Federal Reserve Bank of New York
Staff Reports

Forecasting and Estimating Multiple Change-Point Models with an Unknown Number of Change Points

Gary M. Koop
Simon M. Potter

Staff Report no. 196
December 2004

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in the paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

Forecasting and Estimating Multiple Change-Point Models with an Unknown Number of Change Points

Gary M. Koop and Simon M. Potter

Federal Reserve Bank of New York Staff Reports, no. 196
December 2004
JEL classification: C11, C22, E17

Abstract

This paper develops a new approach to change-point modeling that allows for an unknown number of change points in the observed sample. Our model assumes that regime durations have a Poisson distribution. The model approximately nests the two most common approaches: the time-varying parameter model with a change point every period and the change-point model with a small number of regimes. We focus on the construction of reasonable hierarchical priors both for regime durations and for the parameters that characterize each regime. A Markov Chain Monte Carlo posterior sampler is constructed to estimate a change-point model for conditional means and variances. We find that our techniques work well in an empirical exercise involving U.S. inflation and GDP growth. Empirical results suggest that the number of change points is larger than previously estimated in these series and the implied model is similar to a time-varying parameter model with stochastic volatility.

Key words: Bayesian, structural breaks, Markov Chain Monte Carlo, hierarchical prior

---

Potter: Research and Statistics Group, Federal Reserve Bank of New York (e-mail: simon.potter@ny.frb.org). Koop: Department of Economics, University of Leicester, Leicester, England (e-mail: gary.koop@leicester.ac.uk). The authors thank Edward Leamer, Richard Paap, Hashem Pesaran, Herman van Dijk, and seminar participants at Erasmus University, the University of Amsterdam, and the University of York for helpful comments. 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

Many recent papers have highlighted the fact that structural instability seems to be present in a wide variety of macroeconomic and financial time series [e.g. Ang and Bekaert (2002) and Stock and Watson (1996)]. The negative consequences of ignoring this instability for inference and forecasting has been stressed by, among many others, Clements and Hendry (1998, 1999), Koop and Potter (2001) and Pesaran, Pettenuzzo and Timmerman (2004). This has inspired a wide range of change-point models. There are two main approaches: one can estimate a model with a small number of change-points (usually one or two). Alternatively, one can estimate a time varying parameter (TVP) model where the parameters are allowed to change with each new observation, usually according to a random walk. A TVP model can be interpreted as having $T-1$ breaks in a sample of size T. Recent influential empirical work includes McConnell and Perez (2000) who use a single change-point model to present evidence that the volatility of US economic activity abruptly fell in early 1984. In a TVP framework, Cogley and Sargent (2001) model inflation dynamics in the US as continuously evolving over time. In this paper we seek to blend these two approaches in a model that approximately nests them.

To motivate why such a blend might be useful consider the case of a small number of breaks and, in particular, the variance break in US economic activity found by many authors around 1984 [see Stock and Watson (2002) for a review]. In addition to the obvious question of why the volatility of activity declined, is a more immediate one: whether the decline be sustained or is only a temporary phenomenon. Models with a small number of structural breaks typically do not restrict the magnitude of change in the coefficients that can happen after a break, but implicitly assume that after the last break estimated in the sample there will be no more breaks. In contrast, in the TVP model there is probability 1 of a break in the next new observation. However, for the TVP model the size of the break is severely limited by the assumption that coefficients evolve according to a random walk. In the model we develop, a new break can be forecast after the end of the sample and size of the break is partly dependent on the properties of the previous regime, partly dependent on the history of all previous breaks and partly has a random element.

Bayesian methods are attractive for change-point models since they can allow for flexible relationships between parameters in various regimes and are

computationally simple. That is, if we have a model with M different regimes, then hierarchical priors can be used to allow information about coefficients in the j^{th} regime (or the duration of the j^{th} regime) to depend on information in the other regimes. Such an approach can improve estimation of coefficients. It is particularly useful for forecasting in the presence of structural breaks since it allows for the possibility of out-of-sample breaks. With regards to computation, use of a hierarchical prior allows the researcher to structure the model so that, conditional on unknown parameters (e.g. the change-points) or a vector of latent data (e.g. a state vector denoting the regimes), it is very simple (e.g. a series of Normal linear regression models). Efficient Markov Chain Monte Carlo (MCMC) algorithms which exploit this structure can be developed. This allows for the estimation of models, using modern Bayesian methods, with multiple change-points that appear infeasible under the standard classical approach to change-point problems.[1]

However, with some partial exceptions [e.g. Pesaran, Pettenuzzo and Timmerman (2004) and Stambaugh and Pastor (2001)], we would argue that the existing Bayesian literature in economics has not fully exploited the benefits of using hierarchical priors. In addition, this literature has, following the existing frequentist literature, focussed on either models with a small number of breaks or TVP models. Furthermore, as argued in Koop and Potter (2004), some commonly-used Bayesian priors have undesirable properties. These considerations motivate the present paper where we develop a model which draws on our beliefs that desirable features for a change-point model are:

1. The number of regimes and their maximum duration should **not** be restricted ex-ante.

2. The regime duration distribution should **not** be restricted to be constant or monotonically decreasing/increasing.

3. The parameters describing the distribution of the parameters in each regime should, if possible, have conditionally conjugate prior distributions to minimize the computational complexity of change-point models.

[1] Many tests require evaluation of something (e.g. a likelihood) at every possible breakpoint. If there are $M-1$ possible breakpoints, then $O\left(T^{M-1}\right)$ evaluations are required. Even for moderate values of M this can become computationally infeasible.

4. Durations of previous regimes can potentially provide some information about durations of future regimes.

5. The parameters characterizing a new regime can potentially depend on the parameters of the old regime.

The plan of this paper is as follows. In Section 2 we review the link between change-points and hidden Markov chains. In Section 3 we develop our new model of regime duration. In Section 4 we construct a method for modeling the change in regime coefficients based on a similar heirarchical structure to the TVP model. Section 5 gives an overview of the posterior simulator used in our Bayesian analysis (a technical appendix contains more details). Section 6 contains applications to US GDP growth and inflation as measured by the PCE deflator. We compare the results of our approach with that of a single structural break and a TVP model and find them to be much closer to the latter. In general, we find our methods to reliably recover key data features without making the potentially restrictive assumptions underlying other popular models.

2 Change-Point Models and Hidden Markov Chains

In order to discuss the advantages of our model, it is worthwhile to begin by describing in detail some recent work and, in particular, the innovative model of Chib (1998) which has been used in many applications [e.g. Pastor and Stambaugh (2001), Kim, Nelson and Piger (2002) and Pesaran, Pettenuzzo and Timmerman (2004)]. In terms of computation, 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 have data on a scalar time series variable, y_t for $t = 1, \ldots, T$ and let $Y_i = (y_1, \ldots, y_i)'$ denote the history through time i and denote the future by $Y^{i+1} = (y_{i+1}, \ldots, y_T)'$. 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 $S^{i+1} = (s_{i+1}, \ldots, s_T)'$. The likelihood function is defined by assuming $p(y_t|Y_{t-1}, s_t = m) = p(y_t|Y_{t-1}, \theta_m)$ for a parameter vector θ_m for $m =$

$1, \ldots, M \leq T$. 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 i to regime $i+2$. 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.[2]

To avoid confusion, we stress 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. 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 $S^{61} = 2$, 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 break points. If the number of break 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

[2] 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)].

estimate models with many change-points. Appendix A describes this algorithm (which we use, with modifications, as a component of the posterior simulator for our model).

Chib chose to model the transition probabilities of the states as being constants. One consequence of this is that regime duration satisfies a Geometric distribution, a possibly restrictive choice. For instance, the Geometric distribution is decreasing, implying that $p(d_m) > p(d_{m+1})$ which (in some applications) may be unreasonable. In the model we introduce below, we generalize this restriction by allowing regime duration to follow a more flexible Poisson distribution.

Furthermore, the model of Chib (1998) assumes that exactly M regimes exist in the data. In Koop and Poirier (2004), we show how this implicitly imposes on the prior a very restrictive form which will tend to put excessive weight near the end of the sample. That is, the standard hidden Markov model (i.e. without restrictions such as those given in equation 2.2) will use probabilities

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

To impose that exactly M regimes occur, this has to be changed to the equal probabilities:

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

In our previous work, we explored the consequences of such restrictions and argued that they can have a substantial impact on posterior inference in practice. We further argued that other sensible priors which impose exactly M regimes will also run into similar problems. Partly for this reason, we argued that it is important to develop a hierarchical prior which treats the number of regimes as unknown.

In summary, the pioneering work of Chib (1998) has changed the way many look at change-point models and has had great influence. In terms of posterior computation, Chib (1998) continues to be very attractive and, indeed, we use a modification of this algorithm as part of our posterior simulator. However, the hierarchical prior has some potentially undesirable properties which leads us to want to build on Chib (1998).

3 A Poisson Hierarchical Prior for Durations

The above discussion illustrates some restrictive properties of traditional hierarchical priors used in the change-point literature and leads to our contention that it is desirable to have a model for durations which satisfies the five criteria listed in the introduction. In this section we develop our alternative approach based on a Poisson model for durations.[3] This approach does not restrict the number or maximum durations of regimes ex-ante, it has a convenient conjugate prior distribution in the Gamma family and the regime duration distribution is not restricted to be constant or monotonically decreasing/increasing. It also allows information about the duration of past regimes to affect the duration of the current regime and potentially the magnitude of the parameter change from old to new regime.

We use a hierarchical prior for the regime durations which is a Poisson distribution. That is, $p(d_m|\lambda_m)$ is given by:

$$d_m - 1 = \tau_m - (\tau_{m-1} + 1) \sim Po(\lambda_m) \tag{3.1}$$

where $Po(\lambda_m)$ denotes the Poisson distribution with mean λ_m. With this hierarchical prior it makes sense to use a (conditionally conjugate) Gamma prior on λ_m. If we do this, it can be verified that $p(d_m)$, the marginal prior for the duration between change points, is given by a Negative Binomial distribution.

To provide some intuition, remember that the assumption comparable to (3.1) in the model of Chib (1998) was that the duration had a hierarchical prior which was Geometric (apart from the end-points). Chib (1998) used a Beta prior on the parameters. This hierarchical prior [and, as shown in Koop and Potter (2004), the marginal prior $p(d_m)$] implies a declining probability on regime duration so that higher weight is placed on shorter durations. In contrast, the Poisson form we use for $p(d_m|\lambda_m)$ and the implied Negative Binomial form for $p(d_m)$ which we work with have no such restriction.

However, the prior given in (3.1) also has the unconventional property that it allocates prior weight to change-points outside the observed sample. That is, there is nothing in (3.1) which even restricts $d_1 < T$ much less $d_m < T$ for $m > 1$. We will argue that this is a highly desirable property

[3] Of course, there are many other popular options for modeling durations other than the Poisson. Bracqemond and Gaudoin (2003) offers a good categorization of different possibilities and explains their properties.

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, suppose we allow for $m = 1, ..., M$ regimes. Then, since some or all of the regimes can terminate out-of-sample, our model implicitly contains models with no breaks, one break, up to $M - 1$ breaks (in-sample).[4] The desirable properties of this feature are explored in more detail in Koop and Potter (2004).[5]

Although our model is much more flexible than that used in Chib (1998), computation is complicated by the fact that the matrix of transition probabilities now depends on the time spent in each regime. To see why this complicates computation, note that a key step in the Chib (1996) algorithm (see Appendix A) requires calculating $p(s_{t+1}|s_t, P)$ where P is the matrix of transition probabilities given by (2.2). In the model of Chib (1998) this density is simple due to the constant transition probability assumption. As discussed in Koop and Potter (2004) and developed in more detail in Appendix A, Chib's algorithm can still be applied in the case of a non-time homogenous transition matrix.

To better understand this point, note that under the Poisson hierarchical prior in (3.1) we can construct a finite element Markov transition matrix for any observed sample, under the assumption that regime 1 started some maximum $\chi - 1 < \infty$ periods before our initial observation. Consider the first observation in the sample. This is assumed to be generated by regime 1 but we do not know whether this is the first period of regime 1 or any other period up to χ. Thus, there are χ possible values for d_1. The probability of a transition from regime 1 to regime 2 is:

$$\Pr[s_2 = 2|s_1 = 1, d_1] = \frac{\exp(-\lambda_1)\lambda_1^{d_1-1}}{(d_1-1)!\left(1 - \sum_{j=0}^{d_1-2} \frac{\exp(-\lambda_1)\lambda_1^j}{j!}\right)}, d_1 = 1, \ldots, \chi, \tag{3.2}$$

where $\sum_{s=0}^{-1} \frac{\exp(-\lambda_1)\lambda_1^s}{s!}$ is defined to be 0. Thus, instead of the single p_{21} in the Chib's model we already have χ calculations to make. In the third period,

[4] The specification of a maximum number of regimes, M, is made only for illustrative purposes. In practice, our model does not require specification of such a maximum. In our empirical work we set $M = T$ which allows for a TVP model and an out-of-sample break.

[5] In a different but related context (i.e. a Markov switching model), Chopin and Pelgrin (2004), adopt a different approach to the joint estimation of number of regimes and the parameters.

depending on whether a regime switch occurred in period 2, we would have to calculate $\Pr[s_3 = 2|s_2 = 1, d_1]$ for $d_1 = 1, \ldots, \chi+1$ and $\Pr[s_3 = 3|s_2 = 2, d_2 = 1]$. In the following period, we would have to calculate $\Pr[s_4 = 2|s_3 = 1, d_1]$ for $d_1 = 1, \ldots, \chi+2$, $\Pr[s_4 = 3|s_3 = 2, d_2]$ for $d_2 = 1$ or 2 and $\Pr[s_4 = 4|s_3 = 3, d_3 = 1]$, and so on.

In general, it can be confirmed that

$$\Pr[s_{t+1} = m+1|s_t = m, d_m] = \frac{\exp(-\lambda_m)\lambda_m^{d_m-1}}{(d_m-1)!\left(1-\sum_{j=0}^{d_m-2}\frac{\exp(-\lambda_m)\lambda_m^j}{j!}\right)}, d_m = 1, \ldots, T-m. \tag{3.3}$$

which must be evaluated for all $T-m$ possible value of d_m and every possible value of m from 1 through t. Thus, unlike with Chib's model (which involves $O(TM)$ such calculations), calculating the transition probability matrix will involve $O(T^3)$ calculations.

Another important issue arises which does not arise in models with a known number of change-points. To motivate this issue, suppose that a true data generating process with one change-point exists and data is observed for $t = 1, ..., T$. Assuming that T is large enough for precise estimation of the true DGP, the posterior simulator will yield most draws which imply two regimes within the observed sample (i.e. most draws will have $s_t = 1$ or 2 for $t = 1, ..., T$) and $s_t = m$ for $m > 2$ will mostly occur for $t > T$. In this case, most of the regimes occur out-of-sample and there will be no data information available to estimate their durations. So, if two regimes exist, there will be a great deal of information to estimate λ_1 and λ_2 but apparently none to estimate λ_m for $m > 2$. In a Bayesian analysis we do not necessarily have to worry about this. It is well known that if no data information relating to a parameter exists, then its posterior is equal to its prior if the prior exhibits independence. Thus, if an independent prior is used such that $p(\lambda_1, \ldots, \lambda_T) = p(\lambda_1) \cdots p(\lambda_T)$ with

$$\lambda_m \sim G\left(\underline{\alpha}_\lambda, \underline{\beta}_\lambda\right), \tag{3.4}$$

then posterior for λ_m in many of the regimes will simply be $G\left(\underline{\alpha}_\lambda, \underline{\beta}_\lambda\right)$.[6]

In theory, there is nothing wrong with using an independent prior such as (3.4), and simplified versions of the methods described below can be used

[6] To simplify notation, we are assuming the $\lambda_m s$ to have the same prior. It is trivial to relax this assumption.

for this case. Out-of-sample regimes will have durations which simply reflect the prior, but this is not important insofar as one is interested in in-sample results (e.g. estimating the number and timing of change-points in-sample). However, if one is interested in forecasting, then out-of-sample properties matter. In many applications, it is reasonable to suppose that the duration of past regimes can shed some light on the duration of future regimes. In order to accommodate such a structure, we modify (3.4) and use a hierarchical prior of the form:

$$\lambda_m \left| \beta_\lambda \sim G\left(\underline{\alpha}_\lambda, \beta_\lambda\right), \right. \tag{3.5}$$

where β_λ is an unknown parameter (not a hyperparameter selected by the researcher).[7]

This new parameter, which reflects the degree of similarity of the durations of different regimes, requires its own prior and it is convenient to have:

$$\beta_\lambda^{-1} \sim G(\underline{\xi}_1, 1/\underline{\xi}_2). \tag{3.6}$$

To aid in prior elicitation, note that this configuration implies the prior mean of d_m (after integrating out λ_m) is

$$1 + \underline{\alpha}_\lambda \left(\frac{\underline{\xi}_2}{\underline{\xi}_1 - 1} \right),$$

if $\underline{\xi}_1 > 1$.

It is important to understand the implications of any prior (Appendix C discusses such properties by simulating from the particular prior used in our empirical work). As discussed following (3.1), in the model we propose the hierarchical prior where $p\left(d_m|\lambda_m\right)$ is Poisson, but if we integrate out λ_m, we get $p\left(d_m|\beta_\lambda\right)$ being a Negative Binomial distribution. The unconditional prior distribution, $p\left(d_m\right)$ is found by integrating out β_λ^{-1}. This does not have a closed form (and we do not reproduce it here for reasons of brevity). In general $p\left(d_m\right)$ inherits the flexible form of $p\left(d_m|\lambda_m\right)$ or $p\left(d_m|\beta_\lambda\right)$. However, it is worth mentioning that if $\underline{\alpha}_\lambda = 1$ then we have the restrictive property

[7] We could also treat $\underline{\alpha}_\lambda$ as an unknown parameter. However, we do not do so since our model already has a larger number of parameters and the additional flexibility allowed would not be great. Choosing $\underline{\alpha}_\lambda = 1$ implies λ_m is drawn from the exponential distribution (with mean estimated from the data). Other integral choices for $\underline{\alpha}_\lambda$ imply various members of the class of Erlang distributions.

that $P(d_m = y) > P(d_m = y+1)$. This suggests that, for most applications, it is desirable to avoid such small values for $\underline{\alpha}_\lambda$. It can also be shown that, for very small values of $\underline{\xi}_2$, with $\underline{\alpha}_\lambda = n$ we have a high prior probability of a regime change every n periods. Such considerations can be useful in prior elicitation.

As noted above, unlike in the model of Chib (1998), the transition probabilities in (3.2) and (3.3) depend upon the duration spent in each regime. However, in both Chib (1998) and our model, the durations do not enter the likelihood (i.e. $p\left(y_t|Y_{t-1}, s_t = m\right) = p\left(y_t|Y_{t-1}, \theta_m\right)$ does not depend on the duration of the regime).

In summary, in this section we have developed a hierarchical prior for the regime durations which has two levels to the hierarchy. At the first level, we assume the durations to have Poisson distributions. At the second level, we assume the Poisson intensities (i.e. $\lambda_m s$) are drawn from a common distribution. Thus, out-of-sample $\lambda_m s$ (and, thus, regime durations) are drawn from this common distribution (which is estimated using in-sample data). This is important for forecasting as it allows for the prediction to reflect the possibility that a change-point occurs during the period being forecast.

4 Development of the Prior for the Parameters in Each Regime

In the same way that the change-point framework of Chib (1998) can be used with a wide variety of likelihoods (i.e. $p\left(y_t|Y_{t-1}, s_t = m\right)$ can have many forms , our Poisson model for durations can be used with any specification for $p\left(y_t|Y_{t-1}, s_t = m\right) = p\left(y_t|Y_{t-1}, \theta_m\right)$. Here we choose a particular structure based on a regression or autoregressive model with stochastic volatility which is of empirical relevance.

We adopt a state space framework where the observable time series satisfies the measurement equation

$$y_t = X_t \phi_{s_t} + \exp(\sigma_{s_t}/2)\varepsilon_t, \tag{4.1}$$

where $\varepsilon_t \sim N(0,1)$ and the $(K+1)$ state vector $\theta_{s_t} = \{\phi_{s_t}, \sigma_{s_t}\}$ satisfies the

state transition equations

$$\begin{aligned} \phi_m &= \phi_{m-1} + U_m, \\ \sigma_m &= \sigma_{m-1} + u_m, \end{aligned} \tag{4.2}$$

where $U_m \sim N(0, V)$, $u_m \sim N(0, \eta)$ and and X_t is a K-dimensional row vector containing lagged dependent or other explanatory variables. The initial conditions, ϕ_0 and σ_0 can be treated in the same way as in any state space algorithm.[8]

Note that this framework differs from a standard state space model used in TVP formulations in that the subscripts on the parameters of the measurement equation do not have t subscripts, but rather s_t subscripts so that parameters change only when states change. This difference leads to the state equations having m subscripts to denote the $m = 1, ..., M$ regimes.

To draw out the contrasts with models with a small number of breaks, note that the hierarchical prior in (4.2) assumes that θ_m depends on θ_{m-1}. In contrast, in most traditional models with a small number of breaks, one assumes θ_m and θ_{m-1} are independent of one another.[9] Furthermore, it is usually assumed that the priors come from a conjugate family. For instance, a traditional model might begin with (4.1) and then let θ_m have the same Normal-Gamma natural conjugate prior for all m. This approach, involving unconditionally independent priors, is not reasonable in our model for reasons we have partially discussed above. That is, our approach involves an unknown number of change-points in the observed sample. So it is possible that many of the regimes occur out-of-sample. In traditional formulations, there is no data information to estimate θ_m if the m^{th} regime occurs out-of-sample. The hierarchical prior of (4.2) alleviates this problem. An alternative approach to this issue is given in Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo and Timmerman (2004). They place a hierarchy on the parameters of the conjugate family for each regime. This is a standard approach in the Bayesian literature for cross-section data drawn from different groups. In a time series application it has less merit since one wants the most recent regimes to have the strongest influence on the new regime. This is a feature that our prior incorporates.

[8] In particular, in our state space algorithm the forward filter step is initialized with a diffuse prior.

[9] An exception, Pesaran, Pettenuzzo and Timmerman (2004) assumes that all the $\theta_m s$ are drawn from a common distribution.

The state equations in (4.2) define a hierarchical prior which links θ_m and θ_{m-1} in a sensible way. This martingale structure is standard in the TVP literature and, as we discuss later, it is computationally simple since it allows the use of standard Kalman filter and smoother techniques to draw the parameters in each regime. We use a standard (conditionally conjugate) prior for the innovation variances:

$$\begin{aligned} V^{-1} &\sim W\left(\underline{V}_V^{-1}, \underline{\nu}_V\right) \\ \eta^{-1} &\sim G(\underline{\alpha}_\eta, \underline{\beta}_\eta) \end{aligned} \tag{4.3}$$

where $W(A, a)$ denotes the Wishart distribution[10] and we assume that $\underline{\nu}_V > K + 1$.

Many extensions of this basic model for the link between regimes can be handled in a straightforward fashion by adding extra layers to the hierarchical structure. The innovation variance in the state equations can be allowed to be different (i.e. η and V can be replaced by η_m and V_m and a hierarchical prior used for these new parameters). Furthermore, in some applications, it might be desirable for the duration in a regime to effect θ_m (e.g. if regime $m-1$ is of very short duration, it is plausible that θ_{m-1} and θ_m are more similar to one another than if it was of long duration). Such considerations can be incorporated in a hierarchical prior for θ_m. For instance, in an earlier version of this paper, we had a prior which incorporates both such features as:

$$\begin{aligned} V_m^{-1} &\sim W\left(\frac{\left[\lambda_{m-1}\mathcal{V}_V\right]^{-1}}{\underline{\nu}_V - K - 1}, \underline{\nu}_V\right) \\ \eta^{-1} &\sim G(\underline{\nu}_\eta, \frac{\left[\lambda_{m-1}\mathcal{V}_\eta\right]^{-1}}{\underline{\nu}_\eta - 1}) \end{aligned}$$

where $\mathcal{V}_V$ and $\mathcal{V}_\eta$ are parameters to be estimated. In our applications to macroeconomic time series extension this did not outperform the simpler version. Nevertheless, in some applications such an extension might be warranted and it is worthwhile mentioning that the requisite methods can be es-

[10] We parameterize the Wishart distribution so that if $Z \sim W(A, a)$ and ij subscripts denote elements of matrices, then $E(Z_{ij}) = aA_{ij}$. The scalar a is a degrees of freedom parameter.

timated using straightforward extensions of the MCMC algorithm described in the next section.

Note also that (4.1) and (4.2) assume that regime changes occur at the same time for the coefficients and the volatility. Having separate regime structures for these is conceptually straightforward but practically complicated. In some cases, the researcher may want to simplify our model by having change-points only in some of the parameters. For instance, in an autoregressive model for GDP growth it might be plausible that the AR coefficients are constant and only the volatility is changing over time. We adopt such a specification for GDP growth in our empirical section.

To summarize, the prior we have developed has the form:

$$\begin{array}{l} p\left(\theta_1,..,\theta_M,\lambda_1,..,\lambda_M,V,\eta,\beta_\lambda,\chi\right) \\ = \left\{\prod_{m=1}^{M} p\left(\theta_m|\theta_{m-1},V,\eta\right) p\left(\lambda_m|\beta_\lambda\right)\right\} p(\theta_0) p\left(\beta_\lambda\right) p\left(\chi\right) p\left(V,\eta\right) \end{array} \tag{4.4}$$

where $p\left(\theta_m|\theta_{m-1},V,\eta\right)$ is given by (4.2), $p(\theta_0)$ is diffuse, (V,η) is given by (4.3) is given by (4.4), $p\left(\lambda_m|\beta_\lambda\right)$ is given by (3.5) and $p\left(\beta_\lambda\right)$ is given by (3.6). To simplify our computation, we set $\chi = 0$, and thus we ignore this parameter in the following discussion.[11]

5 Posterior and Predictive Simulation

In this section we outline the general form of the MCMC algorithm used to estimate the model. Precise details are given in Appendices A and B. To simplify notation we define $\Theta = \left(\theta_1',...,\theta_M'\right)'$ and $\lambda = \left(\lambda_1',...,\lambda_M'\right)'$. Note that our MCMC algorithm draws a sequence of states (S_T) that includes the values for the regime durations, d_m. Furthermore, we will set $M = T$ so that we can nest a standard TVP model. However, it is possible to set $M < T$ if one wishes to restrict the number of feasible regimes.

Our MCMC proceeds by sequentially drawing from the full posterior conditionals for the parameters $S_T, \theta, \lambda, V, \eta, \beta_\lambda$. The posterior conditional $p\left(S_T|Y_T,\Theta,\lambda,V,\eta,\beta_\lambda\right) = p\left(S_T|Y_T,\Theta,\lambda\right)$ can be drawn using the modified algorithm of Chib (1996) described in Appendix A with transition probabilities given by (3.2) and (3.3). $p\left(\Theta|Y_T,S_T,\lambda,V,\eta,\beta_\lambda\right)$ can be simulated from using extensions of methods of posterior simulation for state space models

[11] Appendix A describes how one can carry out Bayesian inference for χ.

with stochastic volatility drawing on Kim, Shephard and Chib (1998) and Carter and Kohn (1994). That is, the TVP model is a standard state space model and thus, standard state space simulation methods can be used directly. However, when regimes last more than one period, the simulator has to be altered in a straightforward manner to account for this.

The (conditional) conjugacy of our prior implies that, with one exception, $p(\lambda_m|Y_T, S_T, \Theta, V, \eta, \beta_\lambda)$ for $m = 1, ..., M$ have Gamma distributions. The exception occurs for the last in-sample regime and minor complications are caused by the fact that this last regime may not be completed at time T. For this Poisson intensity we use an accept/reject algorithm.

Standard Bayesian results for state space models can be used to show $p(V^{-1}|Y_T, S_T, \Theta, \lambda, \eta, \beta_\lambda)$ is Wishart and $p(\eta^{-1}|Y_T, S_T, \Theta, \lambda, V, \beta_\lambda)$ is Gamma. $p\left(\beta_\lambda^{-1}|Y_T, S_T, \Theta, \lambda, V, \eta\right)$ can also be shown to be a Gamma distribution.

The predictive distribution of y_{T+J} can be generated by noting that, if one knew what regime y_{T+j} was in, then standard results for simulating a predictive distribution in the Normal linear regression (or autoregressive model if X_t contains lags of the dependent variable). That is, the properties of $p(y_{T+J}|Y_T, S_{T+J}, \Theta, \lambda, V, \eta, \mathcal{V}_V, \mathcal{V}_\eta, \beta_\lambda)$ can be calculated using standard simulation methods using output from the MCMC algorithm. Note that the probability of future states S_{T+J} are provided at each iteration of our MCMC algorithm by using the underlying Markov chain.

6 Application to US Inflation and Output

In economics, many applications of change-point modeling have been to the decline in volatility of US real activity and changes in the persistence of the inflation process. With regards to GDP growth, Kim, Nelson and Piger (2003), using the methods of Chib (1998) (assuming a single change-point), investigate breaks in various measures of aggregate activity. For most of the measures they consider, the likelihood of a break is overwhelming and Bayesian and frequentist analyses produce very similar results.[12] On the

[12] Since such papers consider only a single break, it is relatively easy to evaluate all the possible break points. Kim, Nelson and Piger (2003) assume that the conditional mean parameters remain constant across the break and the only change is in the innovation variance. If one allowed both the conditional mean and variance to break and assumed an exchangeable Normal- Gamma prior then the model can be evaluated analytically. This was the approach followed in Koop and Potter (2001) and it has the advantage that one

other hand, Stock and Watson (2002) present evidence from a stochastic volatility model that the decline in variance might have been more gradual, a thesis first forward by Blanchard and Simon (2001).

Clark (2003) discusses the evidence on time variation in persistence in inflation. Cogley and Sargent (2001, 2003) present evidence of time variation in the inflation process both in the conditional mean and conditional variance of a smooth type. Stock (2001) finds little evidence for variation in the conditional mean of inflation using classical methods and Primiceri (2003) finds some evidence for variation in the conditional variance but little in the conditional mean.

Accordingly, in our empirical work we investigate the performance of our model using quarterly US GDP growth and the inflation rate as measured by the PCE deflator (both expressed at an annual rates) from 1947Q1 through 2004Q2. With both variables we include an intercept and two lags of the dependent variable as explanatory variables. We treat these first two lags as initial conditions and, hence, our data effectively runs from 1947Q3 through 2004Q2.

In addition to our Poisson hierarchical model for durations we also present results for standard TVP with stochastic volatility [see Stock and Watson 2002] and one-break models estimated using Bayesian methods. Both of these can be interpreted as restricted versions of our model. The TVP model imposes the restrictions that the duration of each regime is one (i.e. $s_t = t$ for all t). The one-break model imposes the restriction that there are exactly two regimes with $s_t = 1$ for $t \leq \tau$ and $s_t = 2$ for $t > \tau$ (and the coefficients are completely unrestricted across regimes with a flat prior on the coefficients and error variances).[13]

Appendix C describes our selection of the prior hyperparameters $\underline{\alpha}_\lambda$,$\underline{\xi}_1$,$\underline{\xi}_2$,$\underline{\alpha}_\eta$,$\underline{\beta}_\eta$,$\underline{V}_V$ and $\underline{\nu}_V$ and comparable prior hyperparameters for the TVP model. Here we note only that we make weakly informative choices for these. In a more substantive empirical exercise we would carry out a prior sensitivity analysis. The researcher interested in more objective elicitation could work with a prior based on a training sample.

---

can also integrate out over lag length in a trivial fashion.

[13] We restrict the prior for the change-points such that the change-point cannot occur in the first or last 5% of the sample.

6.1 Empirical Results

Figure 1 presents information relating to the TVP for GDP growth. The posterior means of the coefficients (i.e. ϕ_t for $t = 1, .., T$) are given in Figure 1a and the volatilities (i.e. $\exp(\sigma_t/2)$ for $t = 1, .., T$) in Figure 1b. Figure 2 presents similar information from the one-break model.

Consider first results from the standard TVP and one break models for real GDP growth. The most interesting findings for this variable relate to the volatility. Both models indicate that volatility is decreasing substantially over time, with a particularly dramatic drop occurring around 1984.[14] However, with the TVP model this decline is much more smooth and non-monotonic than with the one break model. The question arises as to whether the true behavior of volatility is as suggested by the TVP model or the one break model. Of course, one can use statistical testing methods which compare these alternatives. However, an advantage of our model is that it nests these alternatives. We can estimate what the appropriate pattern of change is and see whether it is the TVP or the one break model – or something in between.

Our findings relating to volatility of GDP growth are not surprising given previous results starting with McConnell and Perez (2000). There is some evidence from the TVP model that volatility started to decline in the 1950s but this decline was reversed starting in the late 1960s. The single break model (by construction) does not show any evidence of this. It dates the single break to be at or very near to 1984. With regards to the autoregressive coefficients, with both models the posterior means indicate that suggest that little change has taken place. However, posterior standard deviations (not presented) are quite large indicating a high degree of uncertainty. In the literature [e.g. Stock and Watson (2002)] these findings have been interpreted as implying that there has been no change in the conditional mean parameters.

In light of this approximate constancy of the coefficients (and to illustrate our methods in an interesting special case), we estimate our model with variation only in the volatilities and not in the coefficients. That is, the first equation in (4.2) is degenerate (or, equivalently, $V = 0_{K\times K}$). Figure 3 plots features of the resulting posterior. Figure 3a, by definition, implies

[14] Note that, in the one break model, the posterior means of the coefficients and volatilities, *conditional on a particular change-point*, will behave like step functions. However, when we present unconditional results, which average over possible change-points, this step function pattern is lost as can be seen in the figures.

constant values for all the coefficients. Figure 3b plots the posterior mean of the volatility. This figure is slightly smoother but otherwise quite similar to the comparable TVP result in Figure 1b, but differs quite substantially from the one break model result. Thus, we are finding evidence that the TVP model is more sensible than the one break model. However, we found such evidence in the context of a model which could have allowed for very few breaks. In fact, as can be seen in Figure 3c, our model indicates that there are about 40 regimes in-sample, as opposed to the 2 regimes of the one break model or 227 of the TVP model.

Let us now turn to inflation. Given findings by other authors and an interest in the persistence of inflation, we use the unrestricted version of our model and allow the AR coefficients to change across regimes. Figures 4, 5 and 6 present results from the TVP, one break and our models, respectively. Figure 4a, containing the two autoregressive parameters from the TVP model, shows a slight but steady increase in the persistence of inflation up to the late 1970s followed by a steady decrease. The fact that the level of inflation increased throughout the 1970s and early 1980s before declining in the 1990s is picked up partly through the pattern in the intercept.[15] The volatility of inflation shows a similar pattern, with a noted increase in the 1970s and early 1980s. These sensible results are consistent with evidence presented in Cogley and Sargent (2001), although at odds with some of the evidence presented in Primiceri (2003). Note, however, that our model yields a smoother evolution especially of the volatility since it finds smoother changes in persistence than the standard TVP model. Thus it is able to allocate more of the high standard deviation of inflation in the 1970s to changes in the conditional mean and does not require an increase in the variance of the innovations to inflation. The single break model indicates quite different patterns (see Figure 5). It wants to put the single break near the beginning of the sample, totally missing any changes in the level, persistent or volatility of inflation in the 1970s and early 1980s. One could force the break later by adopting a prior that the change-point is later in the sample. As one can see by concavity in Figure 6c our model is able to assign many change-points early in the sample then adapt to a slower rate of regime change later in the sample.

The last panel of all of the figures plots the predictive density one period

[15] Note, of course, that the unconditional mean depends on the intercept and the AR coefficients.

ahead. For both GDP growth and inflation, results using our model and the TVP model are similar to one another. This is not surprising given that both models imply similar coefficients and volatilities at the end of the sample and it is either certain or quite probable that a regime switch occurs out-of-sample. This latter feature accounts for the fact that the one break model exhibits a predictive density which has slightly smaller standard deviation and thinner tails. For inflation, results are particularly strong with the one break model indicating both a much tighter density but also one shifted upwards relative to our model and the TVP model.

When comparing results from the TVP and one-break model to ours, as a general rule we are finding our model supports many change-points rather than a small number and thus the movements of the conditional mean and variance parameters are closer to the TVP model. We take this as evidence that our methods are successfully capturing the properties of a reasonable data generating process, but without making the assumption of a break every period as with the TVP model. That is, we are letting the data tell us what key properties of the data are, rather than assuming them. Our empirical results also show the problems of working with models with a small number of breaks when, in reality, the evolution of parameters is much more gradual.

7 Conclusions

In this paper we have developed a change-point model which nests a wide range of commonly-used models, including TVP models and those with a small number of structural breaks. Our model satisfies the five criteria set out in the introduction. In particular, the maximum number of regimes in our model is not restricted and it has a flexible Poisson hierarchical prior distribution for the durations. Furthermore, we allow for information (both about durations and coefficients) from previous regimes to affect the current regime. The latter feature is of particular importance for forecasting.

Bayesian methods for inference and prediction are developed and applied to real GDP growth and inflation series. We compare our methods to two common models: a single-break model and a time varying parameter model. We find our methods to reliably recover key data features without making the restrictive assumptions underlying the other models.

8 References

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

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.

Blanchard, O. and Simon, J., 2001, The long and large decline in US output volatility, *Brookings Papers on Economic Activity*, 1, 135-174.

Bracqemond, C. and Gaudoin, O., 2003, A survey on discrete lifetime distributions, *International Journal of Reliability, Quality and Safety Engineering*, 10, 69-98.

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

Carter, C. and Kohn R., 1994, On Gibbs sampling for state space models, *Biometrika,* 81, 541-553.

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., 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.

Chopin, N. and Pelgrin, F., 2004, Bayesian inference and state number determination for hidden Markov models: An application to the information content of the yield curve about inflation, *Journal of Econometrics*, 123, 327-344.

Clark, T., 2003, Disaggregate evidence on the persistence of consumer price inflation, Federal Reserve Bank of Kansas City, Working Paper RP#03-11 available at http://www.kc.frb.org/publicat/reswkpap/PDF/RWP03-11.pdf

Clements, M. and Hendry, D., 1998, *Forecasting Economic Time Series.* (Cambridge University Press: Cambridge).

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, forthcoming *Review of Economic Dynamics.*

DeJong, P. and Shephard, N., 1995, The simulation smoother for time series models, *Biometrika*, 82, 339-350.

Durbin, J. and Koopman, S., 2002, A simple and efficient simulation smoother for state space time series analysis, *Biometrika*, 89, 603-616.

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.

Fernandez, C., Osiewalski, J. and Steel, M.F.J., 1997, On the use of panel data in stochastic frontier models with improper priors, *Journal of Econometrics*, 79, 169-193.

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

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.

Kim, S., Shephard, N. and Chib, S., 1998, Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies*, 65, 361-93.

Koop, G. and Poirier, D., 2004, Empirical Bayesian inference in a nonparametric regression model, in *State Space and Unobserved Components Models Theory and Applications*, edited by Andrew Harvey, Siem Jan Koopman and Neil Shephard (Cambridge University Press: Cambridge).

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, Prior elicitation in multiple change-point models, working paper available at http://www.le.ac.uk/economics/gmk6/.

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

Nyblom, J.1989, Testing for the constancy of parameters over time, *Journal of the American Statistical Association,* 84, 223-230.

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.

Primiceri, G., 2003, Time varying structural vector autoregressions and monetary policy, working paper available at http://www.faculty.econ.northwestern.edu/faculty/prim

Stock, J., 2001, Comment on Cogley and Sargent, *NBER Macroeconomic Annual.*

Stock, J. and Watson, M., 1996, Evidence on structural instability in macroecconomic 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.*

9 Appendix A: A Modified Version of Chib (1996)'s Algorithm

Bayesian inference in the model of Chib (1998) is based on a Markov Chain Monte Carlo (MCMC) algorithm with data augmentation. If $\Theta = (\theta_1', \ldots, \theta_M')'$ and $P = (p_1, \ldots, p_{M-1})$ and we expand the definition of the state to include the duration of a regime in addition to the number of the regime then the algorithm proceeds by sequentially drawing from

$$\Theta, P | Y_T, S_T \tag{A.1}$$

and

$$S_T | Y_T, \Theta, P. \tag{A.2}$$

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

$$p\left(S_T|Y_T,\Theta,P\right)=p\left(s_T|Y_T,\Theta,P\right)p\left(s_{T-1}|Y_T,S^T,\Theta,P\right) \qquad \text{(A.3)}$$
$$\cdots .p\left(s_t|Y_T,S^{t+1},\Theta,P\right)\cdots p\left(s_1|Y_T,S^2,\Theta,P\right).$$

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

$$p\left(s_t|Y_T,S^{t+1},\Theta,P\right)\propto p\left(s_t|Y_t,\Theta,P\right)p\left(s_{t+1}|s_t,P\right). \qquad \text{(A.4)}$$

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 this case and a finite number in the general one), we need only to worry about $p\left(s_t|Y_t,\Theta,P\right)$. 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:

$$p\left(s_t=k|Y_t,\Theta,P\right)=\frac{p\left(s_t=k|Y_{t-1},\Theta,P\right)p\left(y_t|Y_{t-1},\boldsymbol{\theta}_k\right)}{\sum_{m=k-1}^{k}p\left(s_t=m|Y_{t-1},\Theta,P\right)p\left(y_t|Y_{t-1},\boldsymbol{\theta}_m\right)}, \qquad \text{(A.5)}$$

using the fact that

$$p\left(s_t=k|Y_{t-1},\Theta,P\right)=\sum_{m=k-1}^{k}p_{mk}p\left(s_{t-1}=m|Y_{t-1},\Theta,P\right), \qquad \text{(A.6)}$$

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 (A.4) for every time period using (A.5) and (A.6) beginning at $t=1$ and going forward in time. If there is uncertainty over χ then the prior over the unobserved duration of the initial regime is used to initialize the algorithm. If there is not uncertainty then it is assumed that the first period of the first regime is the first period of the sample. Then the states themselves are drawn using (A.3), beginning at period T and going backwards in time. Of course, the draw of s_T will be of a regime number and a duration. The duration will locate the time when the regime started, thus it immediately identifies the last change-point in sample. This information is used to find the appropriate time period at which to evaluate (A.4). In the case where $\chi=0$ the draw of the duration of the second regime directly gives the duration of the first regime and the this is the final iteration. In the case where there is uncertainty over χ the final iteration of the algorithm will involve drawing the duration of regime 1 using (A.3) and (A.4).

10 Appendix B: MCMC Algorithm

The posterior conditionals used in our MCMC algorithm are described in Section 5. Further details are provided in this Appendix. The states are drawn from $p\left(S_T|Y_T,\Theta,\lambda\right)$ using a modification of the algorithm of Chib (1996) as described in Appendix A.

$p\left(\Theta|Y_T,S_T,\lambda,V,\eta,\beta_\lambda\right)$ can be drawn from using methods of posterior simulation for state space models with stochastic volatility drawing using standard algorithms for state space models [e.g. Kim, Shephard and Chib (1998), Carter and Kohn (1994), DeJong and Shephard (1995) or Durbin and Koopman (2002)]. These algorithms can be used for our change-point models with simple modifications. To fix ideas consider the simplest case where $\mathbf{X}_t$ contains only an intercept. Then

$$\frac{1}{d_m}\sum_{s_t=m} Y_{s_t} = \overline{Y}_m = \phi_m + \varepsilon_m,$$

with $\varepsilon_m \sim N(0, \exp(\sigma_m)/d_m)$. The draw of the ϕ_m (conditional on σ) could then proceed using the analyst's favorite algorithm in the time scale given by the change points. For σ_m (conditional on ϕ_m) one can apply a stochastic volatility algorithm [e.g. Kim, Shephard and Chib (1998)] using the measurement equation:

$$\sum_{s_t=m}\left(Y_{s_t}-\phi_m\right)^2 = \exp(\sigma_m)\varepsilon_m,$$

where $\varepsilon_m \sim N(0,d_m)$ in the time scale given by the change-points. Since, most algorithms analyze the stochastic volatility conditional on draws for the ϕ_m, the extension to the case where X_t contains more than just an intercept is immediate. We take draws of σ_m with a one-step sampler using an acceptance/rejection approach from Kim, Shephard and Chib (1998).

In the case where X_t contains more than just an intercept we proceed by first forming the predictive distribution for ϕ_m using the transition equation:

$$\phi_{s_t} = \begin{cases} \phi_m & \text{if } s_t = s_{t-1} = m \\ \phi_m + U_{m+1} & \text{if } s_t = m+1, s_{t-1} = m \end{cases},$$

and then use the Kalman filter to derive:

$$p(\phi_{s_t}|Y_t,S_T) \text{ for } t \in \{s_{t+1} \neq s_t\}.$$

We then draw ϕ_M from $p(\phi_{s_T}|Y_T, S_T)$ and proceed back through $M-1, \ldots, 1$. We use the Carter and Kohn (1994) sampler, modified for the time scale given by the change-points. For out-of-sample regimes we use the transition equation and values of V and η to generate draws starting from the value of ϕ_M.

The Poisson intensities are drawn from $p\left(\lambda_m|Y_T, S_T, \Theta, V, \eta, \beta_\lambda\right)$ for $m = 1, .., M$. These posterior conditionals are:

$$\lambda_m|Y_T, S_T V, \eta, \beta_\lambda \sim G\left(\overline{\alpha}_m, \overline{\beta}_m\right)$$

where

$$\begin{aligned} \overline{\alpha}_m &= \underline{\alpha}_\lambda + d_m \\ \overline{\beta}_m &= \left[\beta_\lambda^{-1} + 1\right]^{-1} \end{aligned}$$

Note, however, that for the last in-sample regime d_m is not observed. Suppose the ongoing regime has lasted D_m periods by the end of the sample. Then the conditional posterior is proportional to

$$\begin{aligned} & \exp(-\lambda_m/\overline{\beta}_m)\lambda_m^{\overline{\alpha}_m - 1} \sum_{y=D_m-1}^{\infty} \frac{\exp(-\lambda_m)\lambda_m^y}{y!} \\ = \; & \exp(-\lambda_m(1 + 1/\overline{\beta}_m))\lambda_m^{\overline{\alpha}_m - 1} \sum_{y=D_m-1}^{\infty} \frac{\lambda_m^y}{y!} \\ \leq \; & \exp(-\lambda_m(1 + 1/\overline{\beta}_m))\lambda_m^{\overline{\alpha}_m - 1} \exp(\lambda_m) \\ = \; & \exp(-\lambda_m/\overline{\beta}_m)\lambda_m^{\overline{\alpha}_m - 1}. \end{aligned}$$

Thus, a simple accept/reject step is available to draw λ_m for the regime with on-going duration. For large values of D_m relative to the values of λ_m implied by the Gamma distribution based on $\overline{\alpha}_m, \overline{\beta}_m^{-1}$ the rejection rate will be high. This should not happen to frequently, but if it does this step can be converted into a Metropolis-Hastings one.

We next turn to the posterior conditionals $p\left(V^{-1}|Y_T, S_T, \Theta, \lambda, \eta, \beta_\lambda\right)$ and $p\left(\eta^{-1}|Y_T, S_T, \Theta, \lambda, V, \beta_\lambda\right)$. These are:

$$V^{-1}|Y_T, S_T, \Theta, \lambda, \eta \sim W\left(\overline{V}_V, \overline{\nu}_V\right),$$

where

$$\overline{V}_V = \left[\underline{V}_V + \sum_{m=1}^{M} \left(\phi_m - \phi_{m-1}\right)\left(\phi_m - \phi_{m-1}\right)'\right]^{-1}$$

and

$$\overline{\nu}_V = \underline{\nu}_V + M.$$

Furthermore,

$$\eta^{-1}|Y_T, S_T, \Theta, \lambda \sim G\left(\overline{\alpha}_\eta, \overline{\beta}_\eta\right)$$

where

$$\overline{\alpha}_\eta = \underline{\alpha}_\eta + \frac{M}{2}$$

and

$$\overline{\beta}_\eta = \frac{1}{\underline{\beta}_\eta + \frac{1}{2}\sum_{m=1}^{M}\left(\sigma_m - \sigma_{m-1}\right)^2}$$

Finally, we have

$$\beta_\lambda^{-1}|Y_T, S_T, \Theta, \lambda, V, \eta \sim G\left(\overline{\alpha}_\beta, \overline{\beta}_\beta^{-1}\right)$$

where

$$\overline{\alpha}_\beta = M\underline{\alpha}_\lambda + \underline{\xi}_1$$

and

$$\overline{\beta}_\beta = \sum_{m=1}^{M}\lambda_m + \frac{1}{\underline{\xi}_2}$$

11 Appendix C: Properties of the Prior

In the body of the text, we developed some theoretical properties of the prior. However, given its complexity, it is also instructive to examine its implications using prior simulation. Accordingly, in this appendix, we illustrate some key properties of our prior for the hyperparameter values used in the

empirical work. We use informative priors. For highly parameterized models such as this, prior information can be important. Indeed, results from the Bayesian state space literature show how improper posteriors can result with improper priors [see, e.g. Koop and Poirier (2004) or Fernandez, Ley and Steel (1997) for more general results]. One strategy commonly-pursued in the related literature [see, e.g., Cogley and Sargent (2001, 2003)] is to restrict coefficients to lie in bounded intervals such (e.g. the stationary interval). This is possible with our approach. However, this causes substantial computational complexities (which are of particular relevance in our model where many regimes can occur out-of-sample and reflect relatively little data information). Training sample priors can be used by the researcher wishing to avoid subjective prior elicitation.

In this paper, we choose prior hyperparameter values which attach appreciable prior probability to a wide range of reasonable parameter values. To aid in interpretation, note that our data is measured as a percentage and, hence, changes in σ_m in the interval $[-0.5, 0.5]$ are the limit of plausibility. For AR coefficients, the range of plausible intervals is likely somewhat narrower than this. With regards to the durations, we want to allow for very short regimes (to approach the TVP model) as well as much longer regimes (to approach a model with few breaks). We choose values of the prior hyperparameters, $\underline{\alpha}_\lambda, \underline{\xi}_1, \underline{\xi}_2, \underline{\alpha}_\eta, \underline{\beta}_\eta, \underline{V}_V$ and $\underline{\nu}_V$ which exhibit such properties.

Figure C plots the prior for key features assuming $\underline{\alpha}_\lambda = 12, \underline{\xi}_1 = \underline{\alpha}_\lambda, \underline{\xi}_2 = \underline{\alpha}_\lambda, \underline{\alpha}_\eta = 1.0, \underline{\beta}_\eta = 0.02, \underline{V}_V = 0.1 I_K$ and $\underline{\nu}_V = 3K$. Note that, by construction, the priors for all our conditional mean coefficients are the same so we only plot the prior for the AR(1) coefficient. Figure C1 plots the prior over durations and it can be seen that the prior weight is spread over a wide range, from durations of 1 through more than 50 receiving appreciable prior weight. Figures C2 and C3 plot prior standard deviations for the state equation innovations (see 4.2). It can be seen that these are diffuse enough to accommodate anything from the very small shifts consistent with a TVP model through much bigger shifts of a small break model.

For the TVP model, we make the same prior hyperparameter choices (where applicable). The prior for the one break model has already been described in the text.

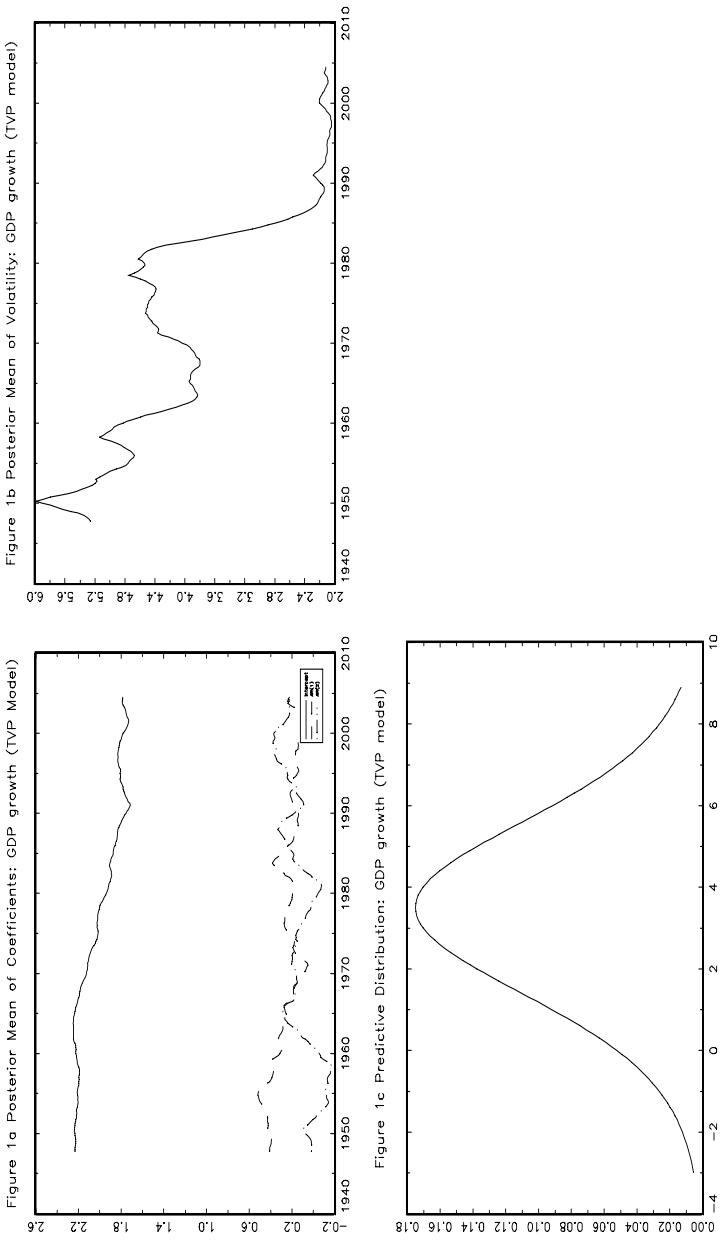
Figure 1a Posterior Mean of Coefficients: GDP growth (TVP Model)
Figure 1b Posterior Mean of Volatility: GDP growth (TVP model)
Figure 1c Predictive Distribution: GDP growth (TVP model)

Figure 1:

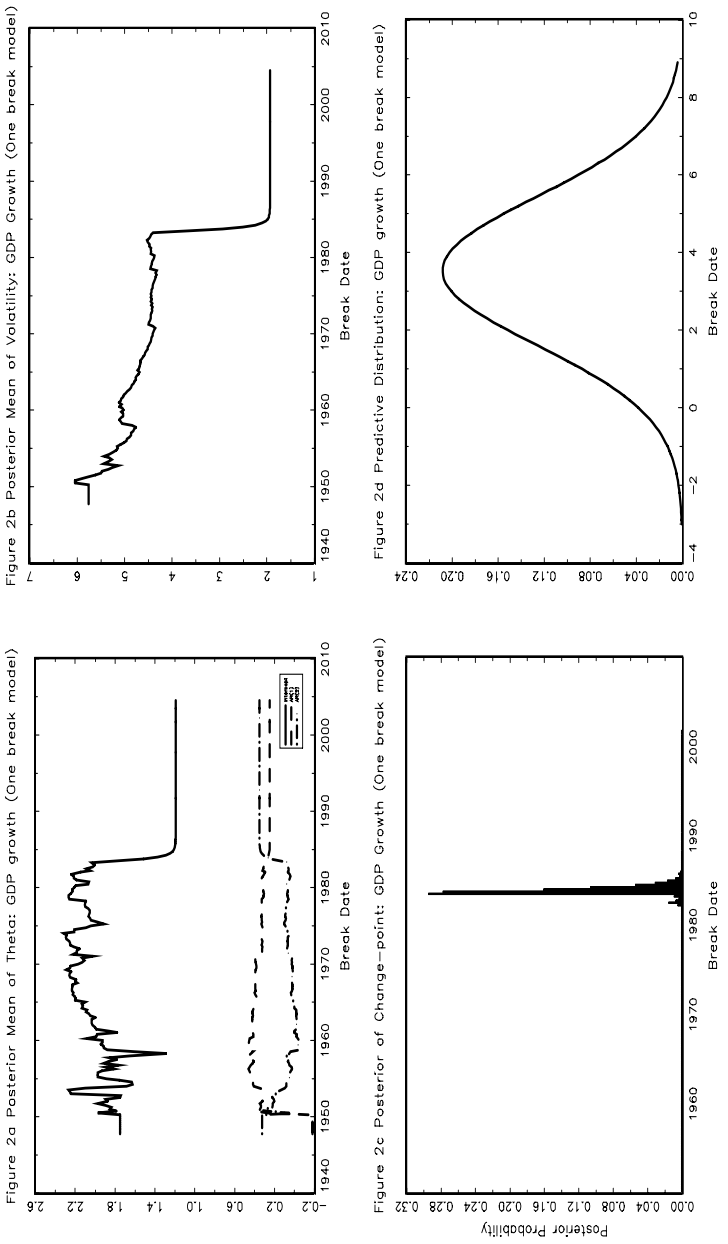
Figure 2a Posterior Mean of Theta: GDP growth (One break model)
Break Date
Figure 2b Posterior Mean of Volatility: GDP Growth (One break model)
Break Date
Figure 2c Posterior of Change-point: GDP Growth (One break model)
Break Date
Posterior Probability
Figure 2d Predictive Distribution: GDP growth (One break model)
Break Date

Figure 2:

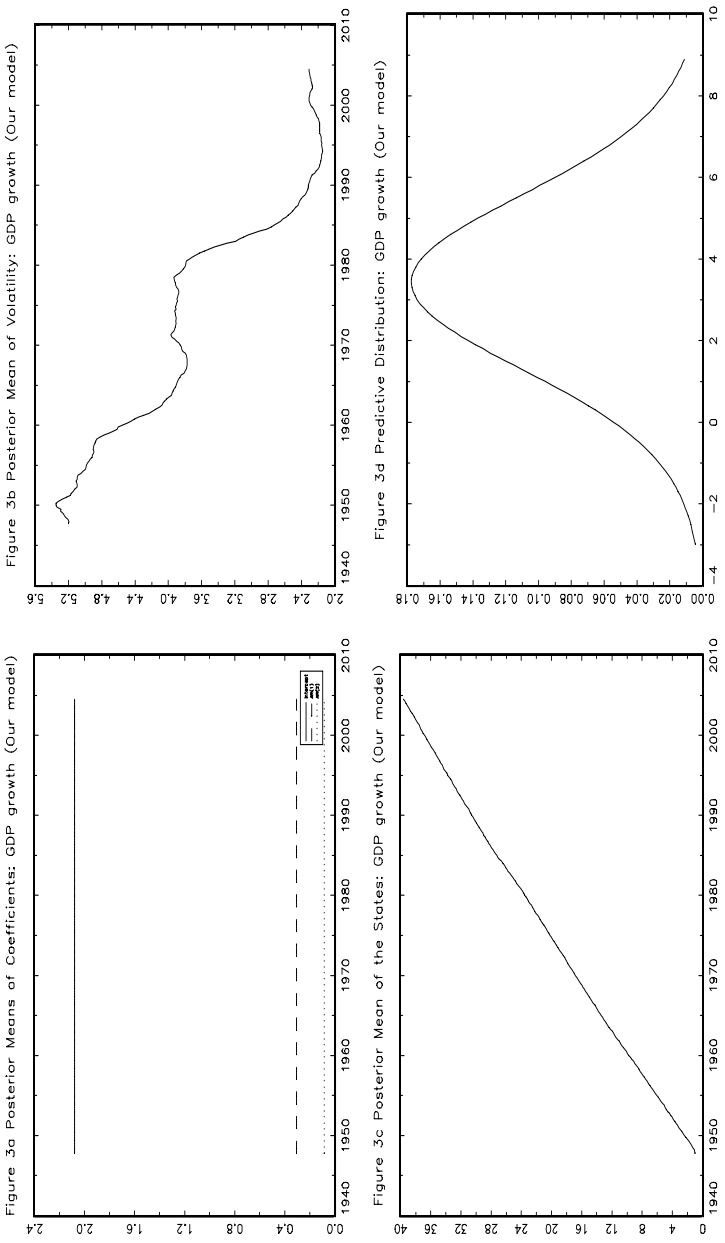
Figure 3a Posterior Means of Coefficients: GDP growth (Our model)
Figure 3b Posterior Mean of Volatility: GDP growth (Our model)
Figure 3c Posterior Mean of the States: GDP growth (Our model)
Figure 3d Predictive Distribution: GDP growth (Our model)

Figure 3:

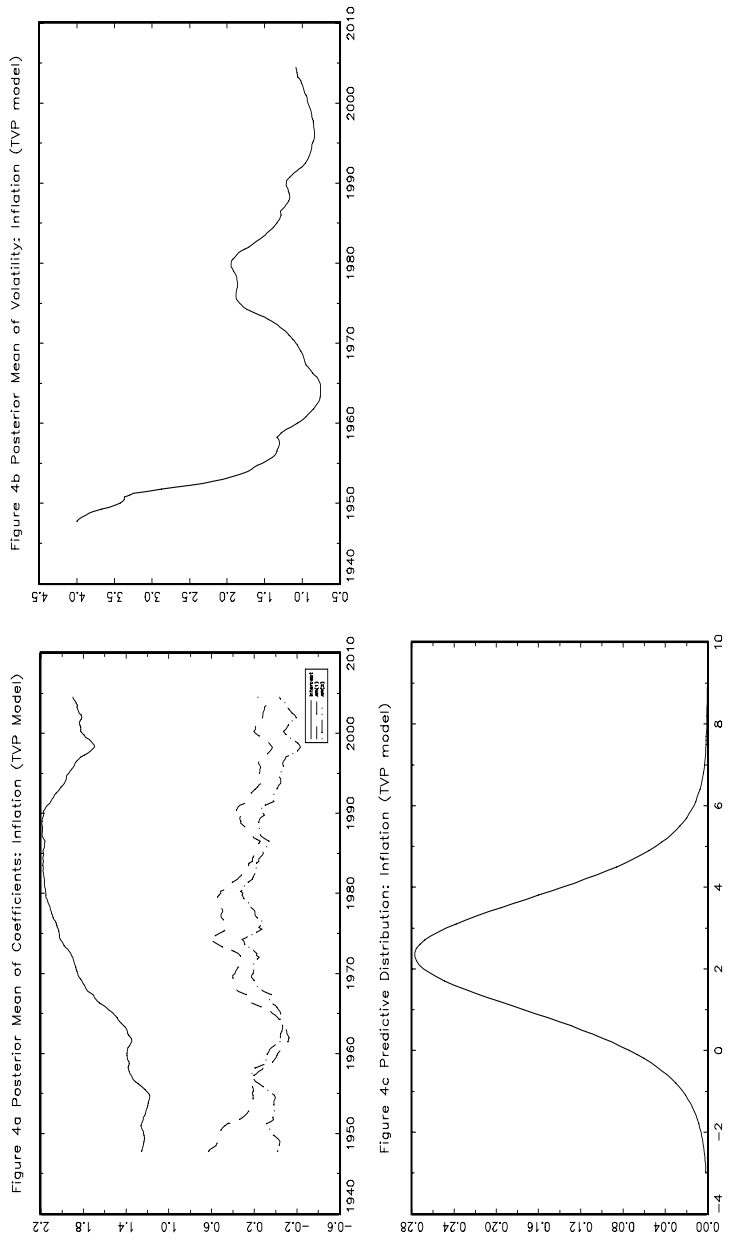
Figure 4a Posterior Mean of Coefficients: Inflation (TVP Model)
Figure 4b Posterior Mean of Volatility: Inflation (TVP model)
Figure 4c Predictive Distribution: Inflation (TVP model)

Figure 4:

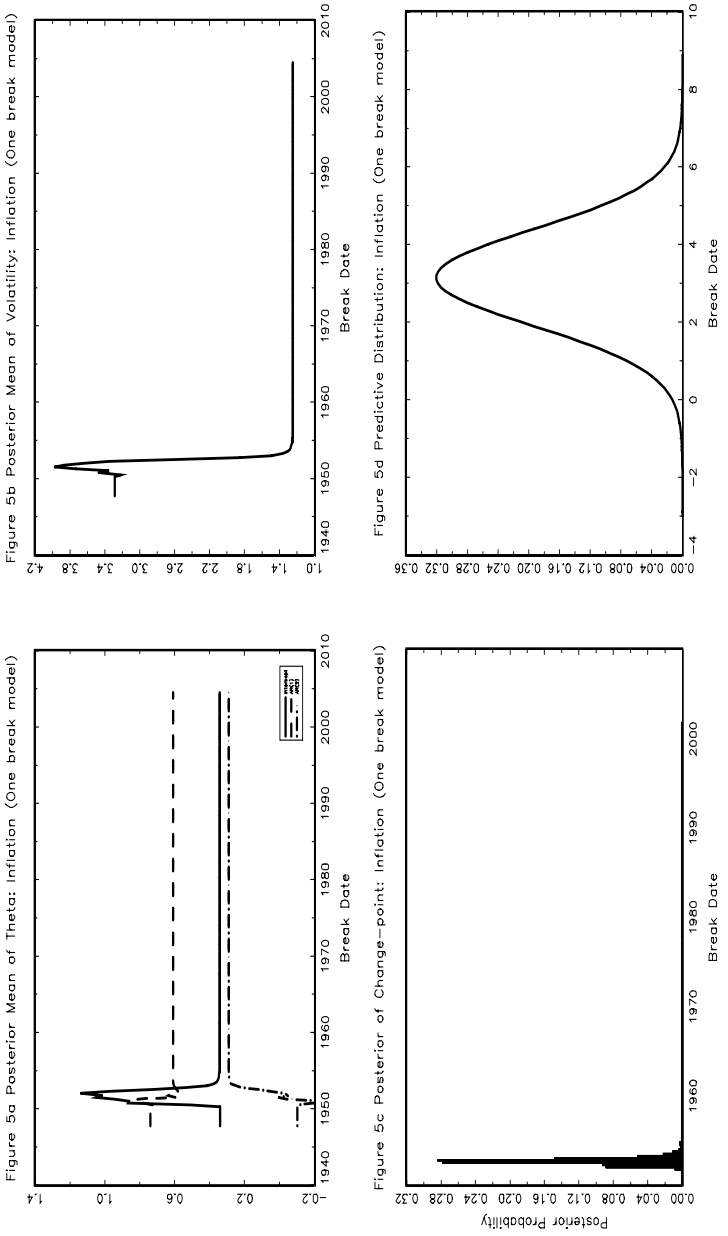
Figure 5a Posterior Mean of Theta: Inflation (One break model)
Break Date
Figure 5b Posterior Mean of Volatility: Inflation (One break model)
Break Date
Figure 5c Posterior of Change-point: Inflation (One break model)
Posterior Probability
Break Date
Figure 5d Predictive Distribution: Inflation (One break model)
Break Date

Figure 5:

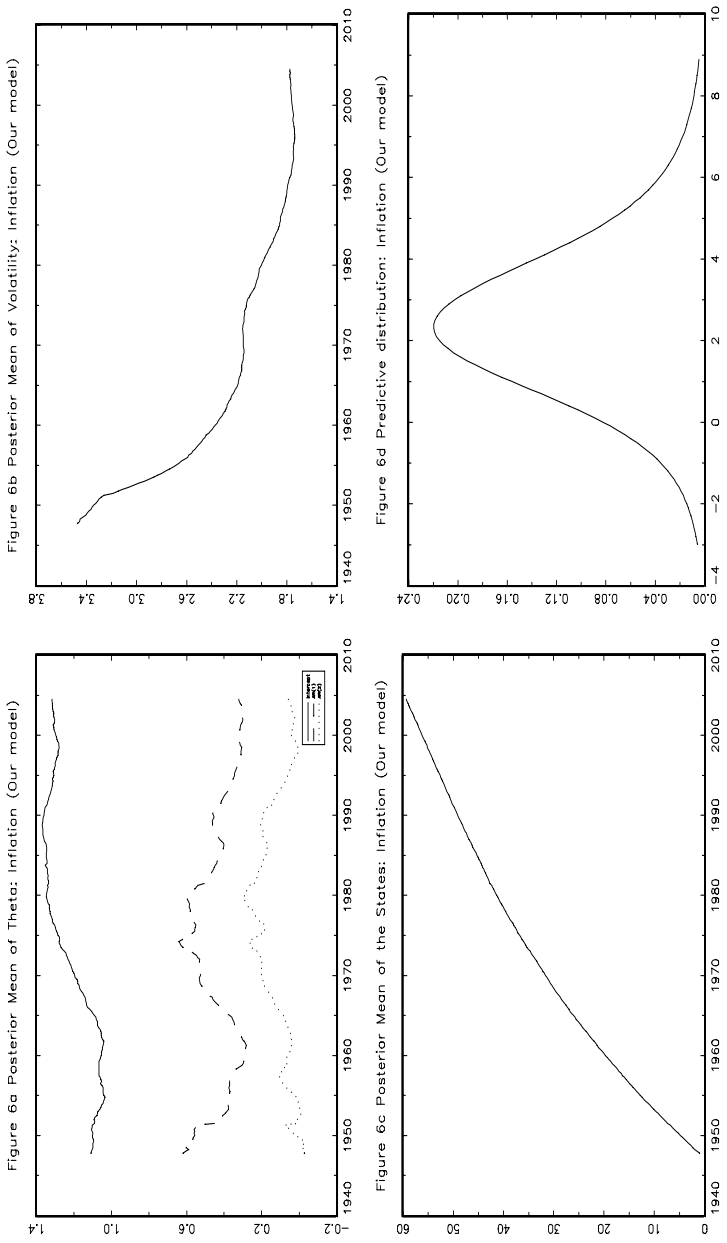
Figure 6a Posterior Mean of Theta: Inflation (Our model)
Figure 6b Posterior Mean of Volatility: Inflation (Our model)
Figure 6c Posterior Mean of the States: Inflation (Our model)
Figure 6d Predictive distribution: Inflation (Our model)

Figure 6:

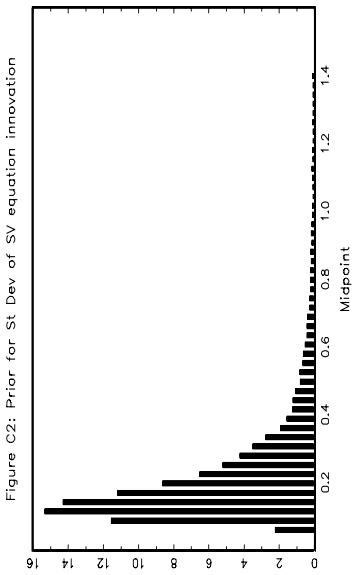
Figure C2: Prior for St Dev of SV equation innovation
Midpoint
0
2
4
6
8
10
12
14
16
0.2
0.4
0.6
0.8
1.0
1.2
1.4

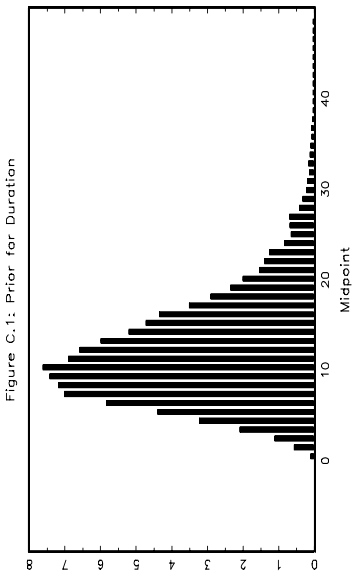
Figure C.1: Prior for Duration
Midpoint
0
1
2
3
4
5
6
7
8
10
20
30
40

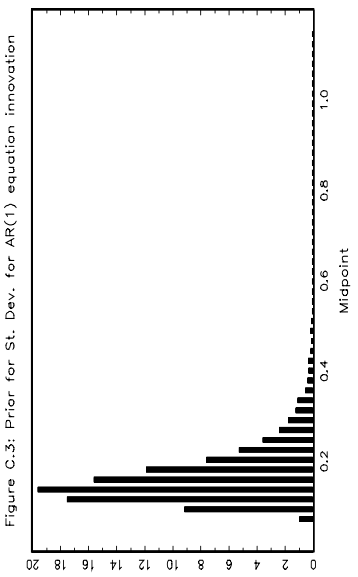
Figure C.3: Prior for St. Dev. for AR(1) equation innovation
Midpoint
0
2
4
6
8
10
12
14
16
18
20
0.2
0.4
0.6
0.8
1.0

Figure 7: