

Prior Elicitation in Multiple Change-point Models*

Gary Koop

Department of Economics
University of Strathclyde
email: Gary.Koop@strath.ac.uk

Simon M. Potter

Macroeconomics and Monetary Studies Function
Federal Reserve Bank of New York
email: simon.potter@ny.frb.org

September 2006, revised May 2007

Abstract

This paper discusses Bayesian inference in change-point models. The main existing approaches either attempt to be noninformative by using a Uniform prior over change-points or use an informative hierarchical prior. Both these approaches assume a known number of change-points. We show how they have some potentially undesirable properties and discuss how these properties relate to the imposition of a fixed number of change-points. We develop a new Uniform prior which allows some of the change-points to occur out-of sample. This prior has desirable properties, can reasonably be interpreted as “non-informative” and handles the case where the number of change-points

*We would like to thank Edward Leamer for useful conversations 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.

is unknown. We show how the general ideas of our approach can be extended to informative hierarchical priors. With artificial data and two empirical illustrations, we show how these different priors can have a substantial impact on estimation and prediction even with moderately large data sets.

1 Introduction

Change-point modeling has become popular due to an increasing awareness of the importance of this issue for empirical practice.¹ For instance, papers such as Ang and Bekaert (2002) and Stock and Watson (1996) have documented the widespread structural instability in many macroeconomic and financial time series. The importance of this for empirical finance and macroeconomics cannot be overstated. Empirical work which ignores this instability can miss important patterns in the data and can result in misleading policy advice. As one example (among many), there has recently been interest in the volatility of US real activity which appears to have been greatly reduced in the last few decades. This finding is sometimes referred to as the Great Moderation of the business cycle. For instance, Kim, Nelson and Piger (2004) investigate breaks in the volatility of various measures of aggregate activity. For most of the measures they consider, they find strong evidence of an abrupt break in the early 1980s. Stock and Watson (2002) find similar evidence for a change in volatility, but find the decline to have been more gradual. Empirical research which ignores this change in volatility risks, e.g., substantially overestimating the uncertainty in GDP growth forecasts.

We should define from the beginning what change-point models are discussed in this paper. We consider ones where the parameters of the likelihood function change at discrete points in time, $\tau_1, \dots, \tau_{M-1}$ and, thus, there the data is divided into M regimes. So, for instance, if the likelihood function is obtained from a regression model using data on a dependent variable y_t and explanatory variables (in the vector x_t), we have

¹In 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.

$$\begin{aligned}
y_t &= x_t' \beta_1 + \sigma_1 \varepsilon_t \text{ if } t \leq \tau_1 \\
y_t &= x_t' \beta_2 + \sigma_2 \varepsilon_t \text{ if } \tau_1 + 1 \leq t \leq \tau_2 \\
&\cdot \\
&\cdot \\
y_t &= x_t' \beta_M + \sigma_M \varepsilon_t \text{ if } t \geq \tau_{M-1}
\end{aligned} \tag{1.1}$$

where ε_t is i.i.d. standard Normal. Equation 1.1 defines a likelihood function for the model (although the methods discussed in this paper can be used with any likelihood function). However, Bayesian inference requires a prior and this is what the present paper is about.

As a digression, we stress that, in this paper, we are focussing on prior elicitation with regards to the change-points, $\tau_1, \dots, \tau_{M-1}$ and not the parameters which characterize the likelihood function within a regime (e.g. β_1, \dots, β_M and $\sigma_1, \dots, \sigma_M$). This is not because we think prior elicitation for the latter parameters is unimportant, but merely to focus the discussion on the change-point issue. Indeed, in Koop and Potter (2007), we have stressed how priors linking coefficients across regimes can be very useful when forecasting (see also Pastor and Stambaugh, 2001, and Pesaran, Pettenuzzo and Timmerman, 2007).² Classical analyses of change-point models almost invariably assume that, after a break occurs, anything can happen (e.g. β_2 and β_1 can potentially be completely different) and the present paper makes a similar assumption. However, we stress that the basic insights of the present paper would still hold if we used more sophisticated priors which link parameters in different regimes in some fashion.

Our starting point, in the next section of this paper, will be a prior which is Uniform over the set of all possible change-points. For the Bayesian, Uniform priors are often used to be noninformative; to express a lack of prior information. This can be thought of as the Bayesian analogue of what the classical econometric literature does. That is, *a priori* every change-point is treated as being equally likely. In the case of a single change-point, the definition of a Uniform prior is relatively non-controversial. But with more than one change-point we show how an apparently sensible “noninformative” prior can be very informative indeed in a potentially undesirable manner. We use this insight to develop a new prior which can be more sensibly interpreted as

²Furthermore, time varying parameter (TVP) models are also a popular choice when faced with structural instability and they involve close links across regimes through state equations (e.g. $\beta_t = \beta_{t-1} + u_t$).

“noninformative”. However, a property of this new prior is that it allows for change-points to occur out-of-sample. We show how this apparently unusual property is actually a very desirable one since it allows us to treat the number of change-points in-sample as unknown (and, thus, estimate it).

In contrast to the classical literature, the existing Bayesian literature often uses informative hierarchical priors for the change-points. Perhaps the most popular Bayesian approach is that of Chib (1998). In the third section of this paper, 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 a prior which imposes a specified number of change-points also leads to potentially undesirable behavior at the end of the sample. We discuss how such behavior can be avoided with an extension of our new prior.

This discussion of prior elicitation may sound abstract and of little practical importance for the classical econometrician and practical Bayesian alike. The former may feel prior elicitation issues are not relevant for classical econometrics. However, the issues which we discuss (e.g. the imposition of a fixed number of change-points) have their parallels in classical econometrics. Furthermore, many of our “prior elicitation” issues can be thought of more generally as modeling issues. For instance, the hierarchical prior of Chib (1998) would be thought of as part of the likelihood function by the classical econometrician. The posterior simulation algorithm of Chib (1998) could (with minor adaptations) be used to carry out maximum likelihood estimation. Thus, the issues we discuss are of relevance for the classical econometrician.

The practical Bayesian may feel that minor changes in the prior are unimportant if one has a moderately large data sets. A key point of this paper is to show that these are not minor changes, but ones that can have substantive effects on posterior and predictive inference, even with sample sizes of relevance in macroeconomics. We make this point by using artificial data sets and by providing two empirical illustrations. The first of these uses US real GDP growth data from 1947Q2 through 2006Q4 to investigate the Great Moderation of the business cycle. The second uses the coal mining disaster data analyzed in Chib (1998).

2 Uniform Priors for Change-points

2.1 Theoretical Considerations

The classical econometric literature on change-points (see, among many others, Bai and Perron, 1998) typically gives equal weight, before seeing the data, to every possible change-point (apart, perhaps from a “prior” restriction that each regime must contain $X\%$ of the observations). The Bayesian wishing to proceed in a comparably “noninformative” manner would turn to the class of discrete Uniform distributions for prior elicitation. In the case where there is a single change-point, τ_1 , such an approach is straightforward. Simply setting:

$$p(\tau_1) = \frac{1}{T-1} \text{ for } \tau_1 = 1, \dots, T-1 \quad (2.1)$$

yields an unambiguously “flat” prior and imposes exactly two regimes in the sample ($M = 2$).

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 (2.1) would write the prior as $p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2|\tau_1)$ where

$$p(\tau_1) = \frac{1}{T-2} \text{ for } \tau_1 = 1, \dots, T-2 \quad (2.2)$$

$$p(\tau_2|\tau_1) = \frac{1}{T-\tau_1-1} \text{ for } \tau_2 = \tau_1 + 1, \dots, T-1. \quad (2.3)$$

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

$$p(\tau_2 = j) = \frac{1}{T-2} \sum_{i=2}^j \frac{1}{T-i} \text{ for } j = 2, \dots, T-1. \quad (2.4)$$

This prior (for a sample size of 100) is shown in Figure 1. 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 this

Uniform prior undesirable? Of course, this depends on the empirical context (something we will investigate below). 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.

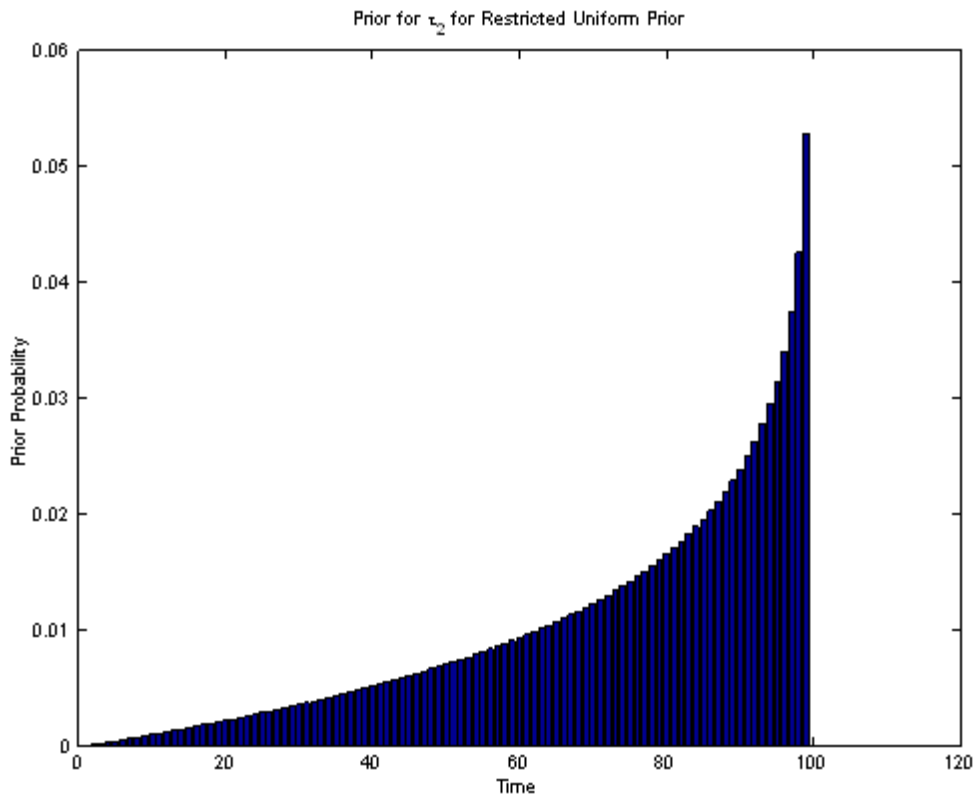


Figure 1

We will refer to priors such (2.2) and (2.3) as Restricted Uniform priors since they restrict the prior to impose a fixed number of change-points on the model. Let us now consider what would happen if we worked with a prior which is “flat” in another sense. We again will illustrate using the two change-point case, with the extension to more change-points being obvious. We refer to this new prior as the Unrestricted Uniform prior since, as we shall see, it does not restrict the model to have a fixed number of change-points

in-sample. As with the Restricted Uniform prior we begin with $p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2|\tau_1)$ and assume

$$p(\tau_1) = \frac{1}{T-2} \text{ for } \tau_1 = 1, \dots, T-2. \quad (2.5)$$

However, we replace (2.3) by

$$p(\tau_2|\tau_1) = \frac{1}{T-2} \text{ for } \tau_2 = \tau_1 + 1, \dots, T + \tau_1 - 2. \quad (2.6)$$

Note that this prior has the very sensible property that both $p(\tau_2|\tau_1)$ and $p(\tau_2)$ are Uniform and have same number of points of support. There is no pile-up of prior probability near the end of the sample as in Figure 1. 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 (2.5) and (2.6) 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. We will elaborate on these points below.

We now describe our general Unrestricted Uniform prior which allows for a maximum of $M-1$ change-points in-sample and adds one possible useful generalization to (2.5) and (2.6). We write the Unrestricted Uniform prior as $p(\tau_1, \tau_2, \dots, \tau_{M-1}) = p(\tau_1) \sum_{j=2}^{M-1} p(\tau_j|\tau_{j-1})$ and assume

$$p(\tau_1) = \frac{1}{[cT]} \text{ for } \tau_1 = 1, \dots, [cT]. \quad (2.7)$$

and

$$p(\tau_j|\tau_{j-1}) = \frac{1}{[cT]} \text{ for } \tau_j = \tau_{j-1} + 1, \dots, \tau_{j-1} + [cT] \quad (2.8)$$

Note that this prior still has the desirable property that $p(\tau_1)$, $p(\tau_j|\tau_{j-1})$ and $p(\tau_j)$ (for $j = 2, \dots, M-1$) are all Uniform (and have the same number of points of support). Thus, it is truly noninformative with respect to the change-points. It also 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 time varying-parameter (TVP) model, whereas as c becomes larger we obtain priors which place more weight on models with fewer regimes. For instance, the prior in (2.5) and (2.6) 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 $[\frac{1}{T}, 2]$ 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 .

In our previous work, Koop and Potter (2007), we argued that it is rarely sensible to restrict the maximum duration of regimes ex ante. With the prior given in (2.7) and (2.8), provided $c \geq 1$, the maximum duration of regimes in-sample will not be restricted. Such a consideration could be useful when selecting c or in eliciting a prior for c .

Values for c and M can be selected by the researcher. However, it is also possible to treat them as unknown parameters and they can either be estimated or integrated out in the standard Bayesian fashion. To see how this would be done, suppose we have data on a time series variable, y_t for $t = 1, \dots, T$ and let $\mathbf{Y}_i = (y_1, \dots, y_i)'$ denote the history through time i . The marginal likelihood for given values of c , M and the change-points is given by:

$$p(\mathbf{Y}_T | \tau_1, \dots, \tau_{M-1}, c, M). \quad (2.9)$$

The exact form of this marginal likelihood will depend on the likelihood and prior for the parameters in each regime (below we will provide examples). Suppose first that the researcher has selected specific values of c and M . It is possible to base empirical results on a particular choices of $\tau_1, \dots, \tau_{M-1}$ (e.g. one can choose the values of $\tau_1, \dots, \tau_{M-1}$ which yield the maximum value for the marginal likelihood). However, the Bayesian would prefer to integrate out the change-points so that empirical results reflect the uncertainty associated with them. The rules of probability imply:

$$p(\mathbf{Y}_T | c, M) = \sum_{\tau_1} \dots \sum_{\tau_{M-1}} p(\mathbf{Y}_T | \tau_1, \dots, \tau_{M-1}, c, M) p(\tau_1, \dots, \tau_{M-1} | c, M), \quad (2.10)$$

where each summation is over all possible values of the appropriate change-point. The marginal likelihood in (2.10) could be used as a basis for empirical work. Provided an analytical form for (2.9) exists and M is not too large, (2.10) can be calculated in a straightforward manner by simply evaluating (2.9) at every possible change-point. Alternatively, various simulation algorithms (see, e.g., Koop and Potter, 2003) for drawing the change-points can be used to evaluate (2.10).

Now let us suppose the researcher is interested in treating c and M as random variables. This requires knowledge of their posterior, $p(c, M | \mathbf{Y}_T)$. But Bayes' rule tells us that:

$$p(c, M | \mathbf{Y}_T) \propto p(\mathbf{Y}_T | c, M) p(c, M). \quad (2.11)$$

Thus, (2.10) and a prior for c and M can be used to evaluate the requisite posterior. Since c and M are scalar parameters, (2.11) can be calculated by evaluating each component on the right-hand side at a grid of values for c and every possible choice for M . Alternatively, a posterior simulation algorithm (e.g. a Metropolis-Hastings algorithm) can be developed for c .

Our preferred empirical strategies are either to simply select values for c and M or to select a value for M (i.e. choose $M - 1$ as the maximum number of in-sample regimes that the researcher thinks is plausible) and treat c as an unknown parameter. For instance, many classical structural break papers using macroeconomic data allow for one or two breaks. Combining this consideration with a desire not to restrict the maximum duration of regimes *ex ante*, would suggest setting $M = 2$ and $c = T$.

We have now derived two priors: the Restricted and Unrestricted Uniform priors. We have seen how the former of these has an undesirable property (i.e. the pile-up of prior probability near the end of the sample) while the latter has an apparently unusual property (i.e. the fact that some of the change-points can occur out-of-sample). We will now provide a theoretical discussion of this latter property (an empirical consideration of both properties will be provided shortly). It is worth stressing that out-of-sample change-points cause no technical nor computational problems for the Bayesian. Bayesian analysis involves a prior and a likelihood function. The likelihood function, of course, reflects data information. If, say, τ_2 , occurs out-of-sample then there will be no data information about τ_2 and, thus, no likelihood information. However, there still will be prior information about τ_2 . Provided the prior is proper (as is the case for all priors used in this paper), a valid posterior

density for τ_2 exists. This is merely a reflection of the well-known Bayesian result that non-identification does not pose a difficulty for Bayesians, but that the posterior for a non-identified region of the parameter space will typically be equal to the prior (see, e.g., Poirier, 1998).³

In one sense, our Unrestricted Uniform prior can be thought of as a simple trick for avoiding the “pile-up of prior probability at the end-of-sample” problem of the Restricted Uniform prior. We would argue that this reason alone makes it a useful addition to the literature. However, the value of the Unrestricted Prior goes well beyond this in that it allows for us treat the number of change-points in-sample as unknown. For the same reason as the Restricted Uniform prior will have the “pile-up of prior probability at the end-of-sample” problem, *any* model with a fixed number of change-points will exhibit this problem. Thus, we argue that it is important to develop models where the number of change-points is unknown and estimated from the data. The discussion after (2.6) makes clear that the Unrestricted Uniform prior treats the number of change-points in-sample as unknown. Other models do exist with this property (e.g. Giordani and Kohn, 2006, Koop and Potter, 2007 and McCulloch and Tsay, 1993). However, the existing literature uses priors which are informative over the change-points. To our knowledge, the Unrestricted Uniform prior is the only existing prior which is noninformative over the change-points. Our empirical section will show how this prior works in practice.

2.2 Empirical Illustrations Using Uniform Priors

We illustrate our Restricted and Unrestricted Uniform priors in the context of the AR(p) model with change-points (i.e. equation 1.1 with x_t containing an intercept and p lags of the dependent variable). A prior is required for the coefficients in each regime and, for these, we choose the natural conjugate prior so that (conditional on $\tau_1, \dots, \tau_{M-1}, c$ and M) analytical results for the marginal likelihood (i.e. equation 2.9), posterior and predictive distributions are available (see, e.g., Koop, 2003, chapter 3). Throughout this section, we use a relatively noninformative prior which reflects a weak belief in stationarity for the AR coefficients. Of course, in a more substantive empirical exercise, a prior sensitivity analysis could be done or a more objective prior

³Formally, the prior for a non-identified parameter is equal to its posterior, unless it is *a priori* correlated with an identified parameter.

elicitation procedure used (e.g. the g-prior or a prior based on a training sample). To be precise, our prior for the intercept and AR coefficients in each regime is:

$$\beta_j | \sigma_j^2 \sim N(0, 0.25\sigma_j^2 I)$$

and

$$\sigma_j^{-2} \sim G(1, 1),$$

where $G(a, b)$ denotes the Gamma distribution with mean ab and variance ab^2 . Note that the prior for the error precision is relatively noninformative (its degrees of freedom parameter is two) but has mean one. In our example using real GDP growth, our data is measured as a percentage so that this prior covers a reasonable region of the parameter space.

There are many ways we could compare our different models and priors (and we illustrate several below). However, especially since many of the problems with priors which impose a fixed number of change-points occur at the end of the sample, predictive results are of particular relevance and these receive much emphasis in all of our empirical illustrations.

2.2.1 Illustrations Using Artificial Data

In this section, we will illustrate several aspects of our priors using artificial data generated from two different data generating processes (DGPs). Both of our DGPs have $T = 100$ and the error variances in each regime are set to one. The conditional mean within each regime is given by an AR(1) model with intercept set to zero. Our DGPs thus differ only in the number of change-points and the value taken for the AR(1) coefficient (which we will label β_j for $j = 1, \dots, M$) in each regime.

We calculate results for the Restricted and Unrestricted Uniform priors for $M = 1, 2, 3$ (with AR(1) processes within each regime). The $M = 1$ case is simply the AR(1) model without a change-point. For the Unrestricted Uniform prior we set $c = 2$. Given the likelihood and natural conjugate prior assumed, analytical results are available and we can simply evaluate the marginal likelihood for given change-points in (2.9) at every possible change-point and average as in (2.10) to produce an overall marginal likelihood for each model.

Our first DGP, DGP_1 , does not have any change-points and sets $\beta_1 = 0.5$ and illustrates how our Unrestricted Uniform prior yields more sensible inference than the Restricted Uniform prior. Figure 2 plots the posterior mean of β_1 for these two priors with $M = 2$. Since $c = 2$, the Unrestricted Uniform prior says that there is a 50% probability that the change-point occurs out of sample. Thus, in essence, it is allocating 50% of the prior probability to the (correct) model with no change-points. In Figure 2, the point estimate of the AR(1) coefficient resulting from the Unrestricted Uniform prior is close to a horizontal line near the true value of $\beta_1 = 0.5$, although the prior is pulling slightly towards the prior mean of zero. The Restricted Uniform prior, which says that there is a 100% probability of a change-point occurring, is far different, especially at the end of the sample (note that the posterior mean of the AR(1) coefficient is pulled down below 0.2 by the end of the sample). This shows the risks of incorrectly imposing a change-point on a model when none exists. We stress that the Unrestricted Uniform prior will never run this risk.

Figure 3, which plots the posterior for the change-point using the Unrestricted Uniform prior with $M = 2$ shows how this prior achieves this sensible result. Most of the posterior probability is allocated out-of-sample indicating that no change-points occur in-sample. The posterior out-of-sample is simply equal to the prior (and it, thus, flat) but this does not cause and problems for estimation (see Figure 1) nor prediction (see Table 1).

Table 1 presents the log of marginal likelihoods for the various models for this DGP. As expected, the true model (with no change-point) has a higher marginal likelihood. Regardless of whether $M = 2$ or $M = 3$ the Unrestricted Uniform prior yields a higher marginal likelihood than the Restricted Uniform Prior, despite the fact that it is noninformative over a much larger support.

Table 1 also presents predictive means and standard deviations for y_{T+1} for the various models and priors we are using. Since DGP_1 has no change-points, we would expect the predictive results for the AR(1) to be most reliable. Note that the Unrestricted Uniform priors are producing predictive results which are much closer to the AR(1) results than those from the Restricted Uniform prior. Clearly, incorrectly imposing a fixed number of change-points in-sample (as the Restricted Uniform prior does) can have negative implications for posterior and predictive inference. The use of the Unrestricted Uniform prior allows us to avoid this risk.

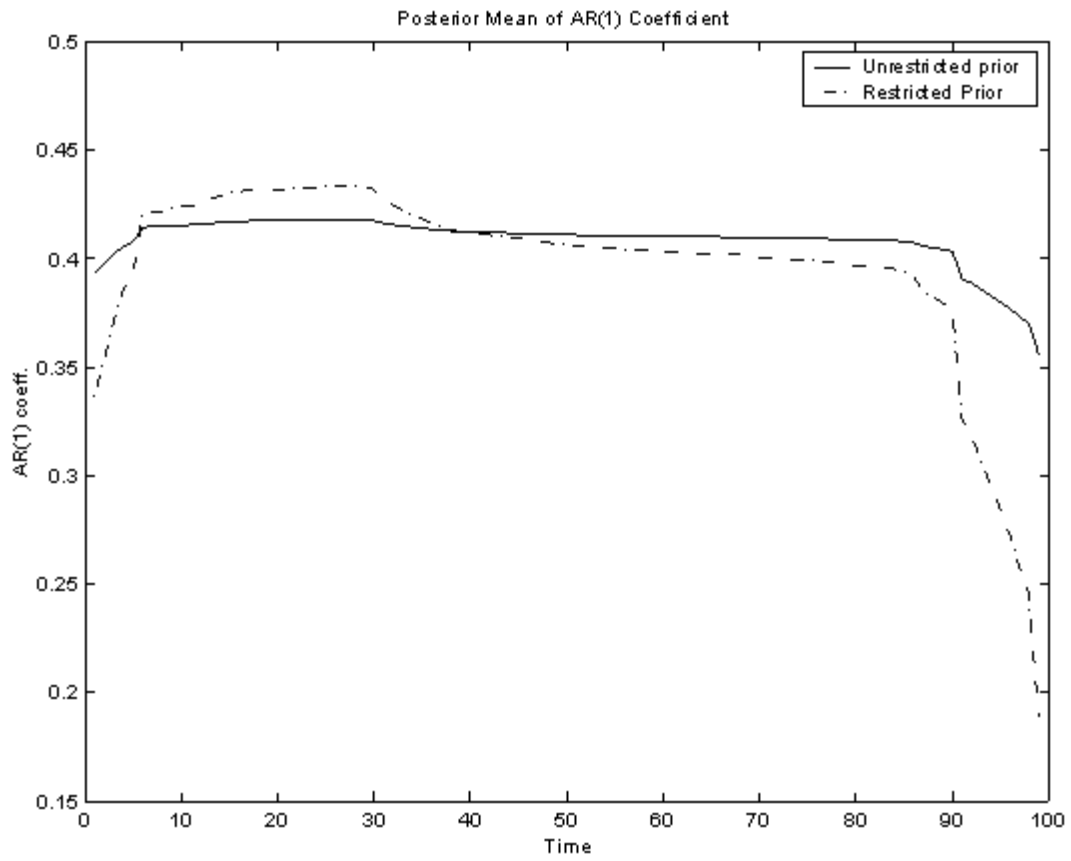


Figure 2: Comparing the AR Coefficient when the DGP Has No Breaks

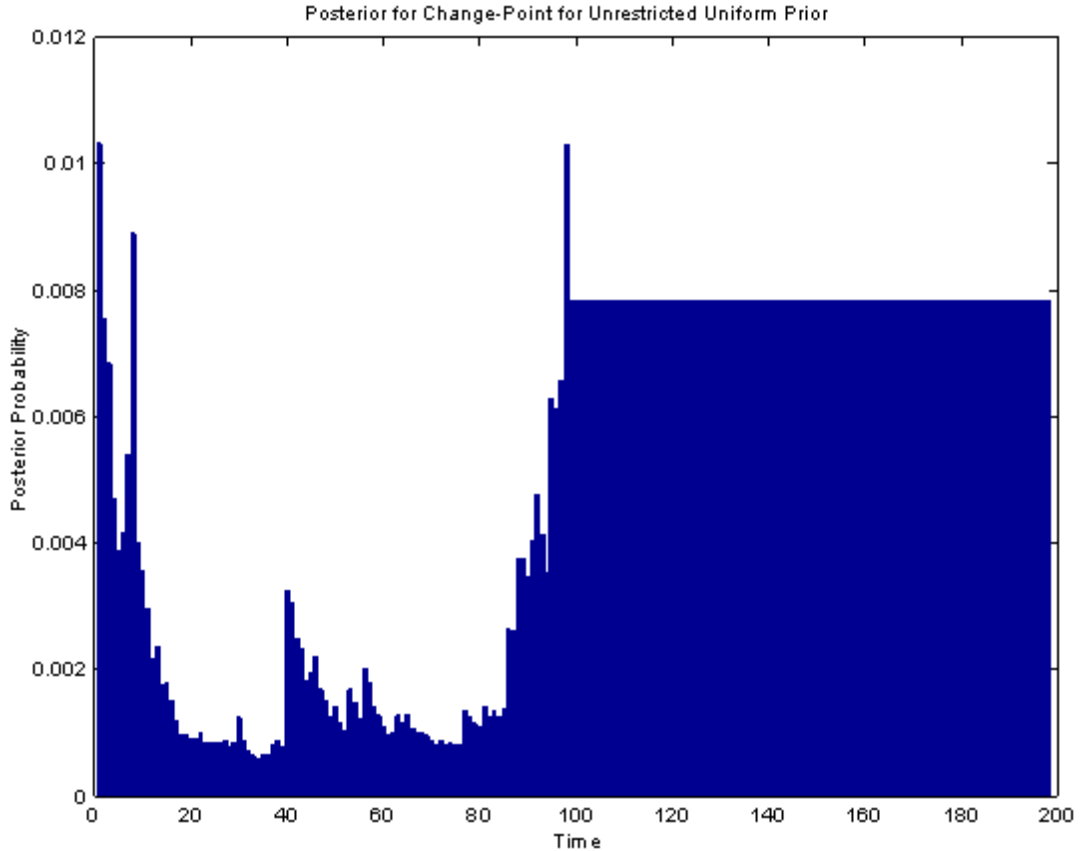


Figure 3: The Posterior of the Change-Point when the DGP Has No Breaks

| | Log Marginal Likelihood | Predictive Mean | Predictive St. Deviation |
|--------------------------|-------------------------|-----------------|--------------------------|
| AR(1) | -143.67 | -0.17 | 1.00 |
| Restricted ($M = 2$) | -145.99 | -0.16 | 1.06 |
| Restricted ($M = 3$) | -147.75 | -0.21 | 1.07 |
| Unrestricted ($M = 2$) | -145.32 | -0.17 | 1.01 |
| Unrestricted ($M = 3$) | -146.13 | -0.18 | 1.05 |

Our second DGP, DGP₂, is the same as DGP₁ with the exception that $\beta_1 = 0.0$ (which holds for $t \leq 50$) and $\beta_2 = 0.75$ (for other values of t). For this DGP, $M = 2$ is the correct choice and results presented in Table 2 indicate this. The Restricted and Unrestricted Priors with $M = 2$ have

the highest marginal likelihoods and their predictive densities have the same mean and standard deviation (to two decimal places). More interesting is what happens when $M = 3$. For this case, the Unrestricted Uniform prior is producing a predictive mean and standard deviation which are very similar to $M = 2$ results, whereas the Restricted Uniform prior is producing very different results. Clearly, the Restricted Uniform prior is yielding reasonable results when M is selected correctly and a well-defined change-point occurs in the middle of the sample. But in other cases the Restricted Uniform prior can be misleading (but the Unrestricted Uniform prior yields sensible results).

It is also worth noting that in DGP_2 with the Restricted Uniform prior, the marginal likelihood for the (correct) $M = 2$ model is only slightly higher than that of the (incorrect) $M = 3$ model. Thus, in a Bayesian model averaging exercise involving only the Restricted Uniform prior, the incorrect $M = 3$ case will get substantive weight. Furthermore, this is only one DGP. If one were to use other randomly generated DGPs it would, of course, be possible to select the wrong model. With the Unrestricted Uniform prior we do not run these risks.

For the sake of brevity, we do not reproduce figures comparable to Figures 2 and 3 for DGP_2 . If we had, they would have shown that the Unrestricted Uniform prior with $M = 2$ would have yielded sensible posterior results for the AR(1) coefficient, capturing the abrupt regime switch in the DGP. Furthermore, the posterior for the second change-point allocates almost all of its probability out-of-sample. Overall, we are finding strong support for our story that the Unrestricted Uniform prior is an effective way of estimating the number of change-points in-sample (as opposed to imposing it on the model) and avoiding the “pile-up of prior probability at the end-of-sample” problem associated with the Restricted Uniform prior.

| | Log Marginal Likelihood | Predictive Mean | Predictive St. Deviation |
|--------------------------|-------------------------|-----------------|--------------------------|
| AR(1) | -153.76 | 0.17 | 1.09 |
| Restricted ($M = 2$) | -149.18 | 0.46 | 1.01 |
| Restricted ($M = 3$) | -150.21 | 0.29 | 1.27 |
| Unrestricted ($M = 2$) | -149.88 | 0.46 | 1.01 |
| Unrestricted ($M = 3$) | -150.06 | 0.45 | 1.04 |

2.2.2 Empirical Illustration Using Real GDP Growth Data

There are many papers which investigate structural breaks in real GDP growth. Of particular interest is the volatility of US real activity and the question of whether it has decreased over time. In this section we shall investigate this issue using U.S. real GDP growth from 1947Q2 through 2006Q4. We use an AR(2) in each regime and the priors for the coefficients are described at the beginning of this section.

Given the findings of Kim, Nelson and Piger (2004) and a wish not to restrict the maximum duration of each regime in-sample, we select $M = 3$ and $c = 1$. This allows for up to three regimes in-sample, and since the maximum regime duration is T it also allows for no change-points to occur in-sample. Given that this paper relates to change-point modeling and the related empirical literature relates to changes in volatility, Figure 4 presents the posteriors for τ_1, τ_2 as well as the posterior mean of the volatility (i.e. the standard deviation of the error) at each point in time.

An examination of Figure 4 show that there does seem to be strong evidence of a change-point in the early 1980s (with $E(\tau_1 | \mathbf{Y}_T) = 1983.2$ with posterior standard deviation of 3.8). This is reflected in the plot of the volatility which shows the Great Moderation decrease found by others in the early 1980s. Most evidence indicates that the second change-point occurs out-of sample, showing how, with the Unrestricted Uniform prior, $M - 1$ can play the role of a maximum number of change-points with the actual number occurring in-sample being estimated. In particular, the posterior for τ_2 implies that there is a 96.5% probability that the second change-point occurs out-of-sample (and, thus, that there is one change-point in-sample). The remaining 3.5% of the posterior probability indicates that there is a second change-point in volatility in the early- to mid-1980s.

It is also worth noting that the fact that our change-points are treated as random parameters means that changes in coefficients over time do not have to be abrupt. For instance, if we set $\tau_1 = 1983$ (with τ_2 being out-of-sample), then the graph of volatility would have been a step function. However, in Figure 4, the decline in volatility is much more gradual than a step function. We are finding appreciable probability (e.g. one percent or more) of a change-point in all quarters between 1982 and 1985. Thus, the stochastic treatment of change-points can be quite useful when modeling gradual evolution of coefficients in that even a small number of change-points can allow for quite flexible behavior.

A good benchmark to compare our Unrestricted Uniform prior (with $M = 3$) results is the Restricted Uniform prior with $M = 3$. Figure 5 presents results for this prior in the same format as Figure 4. If nothing else, a comparison of Figures 4 and 5 makes the point that priors matter. At first pass, the Restricted and Unrestricted Uniform priors might seem like very similar “noninformative” priors. Clearly they are yielding quite different posteriors. In contrast to the Unrestricted Uniform prior, the Restricted Uniform prior is yielding much evidence that the first change-point is occurring at the beginning of the sample and it is the second change-point that is occurring in the early 1980s (although there is also some evidence that τ_2 is occurring at the end of the sample). This difference in posterior inference about change-points between the two priors can, of course, have consequences for posterior inference about other parameters or for prediction. In terms of volatility findings, although the main “Great Moderation in the early 1980s” pattern is found using both Unrestricted and Restricted Uniform priors, non-negligible differences between these two priors are found at both the beginning and the end of the sample. Furthermore, Table 3 presents predictive means and standard deviations for y_{T+1} obtained using the two priors. Although the predictive means are quite similar to one another, the predictive standard deviation is substantially larger using the Restricted Uniform prior. This is due to the small upturn in the stochastic volatility at the end of the sample found using this prior. This small upturn is due to the fact that the Restricted Uniform Prior is finding a small probability of a change-point occurring at the end of the sample. This finding, in turn, is due to the pile up of prior probability at the end of the sample which arises with the Restricted Uniform prior.

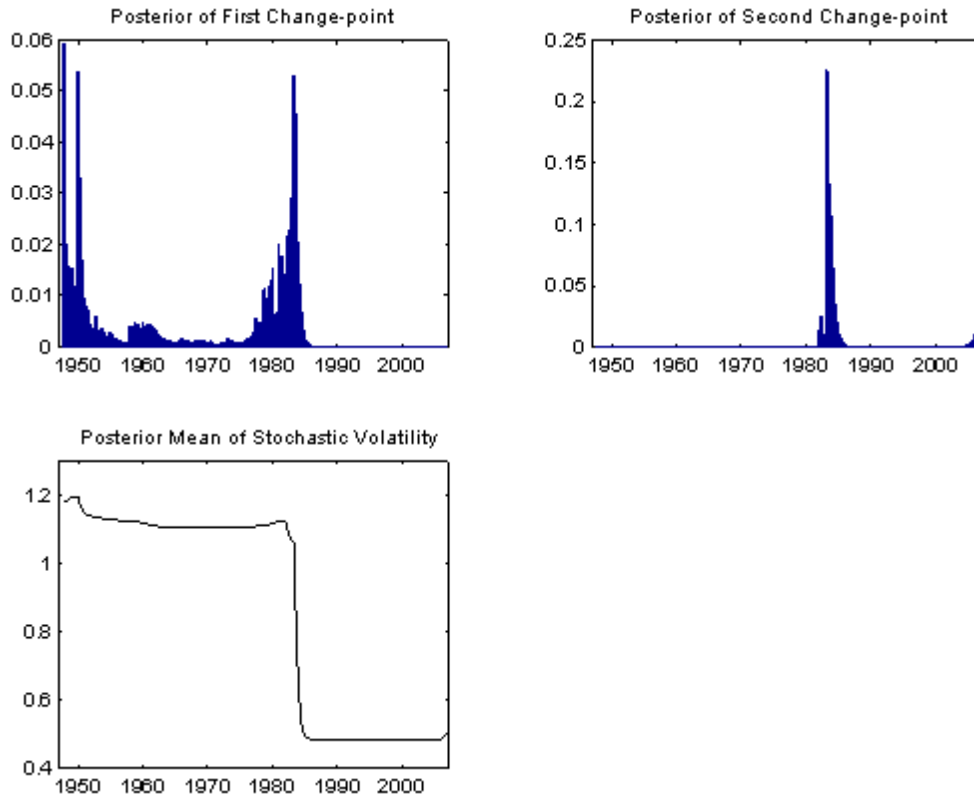


Figure 5: Results Using Restricted Uniform Prior with $M=3$

Table 3 presents marginal likelihoods and predictive results for a wider range of models. Note that, for $c = 1$, the Restricted and Unrestricted Uniform priors with $M = 2$ are the same. Using our approach, there seems strong evidence in favor of one change-point occurring in-sample. One message from this table is that, with the Unrestricted Uniform prior, we can correctly obtain this result (even if we set $M = 3$). Another message is that the Restricted Uniform prior can lead you astray (as evidenced, e.g., by an examination of predictive means and standard deviations) unless you get the the number of change-points exactly correct. Our preferred strategy is to select a maximum number of regimes, M , and then use the Unrestricted Uniform prior to tell us the number of change-points in-sample. This seems to work well with this data set. An alternative strategy would be to use the Restricted Uniform prior with several values for M and then select the value for M which yields the highest marginal likelihood (or do Bayesian model

averaging over different values for M). In this case, these two strategies yield roughly the same results (although Bayesian model averaging would attach almost five percent of the probability to the misleading $M = 3$ results). However, with model selection there is always the chance of selecting an inferior model. Furthermore, this example provides strong evidence for one change-point and, as we have seen, the “pile-up of prior probability at the end of the sample” problem occurs most prominently when more change-points occur.

| | Log Marginal Likelihood | Predictive Mean | Predictive St. Deviation |
|--------------------------|-------------------------|-----------------|--------------------------|
| AR(2) | -327.75 | 0.71 | 0.93 |
| Restricted ($M = 2$) | -302.60 | 0.61 | 0.49 |
| Restricted ($M = 3$) | -305.90 | 0.58 | 0.62 |
| Unrestricted ($M = 2$) | -302.60 | 0.61 | 0.49 |
| Unrestricted ($M = 3$) | -303.05 | 0.61 | 0.49 |

In the previous material, we have chosen particular values for c for the Unrestricted Uniform prior. We remind the reader that it is possible to treat c as an unknown parameter and either choose a particular value for it (e.g. choose the posterior mode of c) or integrate it out (see equation 2.11 and surrounding discussion). For this empirical application we do not present detailed results using either approach since they are similar to those presented above (provided c is big enough to allow for τ_2 to occur out-of-sample). Figure 6 plots the posterior of c (for $M = 3$).

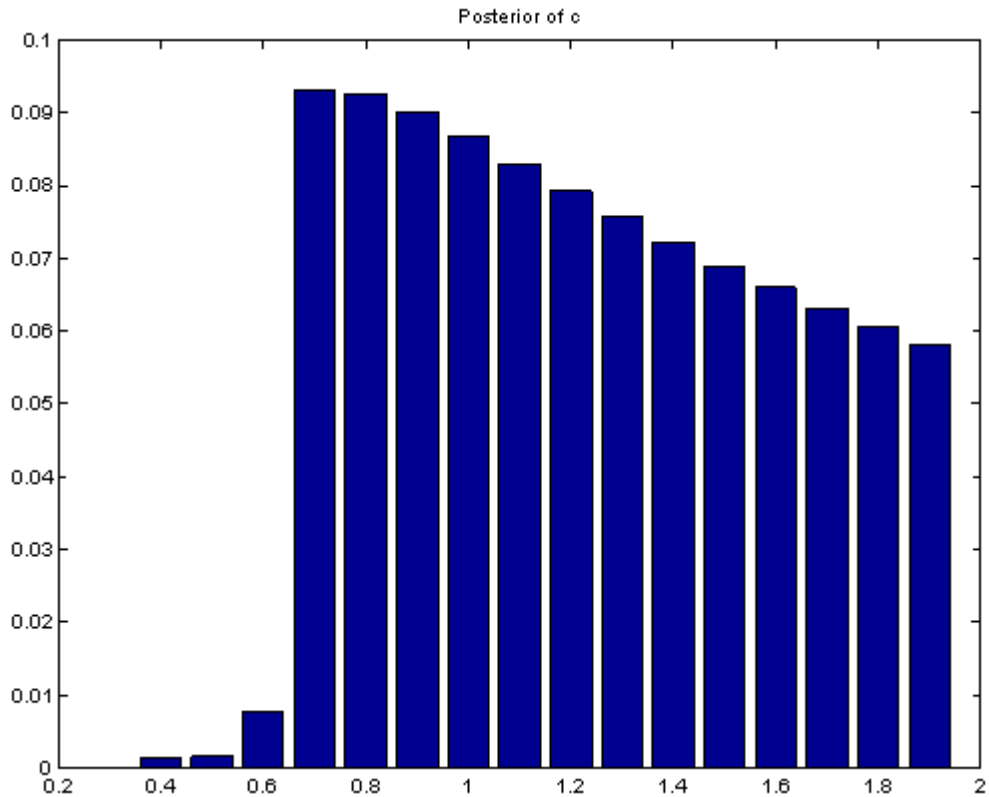


Figure 6: Posterior Results using Unrestricted Uniform Prior with $M=3$

3 Other Priors for Change-Points

The main purpose of this paper is to discuss priors for change-points which are noninformative, comparable to the methods used by classical econometricians. However, it is worth noting that several informative priors are popularly used with change-point models. In this section, we briefly describe some approaches and show how similar issues arise with them. In particular, if a particular number of change-points is imposed in-sample then one can end up with priors which are informative (particularly near the end of the sample) in an undesirable way. This can have a substantial impact on posterior inference in empirically-reasonable data configurations. It is preferable to treat the number of change-points as unknown and estimate it. In a simi-

lar manner as the Restricted Uniform prior has undesirable properties which can be eliminated if we use an Unrestricted Uniform prior, we discuss how other existing approaches can be extended.

Perhaps the most influential Bayesian change-point model in econometrics is developed in Chib (1998).⁴ 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 (2004)]. See also Pesaran, Pettenuzzo and Timmerman (2007) and Maheu and Gordon (2007) for recent extensions of relevance for 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.

To explain the approach of Chib (1998), we extend our earlier notation. Remember that we have data on a time series variable, y_t for $t = 1, \dots, T$ and let $\mathbf{Y}_i = (y_1, \dots, y_i)'$ denote the history through time i . Regime changes depend upon a discrete random variable, s_t , which takes on values $\{1, 2, \dots, M\}$. The likelihood function is defined by assuming $p(y_t | \mathbf{Y}_{t-1}, s_t = m) = p(y_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}_m)$ for a parameter vector $\boldsymbol{\theta}_m$ for $m = 1, \dots, M$. Thus, change-points occur at times τ_m defined as

$$\tau_m = \{t : s_{t+1} = m + 1, s_t = m\} \text{ for } m = 1, \dots, M - 1. \quad (3.1)$$

To avoid confusion, it is worth stressing that change-point models can be parameterized in different ways. With our Uniform priors, we parameterized directly in terms of the change-points (i.e. $\tau_1, \dots, \tau_{M-1}$). But one can also work in terms of states which denote each regime (i.e. s_t). It is also possible to write models in terms of durations of regimes. In the following material,

⁴Other key early Bayesian work in the statistics literature includes Carlin, Gelfand, and Smith (1992), Barry and Hartigan (1993) and Stephens (1994).

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 60th can be expressed as $\tau_1 = 60$, or $s_t = 1$ for $t = 1, \dots, 60$, and $s_t = 2$ for $t = 61, \dots, 100$, or $d_1 = 60$ and $d_2 = 40$ (where d_m denotes the duration of regime m).

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

$$\Pr(s_t = j | s_{t-1} = i) = \begin{cases} p_i & \text{if } j = i \neq M \\ 1 - p_i & \text{if } j = i + 1 \\ 1 & \text{if } i = M \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

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, (3.2) describes a hierarchical prior for the vector of states.⁵

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. This is what we have done with our Uniform priors in the previous section.⁶ If the number of change-points is even moderately large, then computational costs can become overwhelming. 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. As a digression, it is worth acknowledging that not all non-Bayesian approaches require searching over all possible sets of change-points. The influential approach of Bai and Perron (1998) is less computationally burdensome. Bai and Perron start from the observation that there are $T(T+1)/2$ ways of partitioning the sample. Bai and Perron then show how an efficient dynamic programming method can be used to find the global least squares minimizer in the special case of all parameters in a linear conditional

⁵A 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)].

⁶Although, as discussed in Section 2, it is possible to develop posterior simulation algorithms which do not require this.

mean changing at each change-point with no restrictions on the coefficients changes. In this special case they require only $O(T^2)$ computations to find the least squares minimizer. But for more general cases, computational problems can be serious with change-point models. Bayesian inference in the model of Chib (1998) is based on a Markov Chain Monte Carlo (MCMC) algorithm with data augmentation and we refer the reader to that paper for details.

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. Thus, this prior is not noninformative in the sense that the Uniform prior is. For many applications, Chib’s prior 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 discussion suggests we should at least investigate the consequences of this particular choice of hierarchical prior and consider possible alternatives.

Equation (3.2) 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 $\underline{\delta}_1, \underline{\delta}_2$.⁷ In this section, we will refer to the change-point model with hierarchical prior given by (3.2) with a 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 have a 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

⁷See, e.g., Poirier (1995), pages 104-105 for the definition and properties of the Beta distribution.

the regime durations for the Chib model is:

$$p(d_m) = \frac{B(\underline{\delta}_1 + d_m - 1, \underline{\delta}_2 + 1)}{B(\underline{\delta}_1, \underline{\delta}_2)}, d = 1, 2, \dots, \quad (3.3)$$

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_m are more likely than $d_m + 1$. Note that this property is present both both in the hierarchical prior, $p(d_m | p_m)$, and the marginal prior, $p(d_m)$.

The hierarchical prior in (3.2) can be combined with a likelihood function within each regime (e.g. equation 1.1) to produce a change-point model with M regimes. However, note that so far there is nothing in (3.2) alone which imposes that exactly M regimes occur in-sample. Indeed, if $p_i > 0$, (3.2) implies there is some probability that a change will never occur in finite time. 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 . Such a change-point model would be analogous to our Unrestricted Uniform prior where some of the change-points could occur out of sample (and indeed such a prior would be desirable for this very reason).

Nevertheless, the existing literature using the Chib model does impose M regimes in-sample. This can be done through restricting the prior. To see what form these restrictions take, note that (3.2) on its own implies

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

In words, this says “if at time $T - 1$, you are in regime $M - 1$, there is nothing which guarantees you will go to regime M next period”. But this can easily be restricted to:

$$\Pr[s_T = M | s_{T-1} = M] = \Pr[s_T = M | s_{T-1} = M - 1] = 1 \quad (3.4)$$

which says “if at time $T - 1$, you are in regime $M - 1$, then you must switch regimes to ensure that exactly M regimes occur in-sample”. Such a restriction is enough to ensure the model with $M = 2$ has exactly one change-point. For $M = 3$ we require the additional restriction:

$$\Pr[s_{T-1} = M - 1 | s_{T-2} = M - 2] = 1$$

and analogous additional restrictions are required for $M > 3$.

In the working paper version of this paper,⁸ we provide further details on both these priors (i.e. the unrestricted Chib model and the Chib model with prior restrictions imposed to ensure exactly $M - 1$ change-points occur in-sample). In particular, we show how the algorithm of Chib (1996) can be used even with the prior restrictions. Furthermore, we show how these prior restrictions can be undesirable in the sense of placing a great deal of prior weight in favor of change-points near the end of the sample. Since the reasons for this “pile up of prior probability at the end of the sample” are much the same as for the Restricted Uniform prior, we will not repeat this material here. They will be illustrated empirically below. But we do stress that the same issues as discussed in the previous section hold with the Chib model, and indeed for any model which imposes an exact number of regimes in-sample.

Before turning to an empirical illustration, it is worthwhile stressing that extending Chib’s model to allow for change-points to occur out-of-sample is trivial. One can simply not impose prior restrictions such as (3.4). One aspect of the model developed in Koop and Potter (2007) has a similar property.⁹ That is, in our previous work we use a Poisson distribution for the duration distribution. This allows for change-points to occur out-of-sample in a similar manner as our Unrestricted Uniform prior. The Poisson distribution, however, implies a prior distribution for the change-points which is far from flat and, hence, may not be suitable for the Bayesian wishing to have a “noninformative” prior over the change-points. McCulloch and Tsay (1993) and Giordani and Kohn (2006) are other approaches to change-point modeling which do not impose a fixed number of change-points in-sample. The basic idea of both these approaches is that there is some probability of switching regimes each period (e.g. McCulloch and Tsay, 1993, simply have a Bernoulli probability of change, p , which is the same in every period) and,

⁸Available at <http://personal.strath.ac.uk/gary.koop/>.

⁹Much of Koop and Potter (2007) relates to hierarchical prior elicitation relating to regime coefficients (e.g. to link coefficients in different regimes together in a sensible manner). Our previous paper also develops hierarchical priors for parameters of duration distributions (e.g. so that the duration of a past regime can provide some information about the likely duration of the current regime). Such considerations are not discussed in this paper, although they could easily be added to our Unrestricted Uniform prior.

thus, the number of switches which actually occur in-sample is unknown and estimated from the data. Like Koop and Potter (2007), these models do not have flat priors over the change-points.

3.1 Empirical Illustration Using Coal Mining Data

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 using the Chib model as well as the Restricted and Unrestricted Uniform priors.

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 (3.3) 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_m . 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, 2004) $\underline{\delta}_2$ has been set to a small value 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 3.3). Hence, such a choice of prior hyperparameters may get around the possibly unattractive property that the prior is informative for the change-points (e.g. it 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.

As we have seen, the model of Chib requires us to add restrictions analogous to (3.4) to ensure that number of change-points assumed in the model do in fact occur. In the case of $M = 2$, this can be done by truncating the distribution of the duration of regime 1 at $d_1 = T - 1$ 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 (3.3) to get

$$P[\tau_1 = 111] = P[d_1 = 111] = 1 - \sum_{d=1}^{110} \frac{B(7 + d, 1.1)}{B(8, 0.1)} = 0.76. \quad (3.5)$$

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 has a similar property. For this case, 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 3.5). In this case as well, there is a huge spike in the prior at the end 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 $G(\underline{\alpha}_m, \underline{\beta}_m)$ for $m = 1, \dots, M$. 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 $\underline{\alpha}_m = 2, \underline{\beta}_m = 1$. 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 (see Chib, 1998, or the appendix to the working paper version of this paper).

With regards to the Unrestricted Uniform prior we set $M = 3$ and $c = \frac{T-2}{T}$, values consistent with previous work with this data which has indicated one or two change-points. For the Chib model and Restricted Uniform prior we try $M = 2$ and $M = 3$ (as well as the model with no change-points). For the sake of brevity, we do not present figures for the priors, although it is worth stressing that the priors for change-points in the Chib model has the huge spikes near the end noted previously (see equation 3.5).

For the case where a single change-point is assumed to exist, the Chib model 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. The posteriors for τ_1 and τ_2 for the Chib model and Unrestricted Uniform prior are plotted in Figures 7¹⁰ and 8, respectively. For τ_1 , the posteriors under the two priors are not too different, although the informativeness of the prior used in the Chib model is clearly having an impact at the beginning of the sample. However, for τ_2 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 prior (see equations 3.4 or 3.5) which imposes exactly two change-points in-sample. We do not

¹⁰The coal mining data ends in 1962, but Figure 7 is truncated in 1900 since the probability that τ_1 occurs after 1900 is essentially zero.

present results for the Restricted Uniform Prior with $M = 3$, but note that the pile up of probability in the posterior for τ_2 near the end of the sample is much less appreciable than for the Chib model. Thus, we are finding prior sensitivity even when staying in the class of models which impose a precise number of change-points.

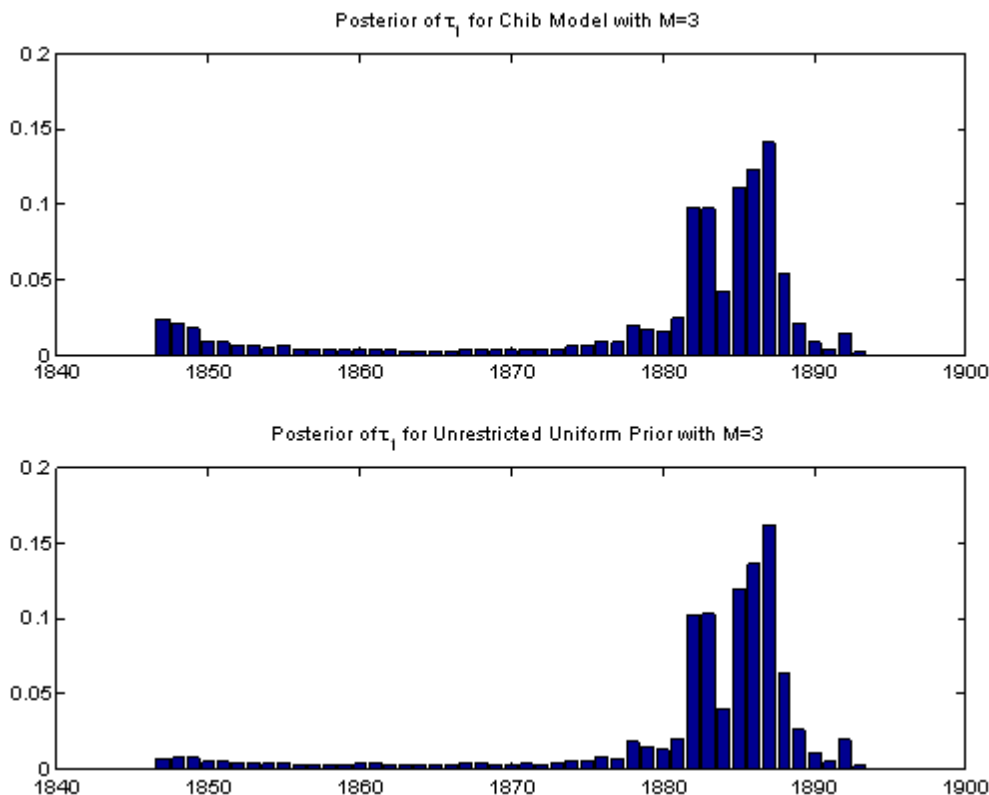


Figure 7: Posteriors of First Change-point using Coal Mining Disaster Data

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, Table 4 reports marginal likelihoods for the various models. Note that, as with the AR model used in the previous empirical illustrations, these can be calculated analytically since, conditional on the change-points, a closed form expression for the marginal likelihood exists (see the appendix to the working

paper version of this paper available at <http://personal.strath.ac.uk/gary.koop/>). These conditional marginal likelihoods can be averaged over the change-points using the appropriate prior to yield an exact (unconditional) marginal likelihood (see equation 2.10).

Table 4 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 Unrestricted Uniform prior) that is receiving more support from the data.

| | |
|--------------------------------|---------|
| No change-points | -206.21 |
| 1 change-point Chib model | -178.35 |
| Restricted Uniform ($M = 2$) | -176.76 |
| 2 change-point Chib model | -178.96 |
| Restricted Uniform ($M = 2$) | -177.35 |
| Unrestricted Uniform | -177.19 |

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 to the working paper version of this paper describes how Bayesian predictive inference can be done in the Chib model with Poisson likelihoods in each regime. For the Chib model and the Restricted Uniform prior, we do Bayesian model averaging across models with differing number of change-points. This can be done in a straightforward fashion by weighting the resulting predictive distribution by the posterior probabilities of the various change-point models. The latter can be directly calculated from the log marginal likelihoods in Table 4. 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. The different panels of Figure 9 are predictive distributions 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 the Chib

model places probability of more than $\frac{1}{3}$ on two or more disasters per year. However, our Uniform priors yield forecast distributions (see the bottom two panels of Figure 9) which indicate roughly $\frac{1}{5}$ 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 (3.4) 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 the top panel of Figure 9 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.

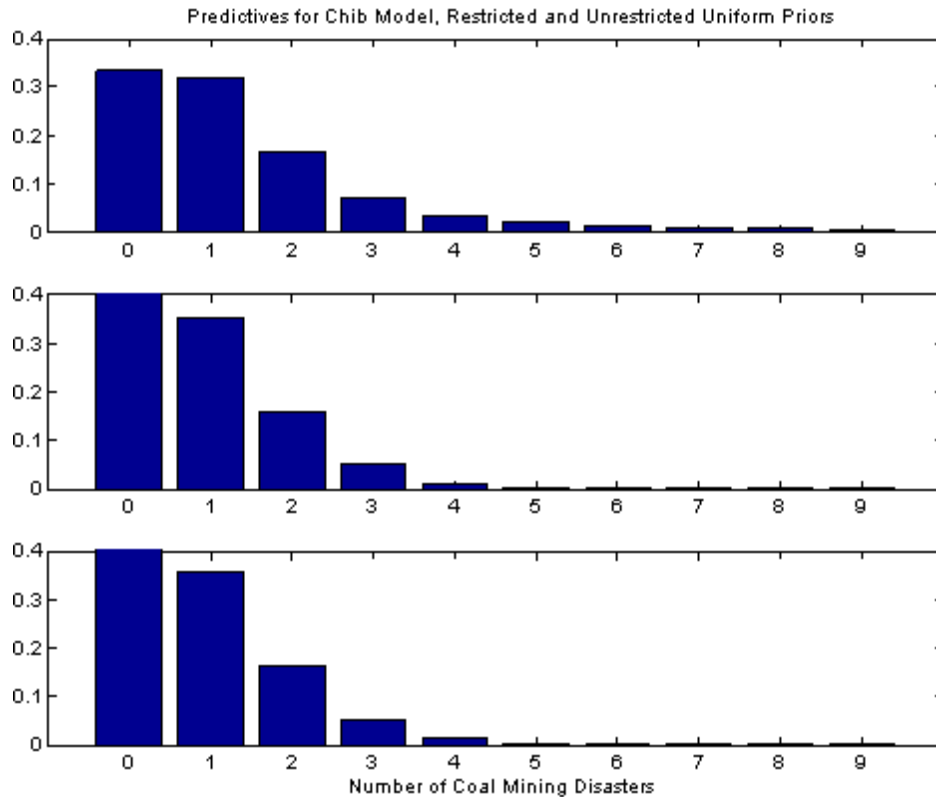


Figure 9: Predictive Distributions for Coal Mining Disaster Data

Finally, as with our GDP growth example, it is worth noting that, instead of selecting c for the Unrestricted Uniform prior, we can treat it as an unknown parameter and estimate it (or integrate it out). For forecasting, the use of such a 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, 2007). 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.¹¹

¹¹Hierarchical 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

In the coal mining disaster data, there is fairly clear evidence of one or two breaks, but not more. If we set $M = 3$ and $p(c|\underline{a}) = U\left(\frac{1}{2}, 2\right)$ we obtain predictive distributions which are quite similar to those given in the bottom two panels of Figure 9. 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-point models despite the fact that Table 4 indicates only weak evidence in favor of the presence of a second change-point. And it is worth stressing that, with our Unrestricted 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.

4 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. Of particular interest is our Unrestricted Uniform prior which can be thought of as a reasonable noninformative prior (what an objective Bayesian may wish to use or comparable to what the classical econometrician does). We present extensive empirical work which shows that the issue of prior elicitation can be of substantive importance in change-point models and that our Unrestricted Uniform prior yields results which are more sensible than other approaches.

5 References

Ang, A. and Bekaert, G., 2002, Regime switches in interest rates, *Journal of Business and Economic Statistics*, 20, 163-182.

(2007) for such an approach.

- 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., 2005, Drifts and volatilities: monetary policies and outcomes on Post WWII US, *Review of Economic Dynamics*, 8, 262-302.
- 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>.
- Giordani, P. and Kohn, R., 2006, Efficient Bayesian inference for multiple change-point and mixture innovation models, manuscript.
- 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., 2004, The less volatile U.S. economy: A Bayesian investigation of timing, breadth, and potential explanations, *Journal of Business and Economic Statistics*, 22, 80-93.
- Koop, G., 2003, *Bayesian Econometrics* (Wiley, Chichester).

- 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., 2003, Bayesian analysis of endogenous delay threshold models, *Journal of Business and Economic Statistics*, 21, 93-103
- Koop, G. and Potter, S., 2007, Estimation and forecasting in models with multiple breaks, *Review of Economic Studies*, forthcoming.
- Maheu, J. and Gordon, S., 2007, Learning, forecasting and structural breaks, manuscript available at <http://www.chass.utoronto.ca/~jmaheu/mg.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.
- McCulloch, R. and Tsay, E., 1993, Bayesian inference and prediction for mean and variance shifts in autoregressive time series, *Journal of the American Statistical Association*, 88, 968-978.
- 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., 2007, Forecasting time series subject to multiple structural breaks, *Review of Economic Studies*, forthcoming.
- Poirier, D.J., 1995, *Intermediate Statistics and Econometrics* (The MIT Press, Cambridge).
- Poirier, D.J., 1998, Revising beliefs in non-identified models, *Econometric Theory*, 14, 483-509.
- Stephens, D., 1994, Bayesian retrospective multiple-changepoint identification, *Applied Statistics*, 43, 159-178.
- Stock, J. and Watson, M., 1996, Evidence on structural instability in macroeconomic time series relations, *Journal of Business and Economic Statistics*, 14, 11-30.
- Stock J. and Watson M., 2002, *Has the business cycle changed and why?*, *NBER Macroeconomic Annual* (University of Chicago Press, Chicago).