  changes and we embed a reversible jump Metropolis-Hastings step within the Gibbs sampler to handle this. Troughton and Godsill (1998) use a reversible jump step similar to our step 2 when determining the lag length of an autoregressive process without structural breaks. The basic problem addressed by the reversible jump step is that the dimension of the parameter vector changes with the lag length. The solution is to embed the parameter vector in a higher dimension which is unaffected by the proposed change of lag length. In practice this is done by generating auxiliary random numbers  $\mathbf{u}$  and let the proposal density for the parameters be implicitly defined by the bijection  $\eta^* = (\mathbf{u}^*, \theta^*) = g(\theta_i, \mathbf{u})$ . This leads to a standard (reversible) Metropolis-Hastings step for the augmented parameter vector  $\eta$ . See Green (1995) for additional details.

The Markov chain of Algorithm 2 will produce an estimate of  $p(\mathbf{b}, \mathbf{p}, \theta | \mathbf{y}, r)$ ,  $p(\mathbf{b} | \mathbf{y}, r)$ ,  $p(\mathbf{p} | \mathbf{y}, r)$  and  $p(\theta | \mathbf{y}, r)$ . In addition we can (at least for a  $\mathbf{p}$  with high posterior probability) obtain estimates of the conditional distributions  $p(\mathbf{b}, \theta | \mathbf{y}, r, \mathbf{p})$ ,  $p(\theta | \mathbf{y}, r, \mathbf{p})$  and  $p(\mathbf{b} | \mathbf{y}, r, \mathbf{p})$ . The output of the chain can also be used to estimate  $m(r | \mathbf{y})$  and by simulating for different  $r$  we can obtain estimates of  $p(r | \mathbf{y})$ .

An additional complication arise when the true number of regimes is greater than  $r$ . In this case the posterior is multimodal and the sampler can easily get stuck at one of the modes and fail to explore the full posterior distribution. For example, the marginal posterior for the break dates is bimodal when the true number of breaks is two but we simulate a model with only one break. In this case a Gibbs sampler drawing the break

---

2.5 minutes. Doubling the number of observations gives 13 million possible break date combinations and allowing for 4 breaks with the longer time series gives  $10^9$  break date combinations corresponding to 23 days of computations with the same time per break date combination. In contrast 100 000 iterations from Algorithm 1 is obtained in just over 1 minute with  $T = 270$ , 1.5 minutes with  $T = 540$  and increases by only 10 seconds if we allow for 4 breaks.

<sup>3</sup>Fat tails e.g a  $t$ -distribution can, as pointed out by a referee, be alternatives to structural breaks. The marginal likelihood is then not available in closed form and simulation methods must then be used. Using the results in Geweke (1993) it is straightforward to adapt our MCMC algorithms to a  $t$  likelihood.



date from the full conditional posterior  $p(b, \mathbf{y}, \theta, r, \mathbf{p})$  will typically quickly locate one of the breaks and never move to the other break. To overcome this difficulty we use a Metropolis-Hastings update for  $\mathbf{b}$  and  $\theta$  every  $n^{\text{th}}$  update. The Metropolis-Hastings step proposes break dates from a uniform distribution over all possible breaks and  $\theta$  from  $p(\theta|\mathbf{y}, r, \mathbf{p}, \mathbf{b})$ .

The estimates of the marginal likelihoods are based on the method of Chib (1995).<sup>4</sup> The relation

$$p(\theta_1, \dots, \theta_{r+1}, \mathbf{b}|\mathbf{y}, r, \mathbf{p}) = p(\mathbf{b}|\mathbf{y}) \prod_{i=1}^{r+1} p(\theta_i|\mathbf{y}_i, \mathbf{b}, \mathbf{p}) = \frac{\pi(\mathbf{b}) \prod_{i=1}^{r+1} L(\mathbf{y}_i|\theta_i, \mathbf{b}, \mathbf{p}) \pi(\theta_i)}{m(\mathbf{y}|r, \mathbf{p})}$$

is used to estimate the marginal likelihood for the model  $\mathcal{M}_{r,p}$  from algorithm 1 as

$$\widehat{m}(\mathbf{y}|r, \mathbf{p}) = \frac{\pi(\mathbf{b}^*) \prod_{i=1}^{r+1} L(\mathbf{y}_i|\theta_i^*, \mathbf{b}^*, \mathbf{p}^*) \pi(\theta_i^*)}{\widehat{p}(\mathbf{b}^*|\mathbf{y}) \prod_{i=1}^{r+1} p(\theta_i^*|\mathbf{y}_i, \mathbf{b}^*)} \quad (4)$$

where  $\widehat{p}(\mathbf{b}^*|\mathbf{y})$  is the MCMC estimate of the posterior probability of the break dates and  $\theta_i^*$  etc. are values with high posterior probabilities (we use the posterior modes for  $\mathbf{b}$  and the conditional posterior means for  $\beta_i$  and  $\sigma_i^2$ ).

The marginal likelihood  $m(\mathbf{y}|r)$  for the model  $\mathcal{M}_r$  is obtained in a similar fashion from the output of algorithm 2, using

$$p(\mathbf{b}, \mathbf{p}|\mathbf{y}) \prod_{i=1}^{r+1} p(\theta_i|\mathbf{y}_i, \mathbf{b}, \mathbf{p}) = \frac{\pi(\mathbf{b}) \pi(\mathbf{p}) \prod_{i=1}^{r+1} L(\mathbf{y}_i|\theta_i, \mathbf{b}, \mathbf{p}) \pi(\theta_i)}{m(\mathbf{y}|r)}$$

and

$$\widehat{m}(\mathbf{y}|r) = \frac{\pi(\mathbf{b}^*) \pi(\mathbf{p}^*) \prod_{i=1}^{r+1} L(\mathbf{y}_i|\theta_i^*, \mathbf{b}^*, \mathbf{p}^*) \pi(\theta_i^*)}{\widehat{p}(\mathbf{b}^*, \mathbf{p}^*|\mathbf{y}) \prod_{i=1}^{r+1} p(\theta_i^*|\mathbf{y}_i, \mathbf{b}^*)}$$

where  $\widehat{p}(\mathbf{b}^*, \mathbf{p}^*|\mathbf{y}) = \widehat{p}(\mathbf{p}^*|\mathbf{y}) \widehat{p}(\mathbf{b}^*|\mathbf{p}^*, \mathbf{y})$  is the MCMC estimate of the posterior probability of the break dates and  $\theta_i^*, \mathbf{p}^*$  etc. are values with high posterior probabilities (we use the posterior modes for  $\mathbf{b}$  and  $\mathbf{p}$  and the conditional posterior means for  $\beta_i$  and  $\sigma_i^2$ ). Both  $\widehat{p}(\mathbf{p}^*|\mathbf{y})$  and  $\widehat{p}(\mathbf{b}^*|\mathbf{p}^*, \mathbf{y})$  can be obtained from the primary Markov chain. The latter might require a large number of draws in order to be accurate. Note that the discrete nature of  $\mathbf{b}$  and  $\mathbf{p}$  simplifies this procedure considerably. With continuously distributed variables we would need to run a second chain, holding  $\mathbf{p}$  fixed, in order to estimate  $p(\mathbf{b}^*|\mathbf{p}^*, \mathbf{y})$ .

---

<sup>4</sup>We have also experimented with the harmonic mean estimator of the marginal likelihood but found this to procedure inferior estimates.

### 3 Simulation experiment

We use a small simulation experiment to assess the model selection properties of our procedure and compare the performance to the structural break tests of Bai and Perron (1998, 2003). As a base case we have a model without structural breaks,

$$y_t = 1 + 0.49y_{t-1} - 0.64y_{t-2} + \varepsilon_t, \sigma_\varepsilon^2 = 0.5, T = 270. \text{ (DGP 1)} \quad (5)$$

This model has complex roots with a modulus of 1.25 and a period of 5 and the unconditional expectation of  $y_t$  is  $E(y_t) = 0.87$ . For the remaining DGPs we introduce structural breaks at  $t = 75$  and  $t = 190$ . The first regime thus contains 75 observations (although some of these will be lost due to lags), the second regime contains 115 observations and the last regime 80 observations. In each case the outer regimes are the same as the base case and we modify the inner regime.

In the first set of structural break DGPs only the constant changes and the dynamics are unaffected by the break. We consider  $\beta_0 = 1.5$  and  $1.75$  corresponding to unconditional expectations for the inner regime of  $E(y_t) = 1.31$  and  $1.56$ . We refer to these as DGPs 2a and 2b.

The second set of structural break DGPs consider changing dynamics. We first consider a relatively small change where the modulus of the inner regime is 5 (smaller amplitude) and the period is unchanged at 5. In addition we consider a change to dynamics where the roots are real in the inner regime with a dominant root of  $-1.25$  and the second root set to match the unconditional expectation in DGP 2a. We thus have two settings for the inner regime:

$$y_t = 1 + 0.12y_{t-1} - 0.04y_{t-2} + \varepsilon_t, \sigma_\varepsilon^2 = 0.5, 75 < t \leq 190, E(y_t) = 1.09 \text{ (DGP 3a)} \quad (6)$$

$$y_t = 1 - 0.22y_{t-1} + 0.46y_{t-2} + \varepsilon_t, \sigma_\varepsilon^2 = 0.5, 75 < t \leq 190, E(y_t) = 1.32 \text{ (DGP 3b)}$$

The third set of structural break DGPs changes both the constant and the dynamics but keep the unconditional expectation constant at  $E(y_t) = 1$  across regimes. For the outer regimes with the dynamics in (5) we set the constant to  $\beta_0 = 1.15$ . For the inner regime we use the same set of dynamics (6) as in the second set of DGPs and set the constant to  $\beta_0 = 0.92$  and  $0.76$ , respectively. We will refer to these as DGPs 4a and 4b.

For each DGP we generate 100 time series and calculate the exact model probabilities for all combinations of number of breaks  $0 \leq r \leq 3$  and lag lengths  $0 \leq p \leq 4$ . Throughout we use the priors of section 2.1 with  $\beta_0 = \mathbf{0}$ ,  $\mathbf{M}_0 = \mathbf{I}$ ,  $S_0 = 6$  and  $v_0 = 8$ . The choice of  $S_0$  and  $v_0$  corresponds to a prior mean and variance of 1.25 and 0.44 for the precision and 1 and 0.5 for the variance of  $\varepsilon$ . The minimum length of a regime is set to  $d = 27$  for comparability with the Bai-Perron test. We use exact calculations rather than estimated model probabilities from Algorithm 1 to reduce the influence of simulation noise. In addition to performing Bayesian model selection and choose the combination of number of breaks with the highest posterior probability out of the  $4 \times 5 = 20$  specifications (the MaxP procedure) we also consider three additional procedures. The first two are similar to frequentist sequential testing, first selecting the number of breaks in a model with zero lags and then selecting lags conditional on the number of breaks (the PBL procedure) or first selecting lags in a model with no breaks and then selecting breaks conditional on the lags (the PLB procedure). The last is more sophisticated and takes account of the

joint posterior distribution of lags and breaks by choosing breaks based on the marginal posterior distribution over the number of breaks and then selecting lags conditional on the breaks (the mPBL procedure). This corresponds to our favoured procedure when the number of lags is allowed to differ between regimes.

For the Bai-Perron tests we follow the recommended testing procedure in Bai and Perron (2003, 2004) and first test the null of no structural breaks against the alternative of at least one structural break with the  $UD$  max test. If the null is rejected the number of breaks is determined using a sequence of  $F(r+1|r)$  tests. A significance level of 5% is used throughout. In order to determine both the number of breaks and lag length we employ two sequential procedures. The first procedure first test for the number of breaks in a model with zero lags and then use BIC to choose lags given the number of breaks (the BP-BIC procedure). The second procedure first finds the number of lags with BIC in a model with no breaks and then tests for breaks given the number of lags (the BIC-BP procedure). In the BP-BIC procedure we use the version of the Bai-Perron test that allow for autocorrelation in the residuals and we allow for different error variances and data distributions (moment matrices) across regimes in both procedures. Finally we use a trimming constant  $\varepsilon = 0.1$ , corresponding to a minimum regime length of 27.

### 3.1 Results

To conserve space we only report a subset of the results and focus on the performance of the MaxP, BP-BIC and BIC-BP procedures.<sup>5</sup> The results for the mPBL procedure are in all cases very close to the MaxP procedure whereas the PBL and PLB procedures performs considerably worse. This confirms our intuition that marginalization and thereby accounting for the uncertainty about other model features provides a valid basis for inference while conditioning on specific features is problematic. Qualitatively the performance of the PBL and PLB procedures are similar to the BP-BIC and BIC-BP procedures.

The results are summarized in Table 1. Overall the procedures perform very well when there is no break in the DGP (DGP 1), selecting zero breaks and two lags in over 90% of the replicates. The BP-BIC procedure never rejects the null hypothesis of zero breaks indicating that the  $UD$  max test is undersized when the nonparametric correction for autocorrelation is used. The BIC-BP test selects zero breaks in 94% of the replicates indicating that the  $UD$  max test has the right size when a sufficient number of lags are included in the model.

Turning to the DGPs with a break in the constant term, the MaxP procedure and the BP-BIC procedure perform poorly with a small break (DGP 2a) selecting two breaks and two lags in 28% or less of the replicates. The poor performance is due to the procedures selecting zero breaks in the majority of the replicates. The BIC-BP procedure performs much better selecting the correct combination of breaks and lags in 59% of the replicates. The picture changes as the magnitude of the break increases. With a larger break (DGP 2b) MaxP and BP-BIC selects the correct combination in 72% of the replicates with too few breaks and too long lag lengths in the remaining replicates. The BIC-BP procedure, on the other hand, selects too long lags in 82% of the replicates and performs poorly with this DGP.

---

<sup>5</sup>A full set of results for a wider range of DGPs are available on request from the authors.

**Table 1** Proportion of number of breaks and lag lengths selected by procedures

		MaxP:Max posterior probability				BP- B-P break test, BIC: then BIC lag				BIC- BIC lag, then B-P break test										
DGP	$r \setminus p$	0	1	2	3	4	Sum	0	1	2	3	4	Sum	0	1	2	3	4	Sum	
DGP 1	0	0	<b>0</b>	<b>94</b>	<b>4</b>	<b>2</b>	<b>100</b>	0	<b>0</b>	<b>95</b>	<b>4</b>	<b>1</b>	<b>100</b>	0	<b>0</b>	<b>91</b>	<b>2</b>	<b>1</b>	<b>94</b>	
	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	4	2	0	6
	2	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0
	3	0	0	0	0	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0
	Sum	0	0	<b>94</b>	4	2	100	0	0	<b>95</b>	4	1	100	0	0	<b>95</b>	4	1	100	
DGP 2a	$r \setminus p$	0	1	2	3	4	Sum	0	1	2	3	4	Sum	0	1	2	3	4	Sum	
	0	0	0	<b>64</b>	15	8	87	0	0	<b>77</b>	12	7	96	0	0	<b>8</b>	2	4	14	
	1	0	0	1	0	0	1	1	0	0	2	0	2	1	0	0	<b>12</b>	3	1	16
	2	<b>0</b>	<b>0</b>	<b>12</b>	<b>0</b>	<b>0</b>	<b>12</b>	2	<b>0</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>2</b>	2	<b>0</b>	<b>0</b>	<b>59</b>	<b>8</b>	<b>3</b>	<b>70</b>
	3	0	0	0	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0	
Sum	0	0	<b>77</b>	15	8	100	0	0	<b>81</b>	12	7	100	0	0	<b>79</b>	13	8	100		
DGP 2b	$r \setminus p$	0	1	2	3	4	Sum	0	1	2	3	4	Sum	0	1	2	3	4	Sum	
	0	0	0	<b>6</b>	5	17	28	0	0	<b>12</b>	7	5	24	0	0	<b>0</b>	0	1	1	
	1	0	0	0	0	0	0	1	0	0	4	0	4	1	0	0	0	2	5	7
	2	<b>0</b>	<b>0</b>	<b>72</b>	<b>0</b>	<b>0</b>	<b>72</b>	2	<b>0</b>	<b>72</b>	<b>0</b>	<b>0</b>	<b>72</b>	2	<b>0</b>	<b>0</b>	<b>18</b>	<b>36</b>	<b>38</b>	<b>92</b>
	3	0	0	0	0	0	0	3	0	0	0	0	0	3	0	0	0	0	0	
Sum	0	0	<b>78</b>	5	17	100	0	0	<b>88</b>	7	5	100	0	0	<b>18</b>	<b>38</b>	<b>44</b>	100		

Bold entries indicate correct choice of number of breaks and or lags

**Table 1** continued. Proportion of number of breaks and lag lengths selected by procedures

DGP	MaxP:Max posterior probability				BP- B-P break test, BIC: then BIC lag				BIC- BP: then B-P break test					
	$r \setminus p$	0	1	2	3	4	Sum	0	1	2	3	4	Sum	
		0	1	2	3	4	Sum	0	1	2	3	4	Sum	
3a	0	0	0	<b>33</b>	2	6	41	0	0	0	<b>79</b>	14	7	100
	1	0	0	<b>6</b>	0	0	6	1	0	0	<b>0</b>	0	0	0
	2	<b>0</b>	<b>0</b>	<b>51</b>	<b>2</b>	<b>0</b>	<b>53</b>	2	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	3	0	0	<b>0</b>	0	0	0	3	0	0	<b>0</b>	0	0	0
	Sum	0	0	<b>90</b>	4	6	100	Sum	0	0	<b>79</b>	14	7	100
3b	0	0	0	0	0	0	0	0	0	0	<b>79</b>	14	7	100
	1	0	0	<b>0</b>	0	0	0	1	0	0	<b>0</b>	0	0	0
	2	<b>0</b>	<b>0</b>	<b>99</b>	<b>0</b>	<b>0</b>	<b>99</b>	2	<b>0</b>	<b>0</b>	<b>27</b>	<b>16</b>	<b>3</b>	<b>46</b>
	3	0	0	<b>1</b>	0	0	1	3	0	0	<b>0</b>	0	0	0
	Sum	0	0	<b>100</b>	0	0	100	Sum	0	0	<b>28</b>	22	50	100
				<b>100</b>					1	1	<b>0</b>	1	97	100

Bold entries indicate correct choice of number of breaks and or lags

**Table 1** continued. Proportion of number of breaks and lag lengths selected by procedures

DGP	MaxP:Max posterior probability					BP- B-P break test, BIC: then BIC lag					BIC- BIC lag, then B-P break test										
	$r \setminus p$	0	1	2	3	4	Sum	$r \setminus p$	0	1	2	3	4	Sum	$r \setminus p$	0	1	2	3	4	Sum
4a	0	0	0	<b>53</b>	8	4	65	0	0	0	<b>75</b>	19	5	99	0	0	0	<b>7</b>	1	2	10
	1	0	0	<b>5</b>	0	0	5	1	0	0	<b>1</b>	0	0	1	1	0	0	<b>13</b>	4	1	18
	2	<b>0</b>	<b>0</b>	<b>30</b>	<b>0</b>	<b>0</b>	<b>30</b>	2	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	2	<b>0</b>	<b>0</b>	<b>56</b>	<b>14</b>	<b>2</b>	<b>72</b>
	3	0	0	<b>0</b>	0	0	0	3	0	0	<b>0</b>	0	0	0	3	0	0	<b>0</b>	0	0	0
	Sum	0	0	<b>88</b>	8	4	100	Sum	0	0	<b>76</b>	19	5	100	Sum	0	0	<b>76</b>	19	5	100
4b	$r \setminus p$	0	1	2	3	4	Sum	$r \setminus p$	0	1	2	3	4	Sum	$r \setminus p$	0	1	2	3	4	Sum
	0	0	0	<b>0</b>	0	0	0	0	0	0	<b>0</b>	22	62	84	0	0	0	<b>0</b>	0	0	0
	1	0	0	<b>0</b>	0	0	0	1	0	0	<b>0</b>	2	6	8	1	0	0	<b>0</b>	0	7	7
	2	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>100</b>	2	<b>0</b>	<b>0</b>	<b>0</b>	<b>6</b>	<b>2</b>	<b>8</b>	2	<b>0</b>	<b>0</b>	<b>0</b>	<b>23</b>	<b>70</b>	<b>93</b>
	3	0	0	<b>0</b>	0	0	0	3	0	0	<b>0</b>	0	0	0	3	0	0	<b>0</b>	0	0	0
Sum	0	0	<b>100</b>	0	0	100	Sum	0	0	<b>0</b>	30	70	100	Sum	0	0	<b>0</b>	23	77	100	

Bold entries indicate correct choice of number of breaks and or lags

Turning to the DGPs with breaks in the dynamics, the signal is relatively weak in DGP 3a with a break from complex dynamics to complex dynamics. The BIC-BP procedure does well, selecting two breaks and two lags in 69% of the replicates followed by MaxP with 51% correct selections. The BP-BIC procedure does not pick up the change in dynamics and always selects zero breaks. The signal is stronger in the DGP with a break to real dynamics (DGP 3b) and the MaxP procedure picks the correct lag length and number of breaks in 99% of the replicates. The BP-BIC procedure does better but there is still a tendency to select too few breaks and too many lags are selected to compensate for the missing breaks. The BIC-BP procedure, on the other hand, selects too long lags and performs worse with stronger signals.

The last set of DGPs repeats the same breaks in the dynamics but combines it with an offsetting break in the constant term to keep the unconditional expectation constant across regimes. The results are qualitatively similar to the results for DGP 3a and 3b. The MaxP and BIC-BP procedures perform well in selecting the number of breaks. With DGP 4a BIC-BP selects the correct number of breaks and lags in 56% of the replicates, MaxP in 30% and BIC-BP in 20% of the replicates. With DGP 4b MaxP selects the correct number of breaks and lags in all replicates whereas BIC-BP selects the correct number of breaks in 93% of the replicates but always selects too long lags leading to zero correct combinations. The BP-BIC procedure, finally, underestimates the number of breaks and overestimates the lag length.

Summing up the results we see that the MaxP procedure consistently performs well. It has the highest or second highest frequency of correct number of breaks and lag lengths for all DGPs. MaxP is beaten by the BIC-BP procedure for the DGPs with small and difficult to detect breaks (DGP 2a, 3a and 4a). But the performance of the BIC-BP procedure varies substantially across the DGPs considered here, never selecting the correct combination of breaks and lags for some of the DGPs, which makes it difficult to recommend this procedure..

### 3.2 Prior sensitivity

In order to investigate the sensitivity of the MaxP procedure to changes in the prior specifications we consider variations on the priors for the number of breaks,  $r$ , the number of lags,  $p$ , and the regression parameters  $\beta_i$ . We do this for two of the DGPs above, the no break DGP 1 and DGP 3a with a relatively difficult to detect break to complex dynamics.<sup>6</sup> For the number of breaks and lags we consider priors of the form

$$\begin{aligned}\pi(r) &\propto (r+1)^{-\delta_r} \\ \pi(p) &\propto (p+1)^{-\delta_p}\end{aligned}$$

with  $\delta_r, \delta_p = 0, 1$  or  $2$ .  $\delta = 0$  corresponds to our uniform priors in section 2.1 and  $\delta > 0$  favours fewer breaks and shorter lag lengths. With the no break DGP 1 the range of values for  $\delta$  considered here has minimal effects on the selection of breaks and lags. For DGP 3a the choice of  $\delta_r$  has a non-trivial effect on the selection of breaks while  $\delta_p$  has a relatively small effect on the already quite good selection of lags. The frequency of selecting the correct number of breaks is reduced by 10 percentage points when  $\delta_r$  is increased from 0 to 1 and again reduced by 10 percentage points when  $\delta_r$  is increased from 1 to 2.

---

<sup>6</sup>Detailed results are available on request from the authors.

Turning to the prior for the regression parameters,

$$\beta_i | \sigma_i^2, p_i \sim N(\beta_0, \sigma_i^2 \mathbf{M}_0^{-1}),$$

we keep the prior mean of zero and study the effect of changes to the prior variance. In addition to our preferred setting  $\mathbf{M}_0 = \mathbf{I}$ , we set  $\mathbf{M}_0 = c\mathbf{I}$  with  $c = 0.1$  for a more uninformative prior,  $c = 10$  for a tighter prior and  $\mathbf{M}_0 = c\mathbf{L}$ , with  $c = 0.1, 1, 10$  and  $\mathbf{L}$  the diagonal matrix  $\mathbf{L} = \text{diag}(0.1, 1, 2, \dots, p)$  for a Litterman type prior which is tighter around the prior mean of zero for higher lags. Compared to the results for DGP 1 and  $\mathbf{M}_0 = \mathbf{I}$  in Table 1, the more disperse priors with  $c = 0.1$  give similar results whereas the tighter prior with  $c = 10$  leads to the selection of too high lag lengths in about 20% of the replicates while still selecting the number of breaks correctly. The Litterman type prior with  $\mathbf{M}_0 = 10\mathbf{L}$  also leads to the selection of too high lag lengths in about 55% of the replicates and thus to a lower frequency of correct break and lag selection.

The effect of the prior specification is larger with DGP 3a. Setting  $c = 0.1$  leads, as discussed in section 2.2, to the selection of too few breaks. Compared to the results for  $\mathbf{M}_0 = \mathbf{I}$  in Table 1 the selection of zero breaks increases by 20 percentage points but has little effect on the choice of lag length. Choosing  $c = 10$  somewhat unexpectedly increases the selection of zero breaks even more to 94% of the replicates and also increases the selection of too large lag lengths with two lags selected in only 50% of the replicates. A possible explanation is that the prior mean of zero does not agree with the data and that the posterior is dominated by the informative prior when there is more than one regime and relatively few observations in each regime.

Increasing the tightness of the prior for higher lags (using  $\mathbf{L}$  instead of  $\mathbf{I}$ ) hardly affects the selection of breaks and lags for  $c = 0.1$  and 1. With  $c = 10$  the prior becomes very tight around zero for large lags and allowing for higher lags hardly contributes to the fit of the model. The maximum lag is consequently selected less often with  $\mathbf{M}_0 = 10\mathbf{L}$  than with  $\mathbf{M}_0 = 10\mathbf{I}$  and this makes it easier to detect the breaks, leading to the selection of two breaks in 20% of the replicates.

## 4 The US ex-post real interest rate

Garcia and Perron (1996) and Bai and Perron (2003) analyzed the U.S. ex-post real interest rate (the three-month treasury bill deflated by the CPI inflation rate) in figure 1<sup>7</sup>. The date is quarterly between 1961:1 and 1986:3. In both cases only the mean and variance was allowed to change, Garcia and Perron used an AR(2) specification whereas Bai and Perron only fitted a constant term and used a non-parametric correction for autocorrelation. Garcia and Perron (1996) found evidence in favour of two breaks and Bai and Perron (2003) found three breaks. Wang and Zivot (2000) analyze a longer version of the real interest rate time series that appear to be ending in 1991. Similar to Garcia and Perron they use an AR(2) specification and allow the intercept and variance to change between regimes. In a Bayesian analysis they find evidence for three brakes.

We start the analysis by considering the case with a common unknown lag length and an unknown number of breaks and set the maximum number of lags and break points to

---

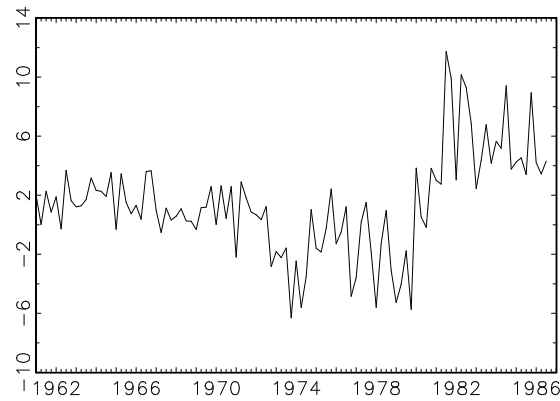
<sup>7</sup>The data is available from the Journal of Applied Econometrics data archive for the Bai and Perron (2003) paper.



---

**Figure 1** US ex-post real interest rate

---



---

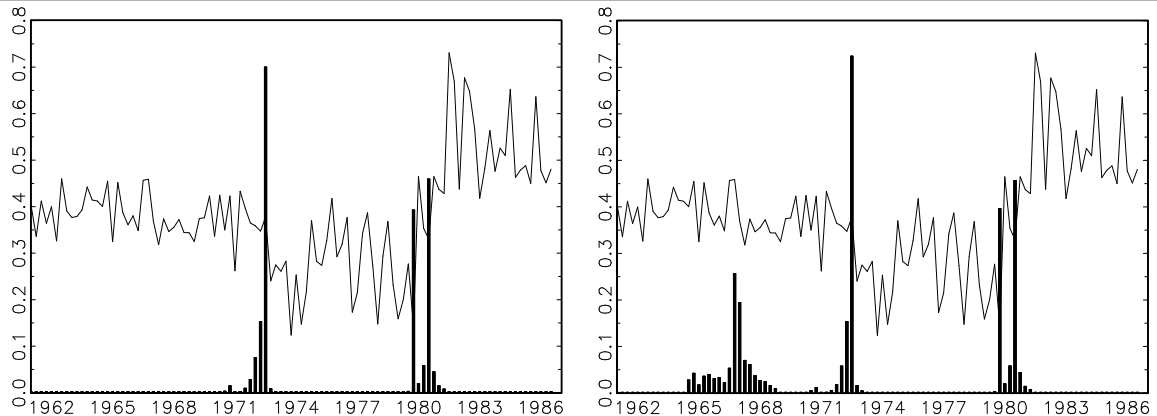
$p_m = 4$  and  $r_m = 4$ . The prior specification is the same as for the simulated data with  $\beta_0 = \mathbf{0}$ ,  $\mathbf{M}_0 = \mathbf{I}$ ,  $S_0 = 6$  and  $v_0 = 8$ . The posterior model probabilities for the models  $\mathcal{M}_{r,p}$  based on exact calculations and the output from algorithm 1, running the algorithm 1 000 000 steps with a burn in of 10 000 for  $p = 0, \dots, p_m$  and  $r = 0, \dots, r_m$ , are shown in table 2. The exact calculations and the MCMC estimates are in very close agreement and indicate a lag length of zero and two or three structural breaks, that is three or four regimes. Figure 2 shows the marginal distribution of the break dates conditional on two or three breaks. For the two break case a joint 80% highest posterior density region includes the break dates (72:3, 80:3), (72:3, 79:4), (72:2, 80:3), (72:2, 79:4), (72:3, 80:2) and (72:1, 80:3) with the first two break dates contributing 60% of the posterior probability. Conditioning on three breaks adds a third, less well determined, break near 1966:4. The joint posterior is naturally more disperse and the posterior probability of the top combination (66:4, 72:3, 80:3) is only 0.082. The combinations of (66:4, 67:1, 67:2, 67:3)  $\times$  72:3  $\times$  (80:3, 79:4) gives a 36% HPD for the break dates. The break dates for the two break case coincide with the breaks reported by Garcia and Perron (1996) and dates for the three break case are identical to the breaks found by Bai and Perron (2003). The results differ slightly from Wang and Zivot (2000) who locate their three breaks at 70:3, 80:2 and 85:4. Their third brake is at the very end of our time series and is ruled out a priori by the requirement that each regime has at least 15 observations.

While our three brake point estimates of the break dates coincide with those of Bai and Perron (2003) other aspects of the inference differ substantially. Highest posterior density regions are shorter and provide more precise information than confidence intervals for the break dates as well as revealing multimodalities in the posterior. Bai and Perron (2003) reports 95% confidence intervals for the break dates as (64:1–69:2), (70:3–72:4) and (79:4–81:1). The corresponding marginal 95% HPD regions are (64:3–64:4, 65:2–68:2), (71:4–72:3) and (79:4, 80:2–80:4).

It is worth noting that sequential procedures, first selecting lags and then breaks would lead to very different conclusions for the real interest rate. Table 3 gives the poster distributions for  $p$  conditional on  $r = 0$  and the posterior of  $r$  conditional on  $p = 3$ . If the analysis is started by selecting lag length in a model without breaks we would select three lags and conditional on this find only one break. On the other hand, choosing lag length and the number of breaks based on the marginal posterior probabilities given in table 2

**Table 2** Posterior model probabilities assuming a common lag length, real interest rate

Exact posterior probabilities						
$r \backslash p$	0	1	2	3	4	Marg prob
0	0	0	0	0	0	0
1	0	0	0	0	0	0.0001
2	0.4130	0.0018	0	0	0	0.4148
3	0.5779	0.0033	0	0	0	0.5812
4	0.0039	0	0	0	0	0.0039
Marg prob	0.9948	0.0052	0	0	0	1
MCMC estimates using algorithm 1						
$r \backslash p$	0	1	2	3	4	Marg prob
0	0	0	0	0	0	0
1	0	0	0	0	0	0.0001
2	0.4140	0.0018	0	0	0	0.4158
3	0.5769	0.0033	0	0	0	0.5802
4	0.0039	0	0	0	0	0.0039
Marg prob	0.9948	0.0052	0	0	0	1

**Figure 2** Marginal posterior distributions of break dates,  $p = 0$ ,  $r = 2$  and  $r = 3$ 

accounts for the uncertainty about the other quantity and leads to the same choice as when the joint posterior is used.

Relaxing the assumption of equal lag length, we start by calculating the marginal posterior probabilities for  $r$  when the lag lengths are allowed to vary independently between 0 and  $p_m$  for each regime. Table 4 shows both exact results based on enumerating all combinations of lag lengths and breaks and Chib type estimates based on 3 000 000 replicates from Algorithm 2 for  $r = 0, \dots, 4$ .<sup>8</sup> The results in Table 4 are similar to the analysis with common lag length and favours two or three breaks but now with a higher probability for two breaks. Table 5 gives the posterior probabilities for the five top combinations of lag lengths given two and three structural breaks. The results for two breaks are in agreement with the analysis conditional on a common lag length indicating a clear

<sup>8</sup>This is one example showing the difficulty with the brute force approach of enumerating combinations of break dates and lag lengths. The total run time for the MCMC based calculations was 1 hour and 28 minutes while the exact calculations required well over 4 hours.

**Table 3** Conditional posterior distributions assuming equal lag length, real interest rate

Posterior for $p$ conditional on $r = 0$					
$p$	0	1	2	3	4
$P(p r=0)$	0	0.0046	0.0218	0.7881	0.1856
Posterior for $r$ conditional on $p = 3$					
$r$	0	1	2	3	4
$P(r p=3)$	0.0211	0.9153	0.0626	0.010	0

**Table 4** Log marginal likelihoods and marginal posterior probabilities for  $r$ , unequal lag lengths.

	$r$	0	1	2	3	4
Exact calculations	$\ln m(\mathbf{y} r)$	-248.33	-241.01	-237.48	-237.81	-243.94
	$P(r \mathbf{y})$	0	0.0167	0.5719	0.4105	0.0008
Estimates based on Algorithm 2	$\widehat{\ln m}(\mathbf{y} r)$	-248.33	-241.02	-237.45	-237.71	-243.98
	$\widehat{P}(r \mathbf{y})$	0	0.0157	0.5532	0.4303	0.0008

preference for no lags in all three regimes. With three breaks the results are not so clear cut and there is an indication that there should be at least one lag in the second regime.

Table 6 gives the posterior means of the parameters for the specification  $r = 2$ ,  $p = 0$  and Table 7 for  $r = 3$  with one lag in the second regime. We give the posterior means accounting for uncertainty about the break dates (i.e. the average of the conditional posterior means weighted by the posterior probability of the break dates or the mean of the posterior marginalized with respect to the break dates) as well as the posterior means conditional on the three combinations of break dates with highest posterior probabilities. We also report 90% equal tail posterior density regions based on the marginal and conditional posteriors. In all cases there are substantial differences between the parameters in all the regimes, indicating that it is inappropriate to assume common parameters. The choice of break dates clearly influence the parameter estimates and averaging over the break dates robustifies against the uncertainty about the break dates.

## 5 Conclusions

The detection of structural change and determination of lag lengths are important topics in time series analysis. We extend the work of Wang and Zivot (2000) and Inclán (1994) by considering autoregressive models with multiple structural changes in level, autoregres-

**Table 5** Posterior probabilities for  $\mathbf{p}$  allowing for different lag lengths

Two breaks					
Lags	0,0,0	0,0,1	1,0,0	0,1,0	2,0,0
Probability	0.5766	0.1106	0.1040	0.0683	0.0329
Three breaks					
Lags	0,1,0,0	0,0,0,0	1,0,0,0	0,2,0,0	1,1,0,0
Probability	0.2480	0.2248	0.0583	0.0571	0.0561

**Table 6** Posterior means and density regions, real interest rate,  $r = 2$  and  $p = 0$ 

	Regime 1	Regime 2	Regime 3
Posterior marginalized over break dates			
$\beta_0$	1.331	-1.809	5.233
	(1.028; 1.634)	(-2.552; -1.059)	(4.275; 6.179)
$\sigma^2$	1.595	5.385	7.584
	(1.150; 2.175)	(3.540; 7.850)	(4.869; 11.362)
Conditional on breaks at $b_1 = 72:3$ , $b_2 = 80:3$ , $p = 0.309$			
$\beta_0$	1.327	-1.742	5.417
	(1.029; 1.625)	(-2.416; -1.067)	(4.541; 6.293)
$\sigma^2$	1.583	5.575	7.123
	(1.145; 2.154)	(3.799; 7.990)	(4.625; 10.648)
Conditional on breaks at $b_1 = 72:3$ , $b_2 = 79:4$ , $p = 0.294$			
$\beta_0$	1.331	-2.055	4.986
	(1.029; 1.625)	(-2.716; -1.394)	(4.105; 5.866)
$\sigma^2$	1.583	4.866	8.070
	(1.145; 2.154)	(3.263; 7.072)	(5.347; 11.856)
Conditional on breaks at $b_1 = 72:2$ , $b_2 = 80:3$ , $p = 0.074$			
$\beta_0$	1.329	-1.654	5.417
	(1.025; 1.633)	(-2.324; -0.985)	(4.542; 6.293)
$\sigma^2$	1.614	5.651	7.123
	(1.163; 2.202)	(3.871; 8.065)	(4.625; 10.648)

90% equal tail posterior density intervals in parenthesis.

sive parameters and variances and combine it with the issue of lag length determination. We demonstrate how a joint determination of breaks and lag length can avoid common pitfalls when these choices are made sequentially. The natural approach is to base the choice on the joint posterior distribution of the number of breaks and lags. Alternatively we can robustify against the uncertainty about the lag length when selecting the number of breaks and vice versa by considering the marginal posterior distributions. In addition, the posterior distributions provides direct evidence of the uncertainty about these quantities and gives us the tool to robustify our analysis against model uncertainty and misspecification by performing Bayesian model averaging.

We provide both analytical expressions for the posterior distributions and MCMC algorithms for exploring the posteriors. The calculations required for the analytical solution becomes impractical if a large number of breaks and/or combinations of lag lengths are considered and MCMC methods are preferred in these cases. We validate the MCMC algorithms against the analytical results and find that the Markov chains give good estimates of the posterior quantities.

A small simulation study demonstrates that both structural breaks and lag lengths are well determined when using our approach. In an application to the U.S. real interest rate we obtain the same results for the number of regimes and location of the breaks as in Bai and Perron (2003).

**Table 7** Posterior means and density regions, real interest rate,  $r = 3$  and  $p = 0, 1, 0, 0$ 

	Regime 1	Regime 2	Regime 3	Regime 4
Posterior marginalized over break dates				
$\beta_0$	1.660 (1.235; 2.088)	1.184 (0.708; 1.673)	-1.829 (-2.565; -1.082)	5.229 (4.271; 6.185)
$\beta_1$		-0.373 (-0.707; -0.040)		
$\sigma^2$	1.538 (0.995; 2.304)	1.176 (0.727; 1.848)	5.367 (3.537; 7.840)	7.592 (4.885; 11.428)
Conditional on breaks at $b_1 = 67:1$ , $b_2 = 72:3$ , $b_3 = 80:3$ , $p = 0.110$				
$\beta_0$	1.710 (1.307; 2.113)	1.154 (0.714; 1.595)	-1.742 (-2.416; -1.067)	5.417 (4.542; 6.292)
$\beta_1$		-0.407 (-0.717; -0.097)		
$\sigma^2$	1.511 (0.981; 2.259)	1.116 (0.714; 1.689)	5.575 (3.799; 7.990)	7.123 (4.625; 10.648)
Conditional on breaks at $b_1 = 67:1$ , $b_2 = 72:3$ , $b_3 = 79:4$ , $p = 0.104$				
$\beta_0$	1.710 (1.307; 2.113)	1.154 (0.714; 1.595)	-2.055 (-2.716; -1.394)	4.986 (4.104; 5.866)
$\beta_1$		-0.407 (-0.717; -0.097)		
$\sigma^2$	1.511 (0.981; 2.259)	1.116 (0.714; 1.689)	4.866 (3.263; 7.072)	8.070 (5.347; 11.856)
Conditional on breaks at $b_1 = 66:4$ , $b_2 = 72:3$ , $b_3 = 80:3$ , $p = 0.074$				
$\beta_0$	1.741 (1.325; 2.157)	1.124 (0.685; 1.563)	-1.742 (-2.416; -1.067)	5.417 (4.542; 6.292)
$\beta_1$		-0.316 (-0.594; -0.038)		
$\sigma^2$	1.544 (0.995; 2.322)	1.122 (0.724; 1.688)	5.575 (3.799; 7.990)	7.123 (4.625; 10.648)

90% equal tail posterior density intervals in parenthesis.

## References

- Albert, J. and Chib, S. (1993), ‘Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts’, *Journal of Business and Economic Statistics* **11**(1), 669–679.
- Bai, J., Lumsdaine, R. and Stock, J. (1998), ‘Testing for and dating common breaks in multivariate time series’, *Review of Economic Studies* **65**, 395–432.
- Bai, J. and Perron, P. (1998), ‘Estimating and testing linear models with multiple structural changes’, *Econometrica* **66**(1), 47–78.
- Bai, J. and Perron, P. (2003), ‘Computation and analysis of multiple structural change models’, *Journal of Applied Econometrics* **18**(1), 1–22.
- Bai, J. and Perron, P. (2004), Multiple structural change models: A simulation analysis, Technical report, Boston College.
- Banerjee, A., Lumsdaine, R. and Stock, J. (1992), ‘Recursive and sequential tests of the unit root and trend break hypotheses: theory and international evidence’, *Journal of Business and Economic Statistics* **10**, 271–288.
- Bauwens, L., Lubrano, M. and Richard, J.-F. (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.
- Casella, G. and George, E. (1992), ‘Explaining the gibbs sampler’, *American Statistician* **46**, 167–174.
- Chib, S. (1995), ‘Marginal likelihood from the gibbs output’, *Journal of the American Statistical Association* **90**(432), 1312–1321.
- Chib, S. and Greenberg, E. (1995), ‘Understanding the metropolis-hastings algorithm’, *American Statistician* **40**, 327–335.
- Garcia, R. and Perron, P. (1996), ‘An analysis of the real interest rate under regime shifts’, *Review of Economics and Statistics* **78**, 111–125.
- Geweke, J. (1993), ‘Bayesian treatment of the independent student-t linear model’, *Journal of Applied Econometrics* **8**, S19–S40.
- Green, P. J. (1995), ‘Reversible jump Markov chain Monte Carlo computation and Bayesian model determination’, *Biometrika* **82**, 711–732.
- Hamilton, J. D. (1990), ‘Analysis of time series subject to changes in regime’, *Journal of Econometrics* **45**, 39–70.
- Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999), ‘Bayesian model averaging: A tutorial (with discussion)’, *Statistical Science* **14**, 382–417. Corrected version available at <http://www.stat.washington.edu/www/research/online/hoeting1999.pdf>.

- Inclán, C. (1994), ‘Detection of multiple changes of variance using posterior odds’, *Journal of Business and Economic Statistics* **11**(3), 289–300.
- Inclán, C. and Tiao, G. (1994), ‘Use of cumulative sums of squares for retrospective detection of changes in variance’, *Journal of the American Statistical Association* **89**(427), 913–923.
- Koop, G. and Potter, S. (2007), ‘Estimation and forecasting in models with multiple breaks’, *Review of Economic Studies* **74**, 763–789.
- Marriot, J. and Newbold, P. (2000), ‘The strength of evidence for unit autoregressive roots and structural breaks: A Bayesian perspective’, *Journal of Econometrics* **98**, 1–25.
- McCulloch, R. and Tsay, R. (1993), ‘Bayesian inference and prediction for mean and variance shifts in autoregressive time series’, *Journal of the American Statistical Association* **88**(423), 968–978.
- Perron, P. (1989), ‘The great crash, the oil price shock and the unit root hypothesis’, *Econometrica* **57**, 1361–1401.
- Pesaran, M. H., Petenuzzo, D. and Timmermann, A. (2006), ‘Forecasting time series subject to multiple structural breaks’, *Review of Economic Studies* **73**, 1057–1084.
- Rappaport, P. and Reichlin, L. (1989), ‘Segmented trends and non-stationary time series’, *Economic Journal* **99**, 168–177.
- Troughton, P. and Godsill, S. (1998), ‘A reversible jump sampler for autoregressive time series’, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing* **4**, 2257–2260.
- Wang, J. and Zivot, E. (2000), ‘A bayesian time series model of multiple structural changes in level, trend and variance’, *Journal of Business and Economic Statistics* **18**(3), 374–386.
- Zivot, E. and Andrews, D. (1992), ‘Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis’, *Journal of Business and Economic Statistics* **10**(3), 251–270.

## Appendix: MCMC Algorithms

---

**Algorithm 1** Simulate  $\theta$  and  $\mathbf{b}$  conditional on  $r$  and  $\mathbf{p}$

---

1. Choose starting values,  $\mathbf{b}^{(0)}$ , for  $\mathbf{b}$ , we use evenly spaced values in  $0, \dots, T$ , and start values for  $\theta$  from  $p(\theta_i | \mathbf{y}_i, \mathbf{b}^{(0)}, \mathbf{p})$  and set  $j = 1$ .
2. For every  $n^{\text{th}}$  update, generate  $\mathbf{b}$  and  $\theta$  using a Metropolis-Hastings step, otherwise do step 3 and 4
  - (a) Propose new break dates  $\mathbf{b}^*$  by drawing from the set of possible break dates with equal probability.
  - (b) For  $i = 1, \dots, r+1$ , draw proposal values  $\theta_i^*$  from the full conditional posterior  $p(\theta_i | \mathbf{y}_i, \mathbf{b}^*, \mathbf{p})$
  - (c) Accept the proposal with probability  $\alpha$  and set  $\theta_i^{(j)} = \theta_i^*$  and  $b_i^{(j)} = b_i^*$  otherwise remain at the current values. The acceptance probability is given by

$$\alpha = \min \left\{ 1, \prod_{i=1}^{r+1} \frac{L(\mathbf{y}_i | \theta_i^*, \mathbf{b}^*, \mathbf{p}) \pi(\theta_i^*, \mathbf{b}^* | \mathbf{p})}{L(\mathbf{y}_i | \theta_i^{(j-1)}, \mathbf{b}^{(j-1)}, \mathbf{p}) \pi(\theta_i^{(j-1)}, \mathbf{b}^{(j-1)} | \mathbf{p})} \prod_{i=1}^{r+1} \frac{p(\theta_i^{(j-1)} | \mathbf{y}_i, \mathbf{b}^{(j-1)}, \mathbf{p})}{p(\theta_i^* | \mathbf{y}_i, \mathbf{b}^*, \mathbf{p})} \right\}.$$

3. For  $i = 1, \dots, r$ , generate  $b_i^{(j)}$  from the full conditional posterior  $p(b_i | \mathbf{y}, b_{i-1}^{(j)}, b_{i+1}^{(j-1)}, \theta_i^{(j-1)}, \theta_{i+1}^{(j-1)}, \mathbf{p})$  in (2).
  4. For  $i = 1, \dots, r+1$ , draw  $\theta_i^{(j)}$  from the full conditional posterior  $p(\theta_i | \mathbf{y}_i, \mathbf{b}^{(j)}, \mathbf{p})$  in (1).
  5. Set  $j = j + 1$  and go to step 2.
- 

The algorithms can be made more computationally efficient in situations where the marginal likelihood  $m(\mathbf{y}_i | \mathbf{b}, p_i)$  is available in closed form. This follows since  $m(\mathbf{y}_i | \mathbf{b}, p_i) = p(\beta_i, \sigma_i^{-2} | \mathbf{y}_i, \mathbf{b}, p_i)^{-1} L(\mathbf{y}_i | \beta_i, \sigma_i^2, \mathbf{b}, p_i) \pi(\beta_i, \sigma_i^{-2} | p_i)$  and the acceptance probability simplifies to  $\alpha = \min \left\{ 1, \frac{\pi(\mathbf{b}^*)}{\pi(\mathbf{b}^{(j-1)})} \prod_{i=1}^{r+1} \frac{m(\mathbf{y}_i | \mathbf{b}^*, p_i)}{m(\mathbf{y}_i | \mathbf{b}^{(j-1)}, p_i)} \right\}$  for the Metropolis-Hastings step and to  $\alpha = \min \left\{ 1, \frac{m(\mathbf{y}_i | \mathbf{b}, p^*) \pi(p^*)}{m(\mathbf{y}_i | \mathbf{b}, p_i^{(j-1)}) \pi(p_i^{(j-1)})} \right\}$  for the reversible jump step. As the acceptance probabilities do not depend on the proposal or current values for  $\theta_i$  we can dispense with generating proposal values unless the proposal is accepted. Instead we generate new draws of  $\theta = (\beta, \sigma^2)$  both when the proposal is accepted and rejected in order to improve mixing. That is, step 4 in Algorithm 1 and step 5 in Algorithm 2 are always executed when the acceptance probabilities simplifies.



---

**Algorithm 2** Simulate  $\theta$ ,  $\mathbf{b}$  and  $\mathbf{p}$  conditional on  $r$ 


---

1. Choose starting values for  $\mathbf{b}$  and  $\mathbf{p}$  and generate start values for  $\theta$  from  $p(\theta_i | \mathbf{y}_i, \mathbf{b}^{(0)}, \mathbf{p}^{(0)})$
2. For  $i = 1, \dots, r + 1$ , generate lag lengths  $p_i^{(j)}$  using a reversible jump step
  - (a) Propose a new lag length  $p^*$  by drawing from  $1, \dots, p_m$  with equal probability.
  - (b) Generate a vector  $\mathbf{u}$  from the proposal density  $q(\mathbf{u} | \theta_i, p_i, p^*)$
  - (c) Obtain proposal values for  $\theta_i$  by setting  $(\mathbf{u}^*, \theta^*) = g(\theta_i, \mathbf{u})$  where  $g$  is a bijection satisfying the dimension matching requirement  $\dim(\mathbf{u}^*) + \dim(\theta^*) = \dim(\mathbf{u}) + \dim(\theta_i)$ .
  - (d) Accept the proposal with probability  $\alpha$  and set  $\theta_i^{(j)} = \theta^*$  and  $p_i^{(j)} = p^*$  otherwise remain at the current values. The acceptance probability is, in general, given by

$$\alpha = \min \left\{ 1, \frac{L(\mathbf{y}_i | \theta^*, \mathbf{b}, p^*) \pi(\theta^*, p^*) \pi(\mathbf{b}) \prod_{k \neq i} L(\mathbf{y}_k | \theta_k, \mathbf{b}, p_k) \pi(\theta_k, p_k)}{\pi(\mathbf{b}) \prod_{i=1}^{r+1} L(\mathbf{y}_i | \theta_i, \mathbf{b}, p_i) \pi(\theta_i, p_i)} \times \frac{q(\mathbf{u}^* | \theta^*, p^*, p_i) j_{*i}}{q(\mathbf{u} | \theta_i, p_i, p^*) j_{i*}} \left| \frac{\partial g(\theta_i, \mathbf{u})}{\partial(\theta, \mathbf{u})} \right| \right\}$$

where  $j_{i*}$  is the probability of proposing  $p^*$  when we are at  $p_i$  and  $j_{*i}$  is the probability of the reverse move. We use the conditional posterior  $p(\theta^* | \mathbf{y}_i, \mathbf{b}, p^*)$  as the proposal distribution  $q$  and  $g(\cdot)$  is the identity defined by  $\mathbf{u}^* = \theta$  and  $\theta^* = \mathbf{u}$ . Consequently the Jacobian of  $g$  is 1 and the acceptance probability simplifies to

$$\alpha = \min \left\{ 1, \frac{L(\mathbf{y}_i | \theta^*, \mathbf{b}, p^*) \pi(\theta^*, p^*) p(\theta_i | \mathbf{y}_i, \mathbf{b}, p_i)}{L(\mathbf{y}_i | \theta_i, \mathbf{b}, p_i) \pi(\theta_i, p_i) p(\theta^* | \mathbf{y}_i, \mathbf{b}, p^*)} \right\}.$$

3. For every  $n^{\text{th}}$  update, generate  $\mathbf{b}$  and  $\theta$  using the same Metropolis-Hastings step as in algorithm 1, otherwise do step 3 and 4.
  4. For  $i = 1, \dots, r$ , generate  $b_i^{(j)}$  from the full conditional posterior  $p(b_i | \mathbf{y}, b_{i-1}^{(j)}, b_{i+1}^{(j-1)}, \theta_i^{(j)}, \theta_{i+1}^{(j)}, \mathbf{p}^{(j)})$  in (2).
  5. For  $i = 1, \dots, r+1$ , generate  $\theta_i^{(j)}$  from the full conditional posterior  $p(\theta_i | \mathbf{y}_i, \mathbf{b}^{(j)}, \mathbf{p}^{(j)})$  in (1).
  6. Set  $j = j + 1$  and go to step 2.
-