University of Leicester

DEPARTMENT OF ECONOMICS

PRIOR ELICITATION IN MULTIPLE CHANGE-POINT MODELS

Gary Koop, **University of Leicester**

Simon M. Potter, **Federal Reserve Bank of New York**

Working Paper No. 04/26
September 2004

Prior Elicitation in Multiple Change-point Models*

Gary Koop
Department of Economics
University of Leicester
email: Gary.Koop@leicester.ac.uk

Simon M. Potter
Business Conditions Function
Federal Reserve Bank of New York
email: simon.potter@ny.frb.org

September 2004

Abstract

This paper discusses Bayesian inference in change-point models. Existing approaches involve placing a (possibly hierarchical) prior over a known number of change-points. We show how two popular priors have some potentially undesirable properties (e.g. allocating excessive prior weight to change-points near the end of the sample) and discuss how these properties relate to imposing a fixed number of change-points in-sample. We develop a new hierarchical approach which allows some of of change-points to occur out-of sample. We show that this prior has desirable properties and handles the case where the number of change-points is unknown. Our hierarchical approach can be shown to nest a wide variety of change-point models, from time-varying parameter models to those with few (or no) breaks. Since our prior is hierarchical, data-based learning about the parameter which controls this variety occurs.

*We would like to thank Edward Leamer for numerous useful conversations on regime modeling and also seminar participants at the Federal Reserve Bank of St. Louis and University of Kansas. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

Change-point modeling has become increasingly popular in the last few years due to an increasing awareness of the importance of this issue for empirical practice.[1] Bayesian methods have proved popular in this field since many change-point models can be written in a simple and computationally convenient form using hierarchical priors. That is, conditional on some parameters or vector of latent data (e.g. the breakpoints or the duration of each regime or a state vector denoting the various regimes), the remainder of the model is very simple (e.g. a sequence of Normal linear regression or Poisson models). A key issue is how these parameters or latent data are modeled. In Bayesian jargon, such model assumptions are made through a (typically hierarchical) prior and various change-point models can be interpreted as implying different priors. However, as with any modelling assumption, it is important that the prior chosen has well-understood properties and be suitably flexible to model empirically-relevant behavior. In this paper, we examine some common approaches and argue that they have properties which may be undesirable in some applications. We use this insight to develop an alternative approach which surmounts this problem. These developments lead to a discussion of the case where the researcher does not know M, the number of regimes. Of course, it is always possible to select the number of regimes by trying various values of M and then using a statistical metric (e.g. a marginal likelihood) to choose M. However, as we shall argue, there are cases where it is appealing to have a model within which M is a parameter to be estimated. We develop a hierarchical prior which allows for this. Throughout, we illustrate our points and new developments using an extended empirical example.

Our starting point will be the influential change-point model developed in Chib (1998).[2] Variants on this model are commonly-used in empirical work in economics and finance [e.g. Pastor and Stambaugh (2001) and Kim, Nelson and Piger (2002)]. See also Pesaran, Pettenuzzo and Timmerman (2004) and Maheu and Gordon (2004) for recent extensions of relevance for

[1] In economics, the terminology "structural break" modeling is often used. We prefer to use the concept of a change-point because the term "structural break" suggests some underlying structure has changed. There are many cases in economics where reduced form relationships can change with the underlying structure remaining constant.

[2] Other key early Bayesian work in the statistics literature includes Carlin, Gelfand, and Smith (1992), Barry and Hartigan (1993) and Stephens (1994).

forecasting in change-point models. Chib's model grew out of early work by Chernoff and Zachs (1964). The latter presented a model where, in each period there is a constant probability of a change to a new regime. If a change occurs the mean of the dependent variable is perturbed by a mean zero Normally distributed shock, if no change occurs the mean remains the same. Chib (1998) generalized the Chernoff and Zachs approach so that the probability of change could vary through time by treating the change-point problem using hidden Markov chains. In the approach of Chib (1998), the problem of locating the change-points is converted into the problem of determining the duration of a Markov regime. As argued by Chib, this allows for the estimation of models, using modern Bayesian methods, with multiple change-points that appear infeasible under the standard approach to change-point problems.

We build on the insight of Chib (1998) in a number of ways. In particular, we show that both Chib's and Chernoff and Zachs' approaches place a severe restriction on the duration of regimes. That is, the regime change is most likely to happen in the first period of the regime, then next most likely in the second period and so on. Below we examine the implications of this aspect of the model of Chib (1998). We show that a further interesting issue arises at the end of the sample relating to the imposition of a fixed number of change-points. In essence, this issue relates to what sort of prior information is required to ensure that the specified number of change-points do in fact occur. We show how the requisite prior has some unusual (and potentially undesirable) properties which can have a major impact on empirical practice. In light of this, we develop other priors which do not have these properties. An additional advantage of our model is that we do not have to assume a fixed number of change-points. Furthermore, we show how such an approach is an attractive one for forecasting in the presence of structural change.

The plan of this paper is as follows. In Section 2 we develop the insight of Chib (1998) on the relationship between change-point models and hidden Markov chains with a particular focus on the role of prior information. In particular, we show that the prior required to impose a specified number of change-points leads to a non-time homogeneous Markov chain at the end of the sample. Chib's (1996) algorithm is valid for non-time homogeneous Markov chains and we show how this algorithm can be used in the change-point model. We also discuss an apparently plausible "noninformative" prior. We discuss some undesirable properties of both these priors. In Section 3 we show how these undesirable properties have important implications for empirical practice using the coal mining disaster data analyzed in Chib (1998). Section 4 shows how these undesirable properties can be corrected through a Uniform prior with support not restricted to in-sample periods. Section 5 generalizes this prior in a hierarchical fashion in a manner which nests a wide variety of specifications and allows for the estimation of the number and timing of change-points. Section 6 concludes.

2 Priors for Change-point Models with Known Number of Change-points

We begin by describing in detail some recent work and, in particular, the popular model of Chib (1998) which has been used in many applications. Our focus is on extending Chib's insight of converting the classical change-point problem into a Markov mixture model and using the algorithm of Chib (1996) to estimate the change-points and the parameters within each regime.

We adopt the following notation (similar to that used in Chib's papers): We have data on a scalar time series variable, y_t for $t = 1, \ldots, T$ and let $\mathbf{Y}_i = (y_1, \ldots, y_i)'$ denote the history through time i and denote the future by $\mathbf{Y}^{i+1} = (y_{i+1}, \ldots, y_T)'$. Regime changes depend upon a discrete random variable, s_t, which takes on values $\{1, 2, \ldots, M\}$. We let $S_i = (s_1, .., s_i)'$ and $\mathbf{S}^{i+1} = (s_{i+1}, \ldots, s_T)'$. The likelihood function is defined by assuming $p(y_t|Y_{t-1}, s_t = m) = p(y_t|Y_{t-1}, \boldsymbol{\theta}_m)$ for a parameter vector $\boldsymbol{\theta}_m$ for $m = 1, \ldots, M$. Thus, change-points occur at times τ_m defined as

$$\tau_m = \{t : s_{t+1} = m + 1, s_t = m\} \text{ for } m = 1, \ldots, M - 1. \tag{2.1}$$

Chib (1998) puts a particular structure on this framework by assuming that s_t is Markovian. That is,

$$\Pr(s_t = j|s_{t-1} = i) = \begin{cases} p_i & \text{if } j = i \neq M \\ 1 - p_i & \text{if } j = i + 1 \\ 1 & \text{if } i = M \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

In words, the time series variable goes from regime to regime. Once it has gone through the m^{th} regime, there is no returning to this regime. It goes through regimes sequentially, so it is not possible to skip from regime m to regime $m+2$. Once it reaches the M^{th} regime it stays there (i.e. it is assumed that the number of change-points in the sample is known). In Bayesian language, (2.2) describes a hierarchical prior for the vector of states.[3]

To avoid confusion, it is worth stressing that change-point models can be parameterized in different ways. Many models indicate when each regime occurs by parameterizing directly in terms of the change-points (i.e. $\tau_1, .., \tau_{M-1}$). Others are written in terms of states which denote each regime (i.e. $\mathbf{S}_T$). It is also possible to write models in terms of durations of regimes. In the following material, we use all of these parameterizations, depending on which best illustrates the points we are making. However, we do stress that they are equivalent. So, for instance, a time series of 100 data points with a break at the 60^{th} can be expressed as $\tau_1 = 60$, or $S_{60} = 1$ and $\mathbf{S}^{61} = 2$, or $d_1 = 60$ and $d_2 = 40$ (where d_m denotes the duration of regime m).

There are many advantages to adopting the framework of Chib (1998). For instance, previous models typically involved searching over all possible sets of change-points. If the number of change-points is even moderately large, then computational costs can become overwhelming [see, for instance, the discussion in Elliott and Muller (2003) of the approach developed in Bai and Perron (1998)]. By using the Markov mixture model, the posterior simulator is recovering information on the most likely change-points given the sample and the computational burden is greatly lowered, making it easy to estimate models with many change-points.

Chib chose to model the transition probabilities of the states as having a constant hazard. This is similar to Chernoff and Zachs (1964) who assumed a constant probability of transition (although Chib allowed the transition probability to be different for different regimes). One consequence of the constant hazard is that regime duration satisfies a Geometric distribution. The Geometric distribution is decreasing in the duration and, thus, the implied distribution of the change-points also adopts this property.[4] For many applications, this might be sensible. However, for others it may be too restrictive. For instance, in the case of a single change-point, τ_1, is it always the case that earlier values of τ_1 should be preferred to later? The classical change-point literature implicitly reveals a preference for priors on τ_1 which are Uniform (i.e. before seeing the data, every value for τ_1, apart from initial conditions and endpoints, is treated as being equally likely). Such informal

[3] A non-Bayesian may prefer to interpret such an assumption as part of the likelihood, but this is merely a semantic distinction with no effect on statistical inference [see, e.g., Bayarri, DeGroot and Kadane (1988)].

[4] This statement will be qualified below in our discussion relating to endpoints.

discussion suggests we should at least investigate the consequences of this particular choice of hierarchical prior and consider possible alternatives.

Equation (2.2) defines a hierarchical prior for the states. To complete the model, a prior for p_m is required. Chib (1998) and subsequent papers have assumed this to be a Beta prior with hyperparameters $\underline{\delta}_1, \underline{\delta}_2$.[5] See, e.g., Poirier (1995), pages 104-105 for the definition and properties of the Beta distribution. In this paper, we will refer to the change-point model with hierarchical prior given by (2.2) with the Beta prior for the transition probabilities as the *Chib model*. Note, also, that in the following material, we discuss *hierarchical priors* for various features (e.g. in the Chib model, the hierarchical prior for durations is Geometric and depends upon the transition probabilities which in turn have their own Beta prior) as well as *marginal priors* (e.g. in the Chib model, we can derive a marginal prior for the durations by integrating out the transition probabilities using their Beta prior). It is important for the reader to keep clear these two types of priors. Note that the marginal prior probability for the regime durations for the Chib model is:

$$p(d_m) = \frac{B(\underline{\delta}_1 + d_m - 1, \underline{\delta}_2 + 1)}{B(\underline{\delta}_1, \underline{\delta}_2)}, d = 1, 2 \ldots, . \tag{2.3}$$

where $B(\underline{\delta}_1, \underline{\delta}_2) = \frac{\Gamma(\underline{\delta}_1)\Gamma(\underline{\delta}_2)}{\Gamma(\underline{\delta}_1 + \underline{\delta}_2)}$ is the Beta function. It can be confirmed that if $\underline{\delta}_2 \leq 1$ then the expected duration does not exist. Further, $p(d_m) > p(d_m + 1)$ so that this distribution is monotonically decreasing. This illustrates a point we have mentioned above: this prior implies that regime durations of d are more likely than $d + 1$. Note that this property is present both both in the hierarchical prior, $p(d|p_m)$, and the marginal prior, $p(d)$.

There is another, possibly undesirable, property of the Chib model which is not as obvious and which a naive researcher could miss. This property can be most easily seen in the context of the posterior simulator commonly used with this model. Bayesian inference in the model of Chib (1998) is based on a Markov Chain Monte Carlo (MCMC) algorithm with data augmentation. If $\Theta = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_M')'$ and $P = (p_1, \ldots p_{M-1})$ then the algorithm proceeds by sequentially drawing from

$$\Theta, P|\mathbf{Y}_T, \mathbf{S}_T \tag{2.4}$$

[5]Throughout this paper, we use a notational convention where lower bars (e.g. $\underline{\delta}_1$) denote prior hyperparameters chosen by the researcher. Thus, when we use hierarchical priors, only the hyperparameters at the final stage of the hierarchy will have lower bars.

and

$$\mathbf{S}_T|\mathbf{Y}_T,\Theta,P. \tag{2.5}$$

Simulation from the latter is done using a method developed in Chib (1996). This involves noting that:

$$\begin{aligned} p\left(\mathbf{S}_T|\mathbf{Y}_T,\Theta,P\right) &= p\left(s_T|\mathbf{Y}_T,\Theta,P\right)p\left(s_{T-1}|\mathbf{Y}_T,\mathbf{S}^T,\Theta,P\right) \\ &\cdots .p\left(s_t|\mathbf{Y}_T,\mathbf{S}^{t+1},\Theta,P\right)\cdots p\left(s_1|Y_T,\mathbf{S}^2,\Theta,P\right). \end{aligned} \tag{2.6}$$

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

$$p\left(s_t|\mathbf{Y}_T,\mathbf{S}^{t+1},\Theta,P\right) \propto p\left(s_t|\mathbf{Y}_t,\Theta,P\right)p\left(s_{t+1}|s_t,P\right). \tag{2.7}$$

Since $p\left(s_{t+1}|s_t,P\right)$ is the transition probability and the integrating constant can be easily obtained (conditional on the value of s_{t+1}, s_t can take on only two values in the change-point case), we need only to worry about $p\left(s_t|\mathbf{Y}_t,\Theta,P\right)$. Chib (1996) recommends the following recursive strategy. Given knowledge of $p\left(s_{t-1}=m|Y_{t-1},\Theta,P\right)$, we can obtain:

$$p\left(s_t=k|\mathbf{Y}_t,\Theta,P\right) = \frac{p\left(s_t=k|\mathbf{Y}_{t-1},\Theta,P\right)p\left(y_t|\mathbf{Y}_{t-1},\boldsymbol{\theta}_k\right)}{\sum_{m=k-1}^{k}p\left(s_t=m|\mathbf{Y}_{t-1},\Theta,P\right)p\left(y_t|\mathbf{Y}_{t-1},\boldsymbol{\theta}_m\right)}, \tag{2.8}$$

using the fact that

$$p\left(s_t=k|\mathbf{Y}_{t-1},\Theta,P\right) = \sum_{m=k-1}^{k} p_{mk}p\left(s_{t-1}=m|\mathbf{Y}_{t-1},\Theta,P\right), \tag{2.9}$$

for $k=1,..,M$. The recursive algorithm is started with $p\left(s_t|Y_1,\Theta,P\right)$.

Thus, the algorithm proceeds by calculating (2.7) for every time period using (2.8) and (2.9) beginning at $t=1$ and going forward in time (we will refer to this as the *forward iteration*). Then the states themselves are drawn using (2.6), beginning at period T and going backwards in time (we will refer to this as the *backward iteration*).

This algorithm was designed for the general Markov mixture model of Chib (1996). However, special considerations apply for the change-point model. In particular, the fact that exactly M regimes are assumed to exist in the observed sample of size T implies that in period 1 we are in regime 1 and in period T we are in regime M. In the algorithm described above, this implies that $p\left(s_T = M | Y_T, \Theta, P\right) = 1$ and $p\left(s_1 = 1 | Y_T, S^2, \Theta, P\right) = 1$ for all possible observed data $\mathbf{Y}_T$. It is tempting to think that one can simply impose these restrictions by fixing $s_1 = 1$ at the start of the forward iteration and, in the backward iteration, fixing $s_T = M$. However, this is not enough. At this point, we explain in detail why this is so, as it illustrates some restrictive properties on this model and motivates our own more flexible specification.

Chib (1996) assumes a standard time homogeneous Markov transition matrix in the development of his algorithm. But for the change-point problem, we have to relax this homogeneity assumption. Under the structure in (2.2) for the change-point problem, there is nothing to ensure that all the regimes are visited with positive probability in all samples of size T starting from regime 1. One way to see this is to consider the duration of each regime. As we have seen, by construction the duration of each regime, d_m, has a Geometric distribution and, thus, the expected duration is given by $1/(1-p_m)$. Without further restrictions, there is positive probability that even the second regime will not be reached in a sample size of T.

One solution to this problem, consistent with the *a priori* belief of M regimes in the observed sample of size T, is to relax the assumption of time homogeneity on the transition probabilities to ensure that all regimes are covered in a sample of T.[6] Thus in the case of $M = 4$, at times $T-3$ through $T-1$ we would have the following structure for the transition matrix:

$$\begin{array}{ccc} T-3 & T-2 & T-1 \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & p_2 & 1-p_2 & 0 \\ 0 & 0 & p_3 & 1-p_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} p_1 & 1-p_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & p_3 & 1-p_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} & \begin{bmatrix} p_1 & 1-p_1 & 0 & 0 \\ 0 & p_2 & 1-p_2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Thus, we must alter the hierarchical prior for s_{T-3}, s_{T-2} and s_{T-1} to ensure that all regimes occur (for other time periods, we maintain the prior given in equation 2.2). However, this new prior requires a slight modification of the algorithm presented Chib (1996).

Consider the backward iteration which begins by fixing $s_T = M$. There are two feasible regimes for s_{T-1}, M or $M-1$. Equation (2.6) involves s_{T-1} being drawn from $p\left(s_{T-1} | \mathbf{Y}_T, \mathbf{S}^T, \Theta, P\right) \propto p\left(s_{T-1} | \mathbf{Y}_{T-1}, \theta, P\right) p\left(s_T | s_{T-1}, P\right)$.

[6] Barry and Hartigan (1993) provide a different solution.

Consider the term $p\left(s_T|s_{T-1},P\right)$. The unmodified version of Chib's algorithm will use probabilities

$$\Pr[s_T=M|s_{T-1}=M]=1, \Pr[s_T=M|s_{T-1}=M-1]=1-p_{M-1},$$

which has to be changed to the equal probabilities:

$$\Pr[s_T=M|s_{T-1}=M]=\Pr[s_T=M|s_{T-1}=M-1]=1 \tag{2.10}$$

in constructing $p\left(s_T|s_{T-1},P\right)$. In many applications the estimates of p_{M-1} tend to be close to 1 suggesting that considerable bias is introduced at this initial step if the unmodified version of Chib's algorithm is used.

Now consider $p\left(s_{T-1}|\mathbf{Y}_{T-1},\theta,P\right)$ from the forward iteration. If a time homogeneous transition matrix is used and $M\geq 3$ then this contains positive probability on regimes $1,\ldots,M-2$ that are infeasible at time $T-1$. Thus, the non-time homogenous transition matrix must ensure that the forward iteration begins removing low numbered states at time $T-M$. This discussion shows how the algorithm must be modified for drawing s_{T-1}. The same sort of modifications must be made at times $t=T-2,\ldots,T-M$ in updating $p\left(s_t|\mathbf{Y}_t,\theta,P\right)$ to $p\left(s_{t+1}|\mathbf{Y}_t,\theta,P\right)$.

In summary, it is not difficult computationally to impose the restriction that exactly M regimes exist in the data by simply adding more prior information of the sort given in (2.10) and the algorithm of Chib (1996) can be used. However, it is possible that such prior information is undesirable in the sense of placing a great deal of prior weight in favor of change-points near the end of the sample.

Such considerations suggests that we may wish to investigate other priors. An apparently plausible thing to do would be to consider a Uniform distribution [see Koop and Potter (2001) for a Bayesian application and Bai and Perron (1998) and subsequent literature for the frequentist application] for the prior for the change-points or, equivalently, the durations. Below, we will refer to this prior as the Restricted Uniform prior since it restricts the prior to impose a fixed number of change-points on the model. In the case of a single change-point, such an approach is straightforward. Simply setting:

$$p\left(\tau_1\right)=\frac{1}{T-1} \text{ for } \tau_1=1,..,T-1 \tag{2.11}$$

is unambiguously "flat" and imposes exactly one regime change in the sample. However, such a prior does not generalize well to more than one change-point. We illustrate this in the case with two change-points, τ_1 and τ_2. The apparently sensible extension of (2.11) would write the prior as $p(\tau_1, \tau_2) = p(\tau_1) p(\tau_2|\tau_1)$ where

$$p(\tau_1) = \frac{1}{T-2} \text{ for } \tau_1 = 1, .., T-2 \tag{2.12}$$

$$p(\tau_2|\tau_1) = \frac{1}{T-\tau_1-1} \text{ for } \tau_2 = \tau_1+1, .., T-1. \tag{2.13}$$

This prior does impose exactly two regime changes in sample and the prior for $\tau_2|\tau_1$ appears noninformative. However, it can be verified that, if we integrate out τ_1, the marginal prior for τ_2 is very non-Uniform, once again giving more weight to change-points late in the sample:

$$p(\tau_2 = j) = \frac{1}{T-2} \sum_{i=2}^{j} \frac{1}{T-i} \text{ for } j = 2, .., T-1. \tag{2.14}$$

This prior (for a sample size of 112) is shown in Figure 1b. Its shape is the reverse of a Geometric distribution. Similar results and prior shapes hold when we have $M > 3$ regimes in our sample. Is this property of our Restricted Uniform prior undesirable? Of course, this depends on the empirical context. However, there is a possibility that a researcher could use this prior, thinking it is "noninformative", but empirical results could be affected by the greater prior weight for τ_2 near the end of the sample.

3 Prior Sensitivity in Change-Point Models in Practice

In the previous section, we showed how two plausible priors: the one underlying the Chib model and what we have called the Restricted Uniform prior have some possibly undesirable properties. In this section, following Chib (1998), we investigate the empirical performance of our priors in a commonly-used data set. We consider the coal mining disaster data of Jarrett (1979) and consider the cases of zero, one or two change-points.

The prior for the Chib model requires the selection of prior hyperparameters for the transition probabilities, $\underline{\delta}_1$ and $\underline{\delta}_2$. To aid in prior elicitation, an examination of (2.3) indicates that, for values of $\underline{\delta}_2$ close to zero and $\underline{\delta}_1$ relatively large, we have $p(d_m) \approx p(d_m+1)$ for larger values of d. Perhaps reflecting a preference for priors which are "flat" over possible change-points, in many applications [for example Chib (1998) Kim and Nelson (1999) and Kim, Nelson and Piger (2003)] $\underline{\delta}_2$ has been set to less than 1. In Chib (1998) the choice $\underline{\delta}_2 = 0.1$ and $\underline{\delta}_1 = 8$ was made. We will also use these values for the case with one change-point. Note that this implies that the marginal prior of regime duration is approximately flat, but the expected duration does not exist (see discussion after equation 2.3). Hence, such a choice of prior hyperparameters may get around the possibly unattractive property that the prior favors shorter durations to longer ones, but raises the possibility that the problems relating to the restrictive prior required to ensure exactly M regimes exist may be exacerbated.

The model of Chib requires us to add restrictions analogous to (2.10) to ensure that number of change-points assumed in the model do in fact occur. This can be done by truncating the distribution of the duration of regime 1 at $d = T - 1$ and assigning all the remaining probability to this point. In the coal mining disaster data there are 112 observations, hence, for the one change-point case we can use (2.3) to set

$$P[\tau_1 = 111] = P[d_1 = 111] = 1 - \sum_{d=1}^{110} \frac{B(7+d, 1.1)}{B(8, 0.1)} = 0.76. \tag{3.1}$$

Thus, the prior required to impose exactly one change-point is allocating a great deal of weight to regime changes at the end of the sample. The prior required to impose exactly two change-points is the obvious extension of (3.1) and has a similar property. The properties of this prior can be seen in Figure 1a for the one change-point case (similar patterns hold for cases with more change-points). There is a huge spike in the prior at the end of the sample. Other than this, these choices for $\underline{\delta}_1$ and $\underline{\delta}_2$ imply that the prior is fairly flat, although the property that the duration distribution is decreasing can be seen at the beginning of the sample.

To carry out our application, we need to specify a likelihood and a method for posterior analysis. Since this data is a count of mining disasters by year a Poisson likelihood is reasonable. Chib assumes the priors on the Poisson intensities in the different regimes, θ_m, to be $G(\underline{\alpha}_m, \underline{\beta}_m)$[7] for $m = 1, \ldots, M$. Throughout the following material, we use the same values for $\underline{\alpha}_m, \underline{\beta}_m$ as in Chib (1998). That is, for the zero and one change-point cases, we set

[7] $G(a, b)$ denotes the Gamma distribution with mean ab and variance ab^2.

$\underline{\alpha}_m = 2, \underline{\beta}_m = 1$. For the two change-point case, we set $\underline{\alpha}_m = 3, \underline{\beta}_m = 1$. Under these assumptions the posterior of the change-points can be found directly as described in the Appendix. Note there is no need to use a Gibbs sampler to do this nor to use a Markov chain interpretation of the regimes.

Posterior results for the Chib model, the change-point model with the Restricted Uniform prior and something which we call the Unrestricted Uniform prior (and will explain later) are given in Table 1 and Figures 2 and 3. Figure 1 plots the priors (except the obvious ones which, e.g., are simply Uniform).

For the case where a single change-point is assumed to exist, the Chib model with (3.1) imposed (see Figure 1a) and the Restricted Uniform prior yield essentially the same posterior so, for the sake of brevity, they are not plotted. However, when we assume two change-points we begin to see the effects of prior assumptions. For this case, with Chib's model we follow Chib (1998) and assume independent Beta priors for the two transition probabilities, each with hyperparameters 5 and 0.1 (and impose prior restrictions on the endpoints analogous to equation 3.1). The posteriors for τ_1 and τ_2 for the Chib model and Restricted Uniform prior are plotted in Figures 2a,b and 3a,b, respectively. Note that the sample ends in 1962. The reason for the X-axis of these figures running past 1962, along with an explanation of Figures 2c and 3c will be given in the next section.

For the first change-point, the posteriors under the two priors are roughly the same (although some differences occur at the beginning of the sample). However, for the second change-point the two priors are yielding substantively different posteriors. Note in particular that the posterior of τ_2 for Chib's model has a huge spike at the end of the sample. This is due to the restriction analogous to (3.1) which imposes exactly two change-points in-sample.[8] Thus, we are finding prior sensitivity even when staying in the class of models which impose a precise number of change-points.

The reader may suspect that our findings of prior sensitivity are occurring in a model which is not supported by this data. To investigate this issue, we report marginal likelihoods for the various models. Note that these can be calculated analytically since, conditional on the change-points, a closed form expression for the marginal likelihood exists (see Appendix). These conditional marginal likelihoods can be averaged over the change-points using the

[8] In the two change-point case, we were unable to exactly re-produce the posteriors in Figure 6 of Chib (1998).

appropriate prior to yield an exact (unconditional) marginal likelihood. Table 1 contains the log marginal likelihoods calculated in this way. It presents strong evidence that at least one change-point is present. The one and two change-point models receive roughly equal support. Note also that the models with Uniform priors receive more support from the data than the Chib model. Thus, we are finding sensitivity to the prior in a model which does receive appreciable support from the data. Furthermore, one might argue that it is the more reasonable prior (i.e. the Restricted Uniform prior) that is receiving more support from the data.[9] We take this as additional motivation for extending the traditional approach.

The last row of Table 1 presents the log of the marginal likelihood for a Uniform prior which treats the number of change-points as unknown. We have not explained what this is. However, note that the model with this prior receives more support from the data than the other 2 change-point models. We now turn to this prior.

Table 1: Log Marginal Likelihoods for Different Priors

Model	**Log Marginal Likelihood**
No change-points	-206.21
1 change-point Chib model	-178.35
1 change-point Restricted Uniform	-176.76
2 change-point Chib model	-178.96
2 change-point Restricted Uniform	-177.35
Unrestricted Uniform	-177.19

4 Uniform Priors on Change-Points: Unknown Number of Change-points

In the previous section, we used what we called a Restricted Uniform prior. This had the conditional priors of $\tau_m|\tau_{m-1}$ being Uniform. However, we argued that, in terms of the marginal prior for τ_m for $m > 1$, it was very non-flat indeed. Let us now consider what would happen if we worked which is "flat" in another sense. We will illustrate using the two change-point case, with

[9] For the case of no change-points the marginal likelihood is very close to the estimate of Chib (1998) which was based on the simulation approach in Chib (1995). However, as one would expect, given the discussion in Section 2, for the log marginal likelihoods of the change-point models there are some differences

the extension to more change-points being obvious. As with the Restricted Uniform prior we begin with $p(\tau_1, \tau_2) = p(\tau_1)\, p(\tau_2|\tau_1)$ and assume

$$p(\tau_1) = \frac{1}{T-2} \text{ for } \tau_1 = 1, .., T-2. \tag{4.1}$$

However, we replace (2.13) by

$$p(\tau_2|\tau_1) = \frac{1}{T-2} \text{ for } \tau_2 = \tau_1 + 1, .., T + \tau_1 - 2. \tag{4.2}$$

Note that this prior, which we refer to as the Unrestricted Uniform prior, now has the very sensible property that both $p(\tau_2|\tau_1)$ and $p(\tau_2)$ are Uniform. However, it also has the unconventional property that it allocates prior weight to change-points outside the observed sample. We will argue that this is a highly desirable property since, not only does this prior not place excessive weight on change-points near the end of the sample, but also there is a sense in which it allows us to handle the case where there is an unknown number of change-points. That is, the prior given by (4.1) and (4.2) does impose that there are two change-points, but since one of them can occur out of sample, it implicitly allows for one (in-sample) change-point as well. The generalization discussed in the following section is even more flexible.

For our coal-mining data, the last row of Table 1 presents the marginal likelihood using the Unrestricted Uniform prior as given in (4.1) and (4.2). It is substantively higher than the marginal likelihoods for the other priors for the two change-point models.[10] As with our other prior, this marginal likelihood is a weighted average of the marginal likelihoods for every possible (τ_1, τ_2) combination. However, unlike the other priors, the set of possible values for (τ_1, τ_2) include some where τ_2 is out-of-sample. The marginal likelihood will be the same for these. For instance, in our data set with $T = 112$, the (conditional) marginal likelihoods for $(\tau_1 = 70, \tau_2 = 115)$ and $(\tau_1 = 70, \tau_2 = 116)$ will be the same (and both will indicate one change-point models). When calculating the (unconditional) marginal likelihood presented in Table 1 we include these two identical (conditional) marginal likelihoods. It can be confirmed that this implies a very attractive property: the set of models with one change-point in sample receive the same amount of prior weight as the set of models with two change-points in sample.

[10] For the one change-point model the Restricted and Unrestricted Uniform priors are identical.

Figures 2c and 3c present the posteriors for τ_1 and τ_2 using the Unrestricted Uniform prior. In-sample these posteriors look quite different from those obtained from the Chib model, but similar in shape to those from the Restricted Uniform prior. Out-of-sample (i.e. after the year 1962) the implications of our Unrestricted Uniform prior can be seen. Note the substantive probability allocated Uniformly in the region of (roughly) 1963-2000. This is the region where the model effectively becomes a single change-point model. Since Table 1 indicates that the single change-point model receives considerable support, it is not surprising that we are recovering this through the posterior for τ_2 allocating substantial weight out-of-sample. After (roughly) 2000, the probability drops off. This is due to the fact that the posterior of τ_1 is concentrated in the late 1880s/early 1890s. Models with τ_1 beyond (roughly) this value receive little support and, since τ_2 is defined on the integers $\tau_1 + 1, .., T + \tau_1 - 2$, models with $\tau_2 > 1890 + T$ receive little support. Note that the Unrestricted Uniform prior as defined in (4.1) and (4.2) does not allocate any prior probability to models with no breaks. In the following section, we discuss a generalization which does.

Another way of examining the effect of the various priors is to examine the predictive distribution, $p(y_{T+1}|\mathbf{Y}_T)$, for an out-of-sample observation. The Appendix describes how Bayesian predictive inference can be done. For the Chib model and the Restricted Uniform prior, Bayesian model averaging across models with differing number of change-points can be done in a straightforward fashion by weighting the resulting forecast distribution by the posterior probabilities of the various change-points models. The latter can be directly calculated from the log marginal likelihoods in Table 1. Predictions using the Unrestricted Uniform prior already implicitly average across models with differing numbers of change-points.

It can be seen that the calculation of the predictive distributions depends crucially on the prior over the change-points assumed. Figure 4a provides a benchmark for comparison. It is the forecast distribution provided by the model with no change-points. Figures 4b, 4c and 4d are forecast distributions provided using our three different classes of prior. For the Chib model and Restricted Uniform prior, we average over one and two change-point models. The Chib model prior yields a forecast distribution which is very different from either of the Uniform priors. Note that Figure 4b (as well as 4a) both place probability of more than 1/3 on two or more disasters per year. However, our Uniform priors yield forecast distributions (see Figures 4c and 4d) which indicate roughly 1/5 chance of two or more mining disasters per year.

This reinforces a central message of this paper: priors matter in change-point models.

It is worth mentioning that an unrestricted version of the prior used in the Chib model can be derived in an analogous manner to what we have done with our Uniform priors (i.e. by not imposing a restriction such as (3.1) and allocating prior weight outside of the observed sample). For brevity we do not do this here. Such a prior does not have the poor properties seen in Figure 1a and the resulting posteriors look a bit more like those found using the Uniform priors. But substantive differences exist and, in this application, marginal likelihoods indicate that Uniform priors are preferred.

5 A Generalization of the Unrestricted Uniform Prior

Another way of looking at these issues is to note that many of the undesirable properties of the priors discussed in the previous section arise since they assume a fixed number of change-points. Any model which makes such an assumption is going to require a prior that imposes this feature and this will typically lead to undesirable properties. By relaxing this assumption we can obtain more sensible priors. As a bonus, this relaxation also allows us to treat the case where the number of change-points is unknown and to forecast breaks out-of-sample. Below we derive a particular prior which has, we argue, attractive properties, but we stress that many other attractive possible priors exist when we enter a world where some change-points can occur out of sample. For instance, in this paper we focus on priors that are Uniform over durations: a common noninformative choice. In Koop and Potter (2004), we develop an alternative model with a Poisson hierarchical prior over durations.

In the previous section, we worked with (4.1) and (4.2), which allowed for two breaks which could occur at any time period. The case where there are possibly many more than two breaks can be handled in a straightforward manner. However, if there are potentially many breaks such priors might not be sensible. Much recent empirical work has found evidence of many change-points. For instance, Pastor and Stambaugh (2001) find 30 change-points in a financial application. In macroeconomics, time-varying parameter (TVP) models have become quite popular [e.g. Cogley and Sargent (2001), (2003)].

TVP models implicitly allow for $T-1$ change-points. Note, however, that the prior given in (4.1) and (4.2) can be interpreted as implicitly favoring models with few breaks. For instance, if the researcher believes many change-points are plausible then it is also plausible that the first change-point will occur relatively early in the sample. In this case, even a prior such as (4.1) could be interpreted as allocating too much weight to the end of the sample. Accordingly, an empirical useful generalization of our Unrestricted Uniform prior allows for the regime durations (or, equivalently, the change-points themselves) to be Uniform over a bounded interval.

For the case where there is a maximum of $M-1$ breaks in-sample, we write our prior as $p(\tau_1, \tau_2, .., \tau_{M-1}) = p(\tau_1) \sum_{j=2}^{M-1} p(\tau_j | \tau_{j-1})$ and assume

$$p(\tau_1) = \frac{1}{[cT]} \text{ for } \tau_1 = 1, .., [cT]. \tag{5.1}$$

and

$$p(\tau_j | \tau_{j-1}) = \frac{1}{[cT]} \text{ for } \tau_j = \tau_{j-1} + 1, .., \tau_{j-1} + [cT] \tag{5.2}$$

Note that this prior has all the desirable properties of the Unrestricted Uniform prior discussed in the previous section, but introduces a scalar parameter c which controls the maximum duration of each regime. The notation $[cT]$ indicates the smallest integer such that $cT \leq [cT]$. Thus, if $c = \frac{1}{T}$ we obtain the TVP model, whereas as c becomes larger we obtain priors which place more weight on models with fewer regimes. For instance, our prior of the previous section had a maximum of two change-points in-sample and set $c = \frac{T-2}{T}$. At the extreme, the researcher might wish to consider values for c in the interval $\left[\frac{1}{T}, 2\right]$ as this would nest everything from the TVP model (with a break every period) through a model which allocates appreciable (i.e. 50%) prior weight to a model with no change-points at all (in-sample). In practice, the researcher would likely wish to consider a much narrow range of values for c.

Note also that the researcher would want to choose M with likely values of c in mind. As an extreme case, note that if the researcher chooses $M = 2$ and $c = \frac{1}{T}$, then the prior imposes that a change-point must occur after $T = 1$, but the choice of M implies there are no additional change-points in the model. Hence, such a prior would imply a very odd pattern of structural change.

The researcher can, of course, simply choose any values of M and c and estimate the model as outlined in previous sections (see Appendix for details).[11] However, in the common case where the number of change-points in-sample is unknown, it is desirable to treat c as an unknown parameter and estimate it from the data. In a Bayesian context, this involves choosing a hierarchical prior for c, $p\left(c|\underline{a}\right)$, where $\underline{a}$ is a vector of hyperparameters selected by the researcher. For instance, the researcher could set $p\left(c|\underline{a}\right) = U\left(\underline{a}_1, \underline{a}_2\right)$ where $\underline{a}_1$ and $\underline{a}_2$ are upper and lower bounds selected to cover a suitable range of change-points.

Using the methods outlined in the Appendix, it is straightforward to derive methods of posterior and predictive inference for this model when the marginal likelihood conditional on the change-points has an analytical form. For the case where no analytical form is known, an MCMC algorithm which builds on Chib (1996) is described in the Appendix. For forecasting, the use of the hierarchical prior has a great advantage in that the updating of c implies that information in-sample can be used for predicting the likelihood of a break out-of-sample. Many authors have argued that poor forecasting performance of many macroeconomic models is largely due to structural breaks [see, among many others, Clements and Hendry (1999) or Pesaran, Pettenuzzo and Timmerman (2004)]. In light of this issue, a model, such as the one introduced here, which attempts to model the probability of out-of-sample change is potentially of great use.[12]

In the coal mining disaster data, there is fairly clear evidence of one or two breaks, but not more. Hence, this empirical example is not well designed to show the advantages of our Generalized Uniform prior. However, if we set $M = 2$ and $p\left(c|\underline{a}\right) = U\left(\frac{1}{2}, 2\right)$ we obtain predictive distributions which are quite similar to those given in Figures 4b and 4c. Furthermore, the posterior mode occurs at $c = \frac{1}{2}$ and, for this value, the log of the marginal likelihood is -176.71 which is better than any of the other priors. Marginal likelihoods have a strong reward for parsimony and it is reassuring to see that this (less parsimonious) model is out-performing the (more parsimonious) one change-

[11] If the researcher does not wish to choose M, it can simply be set to $T-1$, thus allowing for the maximum feasible number of breaks in-sample, and let c alone control the number of breaks. As noted in the Appendix, such an approach can be computationally demanding.

[12] Hierarchical priors for regime-specific likelihood parameters, θ_m, can also be of use in improving forecast performance in the presence of structural change. See Koop and Potter (2004) for such an approach.

point models despite the fact that Table 1 indicates only weak evidence in favor of the presence of a second change-point. And it is worth stressing that, with our Generalized Uniform prior, we did not need to assume a fixed number of change-points (in-sample). We are successfully recovering the reasonable inferences from other models, without making the assumptions that were necessary in those other models.

6 Conclusions

In this paper, we have discussed prior elicitation in change-point models. We have shown how some common and apparently sensible priors have potentially undesirable properties. Relaxing these priors to eliminate these properties results in priors which allocate probability to change-points occurring out-of-sample. Much of the paper is devoted to showing how this apparently odd property actually is highly desirable, leading to a model which effectively allows for the number of change-points to be unknown. Furthermore, by allowing for change-points out-of-sample, we obtain a model which is highly beneficial for forecasting and treats in a sensible manner the issues raised in, e.g., Maheu and Gordon (2004) and Pesaran, Pettenuzzo and Timmerman (2004). Simple methods for predictive and posterior inferences under all our priors have been developed and illustrated using the coal mining disaster data of Jarrett (1979).

7 References

Bai, J. and Perron, P., 1998, Estimating and testing linear models with multiple structural changes, *Econometrica* 66, 47-78.

Barry, D. and Hartigan, J., 1993, A Bayesian analysis for change point problems, *Journal of the American Statistical Association*, 88, 309-319.

Bayarri, M., DeGroot, M. and Kadane, J., 1988, What is the likelihood Function? pp. 1-27 in *Statistical Decision Theory and Related Topics IV*, volume 1, edited by S. Gupta and J. Berger, New York: Springer-Verlag.

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

Chernoff, H. and Zacks, S., 1964, Estimating the current mean of a Normal distribution which is subject to changes in time, *Annals of Mathematical*

Statistics 35, 999-1018.

Chib, S. 1995, Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association,* 90, 1313-1321.

Chib, S., 1996, Calculating posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics* 75, 79-97.

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

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

Cogley, T. and Sargent, T., 2001, Evolving post-World War II inflation dynamics, *NBER Macroeconomic Annual*

Cogley, T. and Sargent, T., 2003, Drifts and volatilities: monetary policies and outcomes on Post WWII US, mimeo UC Davis.

Elliott, G. and Muller, U., 2003, Optimally testing general breaking processes in linear time series models, manuscript available at

http://www.econ.ucsd.edu/~gelliott/RECENTPAPERS.htm.

Jarrett, R., 1979, A note on the interval between coal mining disasters, *Biometrika*, 66, 191-193.

Kim, C. and Nelson C. 1999 Has the US economy become more stable? A Bayesian approach based on a Markov switching model of the business cycle, *Review of Economics and Statistics,* vol. 81: 1-10.

Kim, C., Nelson, C. and Piger. J., 2003, The less volatile U.S. economy: A Bayesian investigation of timing, breadth, and potential explanations, working paper 2001-016C, The Federal Reserve Bank of St. Louis.

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

Koop, G. and Potter, S., 2004, Forecasting and estimating multiple change-point models with an unknown number of change-points, manuscript.

Maheu, J. and Gordon, S., 2004, Learning, forecasting and structural breaks, manuscript available at http://www.nd.edu/~meg/MEG2004/Gordon-Stephen.pdf .

McConnell, M. and Perez, G., 2000, Output fluctuations in the United States: What has changed since the early 1980s? *American Economic Review* 90 1464-76.

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

Pesaran, M.H., Pettenuzzo, D. and Timmerman, A., 2004, Forecasting time series subject to multiple structural breaks, manuscript available at http://www.econ.cam.ac.uk/faculty/pesaran/HMCMultiBreakJune04.pdf.

Poirier, D.J., 1995, Intermediate Statistics and Econometrics (The MIT Press, Cambridge).

Stephens, D., 1994, Bayesian retrospective multiple-changepoint identification, *Applied Statistics*, 43, 159-178.

8 Appendix

Posterior and Predictive Inference with Poisson Likelihoods

We use a Poisson likelihood with $m = 1, ..., M$ regimes. The Poisson intensity in each regime is denoted by θ_m and each of these intensities has an independent $G(\underline{\alpha}_m, \underline{\beta}_m)$ prior. In the case of no breaks the posterior distribution of θ_1 is:

$$p(\theta_1|\mathbf{Y}_T) = \frac{\left\{\prod_{t=1}^{T} \exp(-\theta_1)\theta_1^{y_t}\right\} \theta_1^{\underline{\alpha}_1 - 1} \exp(-\theta_1/\underline{\beta}_1)}{m_0 \prod_{t=1}^{T} y_t! \Gamma(\underline{\alpha}_1) \underline{\beta}_1^{\underline{\alpha}_1}}, \tag{A.1}$$

where m_0 is the marginal likelihood of the observed data (assuming no breaks) and $\{y_t\}$ is the number of coal mining disasters in year t. Noting that the numerator is in the form of a Gamma density with parameters $\underline{\alpha}_1 + \sum_{t=1}^{T} y_t, [1/\underline{\beta}_1 + T]^{-1}$, the marginal likelihood can be expressed as

$$m_0 = \frac{\Gamma(\underline{\alpha}_1 + \sum_{t=1}^{T} y_t)[1/\underline{\beta}_1 + T]^{-(\underline{\alpha}_1 + \sum_{t=1}^{T} y_t)}}{\prod_{t=1}^{T} y_t! \Gamma(\underline{\alpha}_1) \underline{\beta}_1^{\underline{\alpha}_1}} \tag{A.2}$$

Conditional on knowing the change-points, comparable results for $m > 1$ regimes involve using equations like (A.1) and (A.2) defined over the relevant sub-periods. For instance, in the case of a single change-point, the marginal likelihood conditional on the change-point τ_1 is given by:

$$m_1 = \frac{\Gamma(\sum_{t=1}^{\tau_1} y_t + \underline{\alpha}_1)\Gamma(\sum_{t=\tau_1+1}^{T} y_t + \underline{\alpha}_2)\left[1/\underline{\beta}_1 + \tau_1\right]^{-\sum_{t=1}^{\tau_1} y_t + \underline{\alpha}_1}\left[1/\underline{\beta}_2 + T - \tau_1\right]^{-\sum_{t=\tau_1+1}^{T} y_t + \underline{\alpha}_2}}{\Gamma(\underline{\alpha}_1)\underline{\beta}_1^{\underline{\alpha}_1}\Gamma(\underline{\alpha}_2)\underline{\beta}_2^{\underline{\alpha}_2}\prod y_t!} \tag{A.3}$$

This defines the marginal likelihood conditional on τ_1. To get an unconditional marginal likelihood, the appropriate prior can be used to integrate out τ_1. Since τ_1 is a discrete random variable this integration is straightforward.

Noting that $m_1 = p(\mathbf{Y}_T|\tau_1)$ and that Bayes rule implies:

$$p(\tau_1|\mathbf{Y}_T) \propto p(\mathbf{Y}_T|\tau_1)\,p(\tau_1)\,, \tag{A.4}$$

it follows that knowledge of m_1 and $p(\tau_1)$ are all that is required to carry out posterior inference on change-points. The case with many change-points is the obvious extension.

With regards to prediction, it can be shown that, conditional on the last change-point τ_m, $p(y_{T+1}|\mathbf{Y}_T, \tau_m)$ has a Negative Binomial distribution with probability of success:

$$\frac{1/\underline{\beta}_m + T - \tau_m}{1/\underline{\beta}_m + 1 + T - \tau_m},$$

and number of successes:

$$\underline{\alpha}_m + \sum_{t=\tau_m+1}^{T} y_t.$$

Using the fact that $p(y_{T+1}|\mathbf{Y}_T) = \sum p(y_{T+1}|\mathbf{Y}_T, \tau_m)\,p(\tau_m|\mathbf{Y}_T)$, it follows that the unconditional predictive distribution, $p(y_{T+1}|\mathbf{Y}_T)$, is a probability weighted combination of the Negative Binomial distributions for $p(y_{T+1}|\mathbf{Y}_T, \tau_m)$ with the weights being given by $p(\tau_m|\mathbf{Y}_T)$.

If a change-point occurs out-of-sample there is no data information to learn about and, hence, its prior equals its posterior. For instance, if $\tau_1 > T$, (A.4) would imply $p(\tau_1|\mathbf{Y}_T) \propto p(\tau_1)$. In this case, we would also have $m_1 = m_0$. That is, the one-break model with change-point occurring out-of-sample reduces to the no-break model. Such considerations do not effect the methods of posterior and predictive inference described above. Furthermore, as discussed in the body of the text, for the Uniform prior such an approach has the desirable property that the classes of models with differing numbers of change-points receive equal prior weight.

Generalization of Uniform Prior

For the Generalization of the Uniform prior involving a hierarchical prior for c, minor modifications of the preceding material are required. All of the probability distributions above are the same if we condition on c. For instance, $m_1 = p(\mathbf{Y}_T|\tau_1, c), p(\tau_1|\mathbf{Y}_T, c)$ is given by (A.4), etc.. The methods described previously involved integrating over change-points using their prior (i.e. to get the unconditional marginal likelihood) or using (conditional) marginal likelihoods to work out the posterior for change-points (i.e. see A.4). The same methods can be used to integrate out c or calculate its posterior.

In cases where analytical integration results are not available or the maximum number of change-points is very large, then the following MCMC algorithm based on Chib (1996, 1998) can be used.

Our Generalized Uniform prior is flat over durations. It implies the following transition probabilities (see Section 2):

$$P[s_{t+1} = m+1|s_t = m, d_m] = \begin{cases} 0 & \text{if } m > t \\ \frac{1}{cT - d_m + 1} & \text{if } d_m < cT \\ 1 & \text{if } d_m = cT \end{cases}, m = 1, \ldots, M$$

where d_m is the duration of regime m. If $M < T$ then we also need to add in the restriction $P[s_t = M|s_{t-1} = M] = 1$. As discussed in Section 2, Chib's 1996 algorithm applies to this non-time homogeneous Markov chain. Hence, the formulae in Section 2 are relevant with the transition probabilities given above. Unfortunately, the computational burden is greater since we have to follow both the regime and duration (i.e. the transition probabilities now depend on d_m). Koop and Potter (2004) shows how the initial condition for the chain can be found when this is unknown. Here we simplify and assume $P[|s_1 = 1, d_1 = 1] = 1$.

In order to understand the computational difficulties of this algorithm, consider the case $c = (T-1)/T$ and $M = T$. Thus, in time period 1 we only have 1 possible state, in time period 2 we have 2 possible states, in time period 3 we have 4 possible configurations to track in our algorithm:

$$[\{s_3 = 1, d_1 = 3\}, \{s_3 = 2, d_2 = 2\}, \{s_3 = 2, d_2 = 1\}, \{s_3 = 3, d_3 = 1\}].$$

In a sample of size T we would have $T - 1 + T - 2 + \cdots + 1 = T(T-1)/2$ configurations by the last observation. This, of course, is many more than with Chib (1998)'s model. Nevertheless, the algorithm is computationally feasible for moderate T and, if desired, larger/smaller values of c/M can be chosen to greatly reduce the computational complexity.

When we treat c as an unknown parameter, the MCMC algorithm can be extended appropriately. Conditional on the durations, we can work out a convenient conditional posterior for drawing c. Define

$$\underline{c} = \max(d_1/T, d_2/T, \ldots, D_r/T, \underline{a}_1),$$

where d_m is the realized duration of (completed) regime m, D_r is the ongoing duration of the last regime observed (for $r = 1, 2, .., M$) and $\underline{a}_1$ is the lower bound of support of c in the prior. The likelihood of a completed regime, d_m/T is given by

$$\frac{1}{c}, c \geq \underline{c}$$

and for an ongoing regime

$$1 - \sum_{d=1}^{D_r/T} \frac{1}{c}, c \geq \underline{c}$$

thus the overall likelihood is

$$\left(\frac{1}{c}\right)^{r-1} \left(c - \frac{D_r}{T}\right), \; c \geq \underline{c}.$$

This likelihood can be combined with a prior to form a conditional posterior. Depending on the exact for of the prior, a suitable algorithm (e.g. rejection sample, Metropolis-Hastings or evaluating the posterior at a grid of values for c) can easily be obtained.

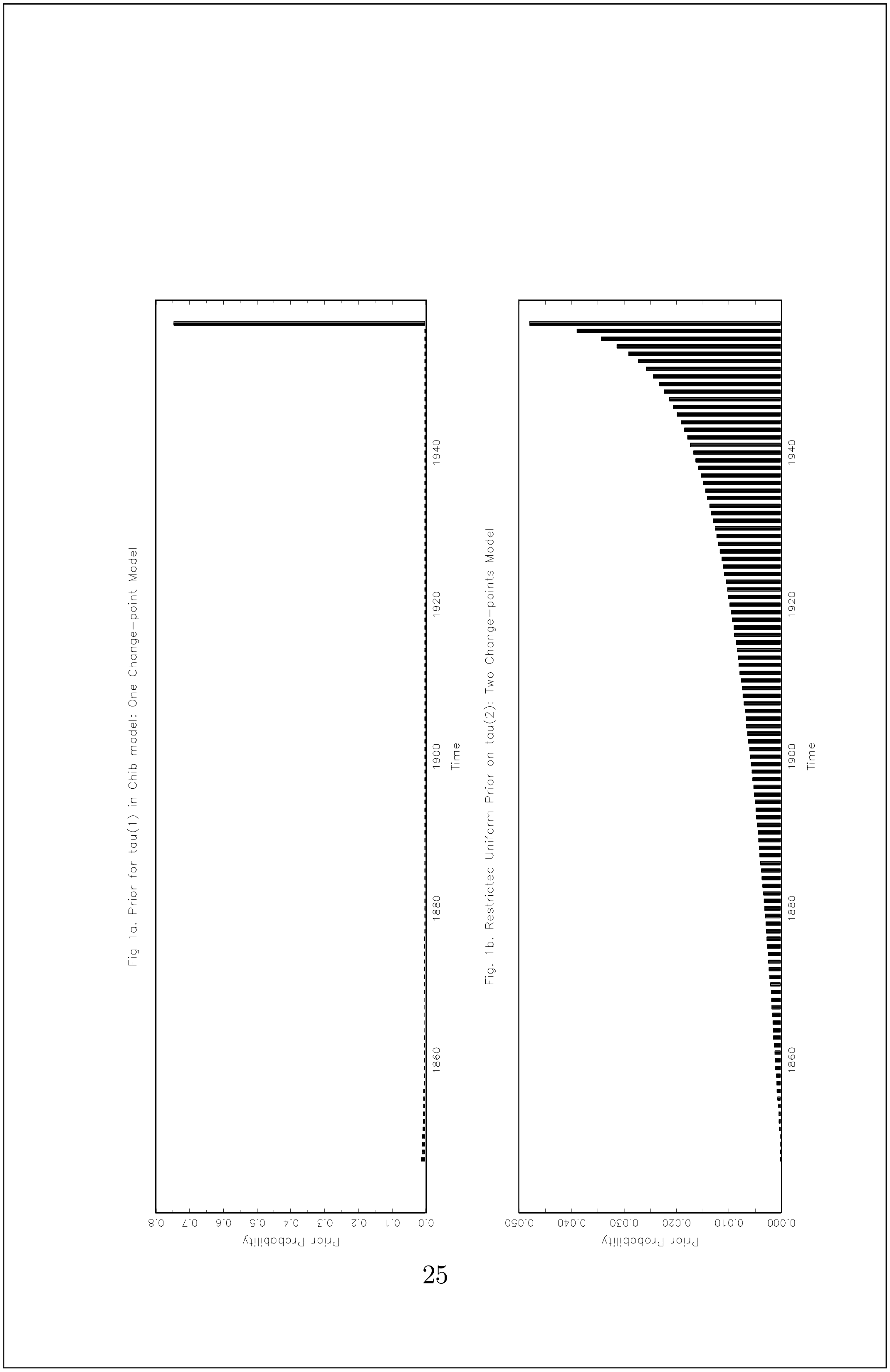
Fig 1a. Prior for tau(1) in Chib model: One Change-point Model
Prior Probability
Time
Fig. 1b. Restricted Uniform Prior on tau(2): Two Change-points Model
Prior Probability
Time
25

Figure 1:

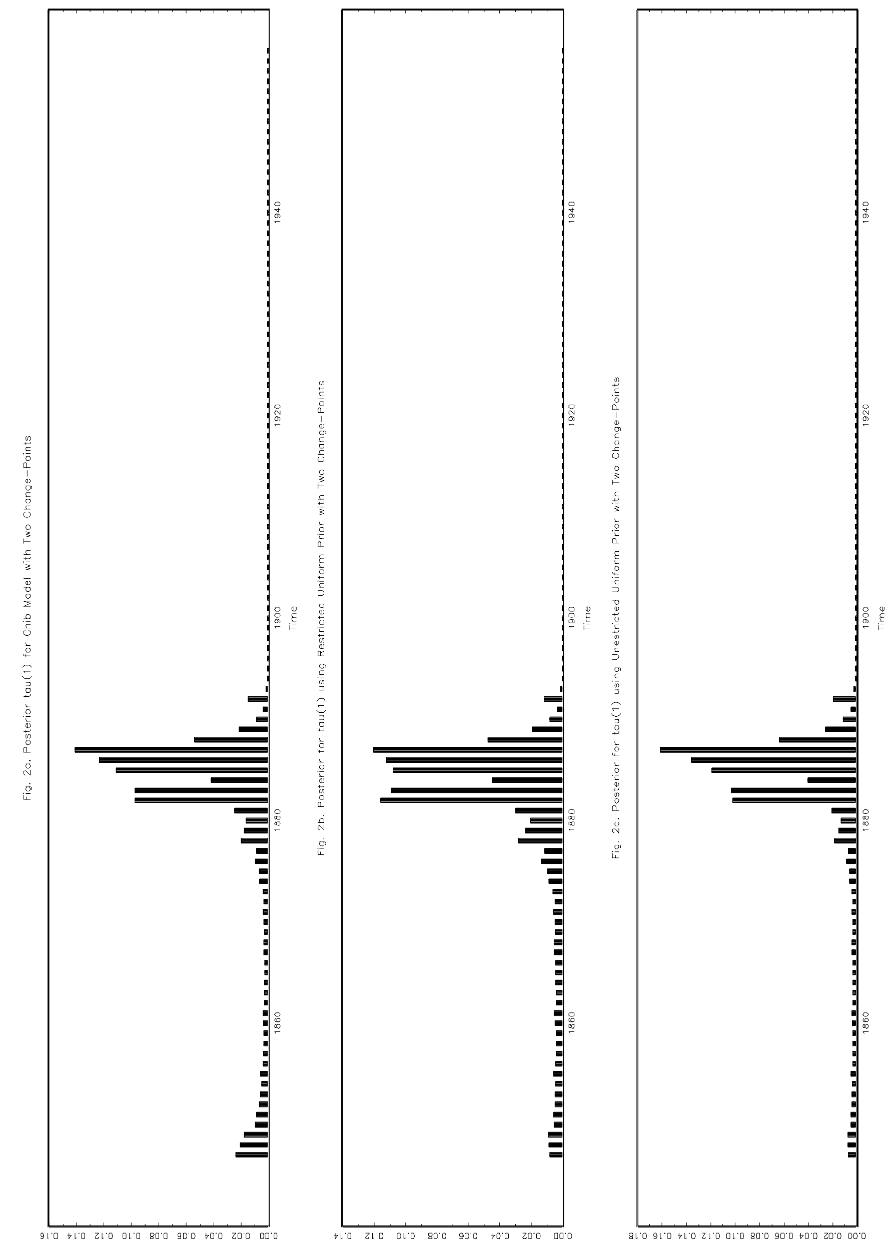
Fig. 2a. Posterior tau(1) for Chib Model with Two Change-Points
Fig. 2b. Posterior for tau(1) using Restricted Uniform Prior with Two Change-Points
Fig. 2c. Posterior for tau(1) using Unestricted Uniform Prior with Two Change-Points
Time
Posterior Probability

Figure 2:

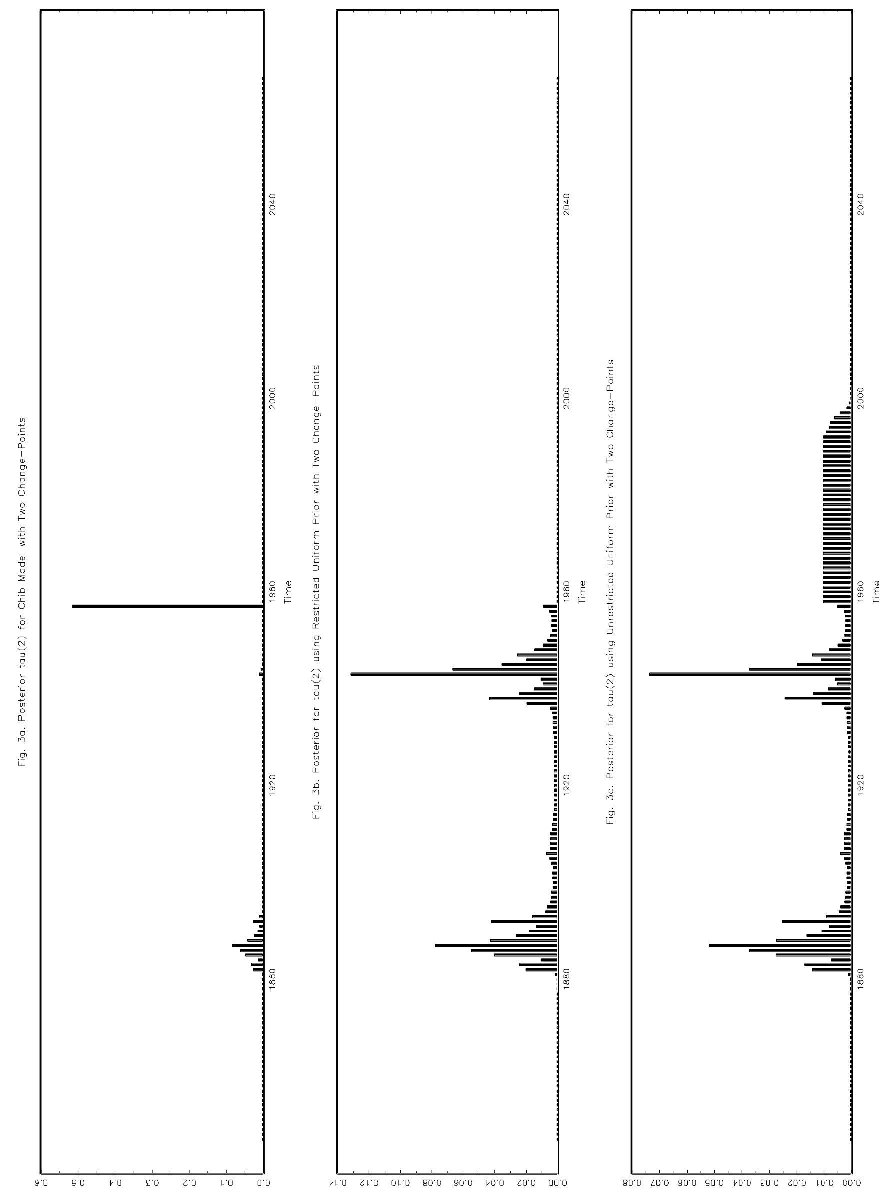
Fig. 3a. Posterior tau(2) for Chib Model with Two Change-Points
Fig. 3b. Posterior for tau(2) using Restricted Uniform Prior with Two Change-Points
Fig. 3c. Posterior for tau(2) using Unrestricted Uniform Prior with Two Change-Points
Posterior Probability
Time
1880
1920
1960
2000
2040

Figure 3:

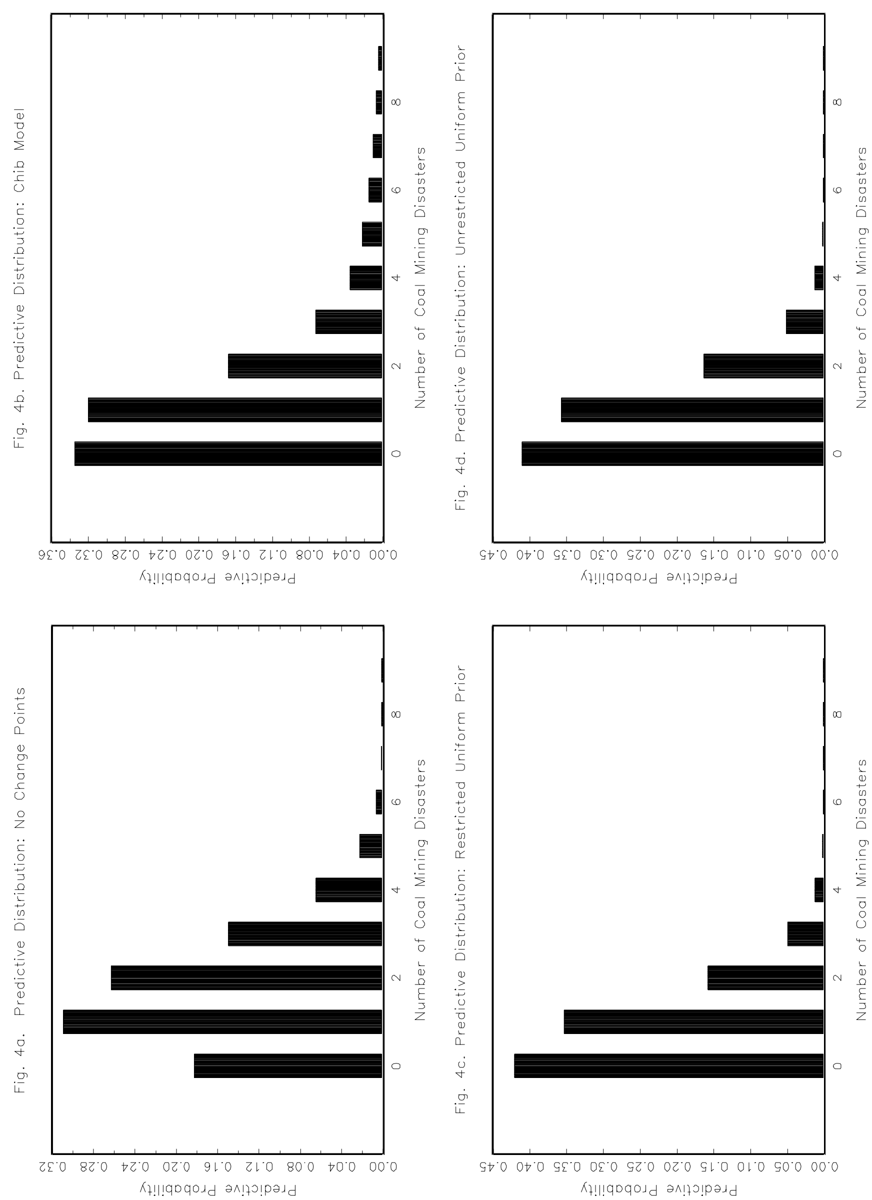
Fig. 4a. Predictive Distribution: No Change Points
Fig. 4b. Predictive Distribution: Chib Model
Fig. 4c. Predictive Distribution: Restricted Uniform Prior
Fig. 4d. Predictive Distribution: Unrestricted Uniform Prior
Predictive Probability
Number of Coal Mining Disasters

Figure 4: