Estimation and Forecasting in Models with Multiple Breaks*

Gary Koop† and Simon M. Potter‡

November 2004, revised September 2006

Abstract

This paper develops a new approach to change-point modeling that allows the number of change-points in the observed sample to be unknown. The model we develop assumes regime durations have a Poisson distribution. It 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 considerable attention on the construction of reasonable hierarchical priors both for regime durations and for the parameters which characterize each regime. A Markov Chain Monte Carlo posterior sampler is constructed to estimate a version of our model which allows for change in conditional means and variances. We show how real time forecasting can be done in an efficient manner using sequential importance sampling. Our techniques are found to work well in an empirical exercise involving US GDP growth and inflation. 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 (with stochastic volatility) model.

JEL classification: C11, C22, E17

Keywords: Bayesian, structural break, Markov Chain Monte Carlo, hierarchical prior, sequential importance sampling

*We would like to thank Paolo Giordani, Edward Leamer, Richard Paap, Hashem Pesaran, Herman van Dijk, two referees, the editor and seminar participants at the Swedish Central Bank, the Federal Reserve Bank of St. Louis, Erasmus University and the Universities of Amsterdam, Kansas and 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.

†Department of Economics, University of Strathclyde, Glasgow U.K., Gary.Koop@strath.ac.uk

‡Federal Reserve Bank of New York, Simon.Potter@ny.frb.org

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 (2006). 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.

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 is 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.

The previous paragraphs highlight the two main issues which must be addressed when building a change-point model: how to model the duration of each regime and how to model the change in coefficients after a break occurs. In this paper, we develop a new model which, we argue addresses both of these issues in a empirically-sensible manner. For reasons outlined below, we use Bayesian methods. The model we develop 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.

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.

6. The change-point model should be easy to update in real time as new data arrives on the time series of interest.

It can immediately be seen that standard implementations of the TVP model and models with small numbers of breaks do not have these features. Furthermore, as we shall see, common Bayesian approaches to these models (which augment standard implementations with hierarchical priors) also do not have these features.

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, in a model with $M > 1$ regimes, 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. That is, 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. 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 are difficult under the standard classical approach to change-point problems. However, with some partial exceptions [e.g. Chib (1998), Pesaran, Pettenuzzo and Timmerman (2006) 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.

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 hierarchical structure to the TVP model. Section 5 gives an overview of the posterior simulator used in our Bayesian analysis.[1] This section also shows how sequential importance sampling (i.e. particle filtering[2]) methods can be used with our model to carry out real time forecasting in a computationally efficient manner. 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 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 (2006)].[3] 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 dis-

[1] Complete details are provided in the working paper version with title "Forecasting and Estimating Multiple Change-point Models with an Unknown Number of Change-points" available at http://personal.strath.ac.uk/gary.koop/.

[2] Particle filtering is sequential importance sampling with resampling. For our purposes, resampling is not required (although we describe how it can be done).

[3] In contrast, a recent influential Bayesian paper, Maheu and Gordon (2005), does not apply such a hiearchical prior structure and, thus, is not directly comparable to the papers discussed in this section.

crete random variable, s_t, which takes on values $\{1, 2, \ldots, M\}$. We let $S_i = (s_1, \ldots, s_i)'$ 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

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

The durations of regimes are defined as:[4]

$$d_m = \tau_m - \tau_{m-1}.$$

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

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

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.[5]

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. $\tau_1, \ldots, \tau_{M-1}$). 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 (i.e. $d_1, .., d_m$). 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.

There are many advantages to adopting the framework of Chib (1998). For instance, previous models often involved searching over all possible sets

[4] In this definition we set $\tau_0 = 0$ and, thus, the first observation signifies the beginning of the first regime. An alternative treatment, pursued in the working paper version of this paper, is to let τ_0 be an unknown parameter.

[5] 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)]. In the working paper version, we discuss how Bai and Perron's (1998, 2003) methods can be re-interpreted in this way.

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 (2006)]. 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. 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 states' transition probabilities 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 Potter (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

$$\Pr[s_T = M|s_{T-1} = M] = p_M, \Pr[s_T = M|s_{T-1} = M-1] = 1-p_{M-1}. \quad (2.3)$$

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

$$\Pr[s_T = M|s_{T-1} = M] = \Pr[s_T = M|s_{T-1} = M-1] = 1. \quad (2.4)$$

If $M > 2$, additional restrictions are required. To express these restrictions in words, consider the case $M = 3$. If, in period $T-1$, we are not already in the third regime, then it must the case that a regime switch occurs in period T and this must be imposed on the model. Similarly, if, in period $T-2$, we are still in the first regime, then we must impose that regime switches occur in both periods $T-1$ and T, in order to ensure that $M = 3$. 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. For the issue of forecasting, these issues are

even more important, since these prior restrictions occur at the end of the sample, precisely where forecasting begins.

Chib (1998) did not consider the question of out-of-sample forecasting. If one were to assume no additional breaks occur out-of-sample, forecasting could be done in a straightforward fashion, based on the likelihood and prior which hold at the end of the sample (i.e. $p\left(y_T|Y_{T-1},\theta_M\right)$ and $p\left(\theta_M\right)$). Such an approach, of course, does not address the issue of forecasting when breaks can occur out of sample. Pesaran, Pettenuzzo and Timmerman (2006) took up the challenge of extending the Chib approach to address this latter issue. They assume a constant transition probability matrix out-of-sample which allows for a probability of a break occurring in each out-of-sample period. Their approach is attractive and, in many ways, a sensible one. However, in adapting the approach of Chib (1998) in this manner, some problems arise. Remember that, to impose exactly M regimes in-sample, restrictions such as (2.4) must be imposed. But, out-of-sample, to allow for breaks to occur, it is desirable to revert to an unrestricted transition probability such as (2.3). Pesaran, Pettenuzzo and Timmerman (2006) explicitly assume that

$$P[s_t = M|s_{t-1} = M] = \left\{ \begin{array}{cc} 1 & \text{if } t \leq T \\ p_M & \text{if } t > T \end{array} \right. , \tag{2.5}$$

with the restriction that $P[s_T = M] = 1$. This assumption, since it is at the end of sample, could conceivably have an important influence on forecasting. To try and understand this assumption better, consider what would happen if we increase the observed sample by one observation. Most Bayesians would argue that any statistical procedure for updating in response to the addition of an extra observation should satisfy Bayes' rule. However, the updating of Pesaran, Pettenuzzo and Timmerman (2006) with an extra observation could be taken to imply:

$$P[s_t = M|s_{t-1} = M] = \left\{ \begin{array}{cc} 1 & \text{if } t \leq T+1 \\ p_M & \text{if } t > T+1 \end{array} \right. , P[s_{T+1} = M] = 1,$$

which is inconsistent with (2.5) and violates Bayes' Rule.

These problems arise due to the imposition of exactly M breaks in-sample. Pesaran, Pettenuzzo and Timmerman (2006) partly address this problem through considering models with different values for M and then doing Bayesian model averaging. This is a sensible thing to do, but does not fully address the problems noted above (i.e. a pile-up of probability near the end of the sample and the difficulty in adapting the approach to out-of-sample forecasting in a way which does not violate Bayes' rule). In

the next section, we will propose our model which does not impose a fixed number of breaks in-sample and, hence, does not run into these problems.

Another relevant paper is McCulloch and Tsay (1993). The model used in this paper is different in many ways from Chib (1998) and Pesaran, Pettenuzzo and Timmerman (2006). However, it does have regimes changing with a certain probability. McCulloch and Tsay do not assume a fixed number of breaks and do not face the problems noted above since they assume that the probability of a break occurring is constant for all observations (in an extension, they allow this probability to depend on additional covariates). In essence, whereas Chib (1998) allows the probability of a break in regime j to be p_j, in McCulloch and Tsay (1993) this is simplified to a single p. This p can be estimated in-sample and then used in out-of-sample forecasting, thus precluding the need for an assumption such as (2.5). However, the assumption of a time- and regime-invariant transition probability is restrictive in many macroeconomic contexts. Furthermore, it will share many of the restrictive features of Chib (1998). For instance, the hierarchical prior for regime duration will be a Geometric distribution.

In summary, the pioneering work of Chib (1998) followed by the work of Pesaran, Pettenuzzo and Timmerman (2006) has changed the way many look at change-point models. Both papers have had great influence and have many attractive features. 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 these previous approaches.

3 A Poisson Hierarchical Prior for Durations

The previous 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 six criteria listed in the introduction. In this section we develop our alternative approach based on a Poisson model for durations.[6] 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

[6] 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. Our methods could be extended to deal with any of these in a straightforward manner.

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 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).[7] The desirable properties of this feature are explored in more detail in Koop and Potter (2004).[8]

[7] 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 anything up to a TVP model and an out-of-sample break.

[8] 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.

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 requires calculating $p\left(s_{t+1}|s_t, P\right)$ where P is the matrix of transition probabilities given in (2.2). In the model of Chib (1998) this density is simple due to the constant transition probability assumption. However, in our model the transition probability is not constant. As shown in the working paper version of this paper, 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 with the initial observation.[9] If $\widetilde{d}_m$ denotes the current duration of regime m, we can derive the transition probability

$$\Pr[s_{t+1} = m+1|s_t = m, \widetilde{d}_m] = \frac{\exp(-\lambda_m)\lambda_m^{\widetilde{d}_m-1}}{(d_m-1)!\left(1-\sum_{j=0}^{\widetilde{d}_m-2}\frac{\exp(-\lambda_m)\lambda_m^j}{j!}\right)}, \widetilde{d}_m = 1, \ldots, T-m, \tag{3.2}$$

where $\sum_{s=0}^{-1}\frac{\exp(-\lambda_m)\lambda_m^j}{j!}$ is defined to be 0. This must be evaluated for all $T-m$ possible values of $\widetilde{d}_m$. Thus, unlike with Chib's model, the transition probability matrix depends on the durations of the regimes.

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

[9] This can be generalized to allow for the first regime to begin $\chi < \infty$ periods before the initial observation. In an earlier version of the paper, we adopted such a specification and showed how χ could be treated as an unknown parameter.

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

$$\lambda_m \sim G\left(\underline{\alpha}_\lambda, \underline{\beta}_\lambda\right), \tag{3.3}$$

then the posterior for λ_m in many of the regimes will simply be $G\left(\underline{\alpha}_\lambda, \underline{\beta}_\lambda\right)$.[10]

In theory, there is nothing wrong with using an independent prior such as (3.3), and simplified versions of the methods described below can be used 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 (3.3) and use a hierarchical prior of the form:

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

where β_λ is an unknown parameter (not a hyperparameter selected by the researcher).[11]

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:

$$\beta_\lambda^{-1} \sim G(\underline{\xi}_1, 1/\underline{\xi}_2). \tag{3.5}$$

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

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

[11] 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.

It is important to understand the implications of any prior (the appendix discusses such properties by simulating from the particular prior used in our empirical work). As discussed following (3.1), we propose a hierarchical prior where $p(d_m|\lambda_m)$ is Poisson, but if we integrate out λ_m, we get $p(d_m|\beta_\lambda)$ being a Negative Binomial distribution. The unconditional prior distribution, $p(d_m)$ is found by integrating out β_λ^{-1}. This does not have a closed form.[12] In general $p(d_m)$ inherits the flexible form of $p(d_m|\lambda_m)$ or $p(d_m|\beta_\lambda)$. However, it is worth mentioning that if $\underline{\alpha}_\lambda = 1$ then we have the restrictive property 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, by suitable choice of $\underline{\xi}_1$ and $\underline{\xi}_2$, 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.

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 predictive distribution 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. $p(y_t|Y_{t-1}, s_t = m)$ can have many forms), our Poisson model for durations can be used with any specification for $p(y_t|Y_{t-1}, s_t = m) = p(y_t|Y_{t-1}, \theta_m)$. Here we choose a particular structure based on a regression or autoregressive model with stochastic volatility which is of empirical relevance.

Many change-point models assume that the anything can happen to the parameters after a regime change occurs. The issues which arise with such an approach are elegantly expressed by Pastor and Stambaugh (2001) in an application which used stock return data to investigate the equity premium.

[12] The working paper version of this paper contains an expression for it that could be evaluated numerically.

> "In standard approaches to models that admit structural breaks, estimates of current parameters rely on data only since the most recent estimated break. Discarding the earlier data reduces the risk of contaminating an estimate of the equity premium with data generated under a different [process]. That practice seems prudent, but it contends with the reality that shorter histories typically yield less precise estimates. Suppose... a shift in the equity premium occurred a month ago. Discarding virtually all of the historical data on equity returns would certainly remove the risk of contamination by pre-break data, but it hardly seems sensible in estimating the current equity premium. Completely discarding the pre-break data is appropriate only when the premium might have shifted to such a degree that the pre-break data are no more useful ..., than, say, pre-break rainfall data, but such a view almost surely ignores economics." Pastor and Stambaugh (2001, pages 1207-1208).

The case for adopting a hierarchical prior which allows for some sort of link between pre- and post-break parameters is, we believe, compelling in many empirical contexts. The question arises as to what sort of hierarchical prior is appropriate. 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 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.[13]

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

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

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}. A similar approach is adopted in McCulloch and Tsay (1993) for the intercept and error variance in an autoregressive model. In most traditional models with a small number of breaks, one assumes θ_m and θ_{m-1} are independent of one another [e.g. Chib (1998) or Maheu and Gordon (2005)]. 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 (2006). They place a hierarchy on the parameters of the conjugate family for each regime. For instance, Pesaran, Pettenuzzo and Timmerman (2006) assume that all the $\theta_m s$ are drawn from a common distribution. This is a standard approach in the Bayesian literature for cross-section data drawn from different groups. In a time series applications 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 hierarchical prior incorporates.

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[14] and we assume that $\underline{\nu}_V > K + 1$.

[14] We parameterize the Wishart distribution so that if $Z \sim W(A, a)$ and ij subscripts

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:

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

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 estimated 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:

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

denote elements of matrices, then $E\left(Z_{ij}\right) = aA_{ij}$. The scalar a is a degrees of freedom parameter.

where $p(\theta_m|\theta_{m-1}, V, \eta)$ is given by (4.2), $p(\theta_0)$ is diffuse, (V, η) is given by (4.3), $p(\lambda_m|\beta_\lambda)$ is given by (3.4) and $p(\beta_\lambda)$ is given by (3.5).

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 provided in the working paper version of this paper.[15] 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 implies 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 $p(S_T|Y_T, \Theta, \lambda, V, \eta, \beta_\lambda) = p(S_T|Y_T, \Theta, \lambda)$ can be drawn using the modified algorithm of Chib (1996) with transition probabilities given by (3.2). $p(\Theta|Y_T, S_T, \lambda, V, \eta, \beta_\lambda)$ can be simulated using extensions of methods of posterior simulation for state space models 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 $p\left(V^{-1}|Y_T, S_T, \Theta, \lambda, \eta, \beta_\lambda\right)$ is Wishart and $p\left(\eta^{-1}|Y_T, S_T, \Theta, \lambda, V, \beta_\lambda\right)$ 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.

5.1 Predictive Distributions

Suppose interest centers on the predictive density for y_{T+h}, given data through time T. The basic idea of our simulation algorithm is that, if we knew the parameters chracterizing the duration distribution (i.e. λ, β_λ), the

[15] This is available at http://personal.strath.ac.uk/gary.koop/

relevant regime (i.e. s_{T+h}) and the coefficients which applied in this regime (i.e. θ_h), then the distribution of y_{T+h} would simply be Normal. That is,[16]

$$\begin{aligned} & p\left(y_{T+h}|Y_T, s_{T+h}=\{m,d_m\},\theta_m,\lambda,V,\eta,\beta_\lambda\right) \qquad (5.1) \\ = & \frac{1}{\sqrt{2\pi\exp(\sigma_m)}}\exp\left[-\frac{(y_{T+h}-X_{t+h}\phi_m)^2}{2\exp(\sigma_m)}\right]. \end{aligned}$$

Thus, if we can obtain posterior draws of $s_{T+h}, \theta_m, \lambda, V, \eta, \beta_\lambda$, we can obtain an estimate of the predictive density as:

$$\widehat{p\left(y_{T+h}|Y_T\right)} = \frac{\sum_{r=1}^{R} p\left(y_{T+h}|Y_T, s_{T+h}=\{m^{(r)},d_m^{(r)}\},\theta_m^{(r)},\lambda^{(r)},V^{(r)},\eta^{(r)},\beta_\lambda^{(r)}\right)}{R}.$$

where (r) superscripts, for $r=1,..,R$, denote these draws (after dropping an appropriate number of burn-in replications). Our MCMC algorithm provides draws of $\lambda, V, \eta, \beta_\lambda$. In this subsection, we describe how draws of s_{T+h} and θ_m are taken (i.e. how out-of-sample draws of regimes and accompanying coefficients are done). We focus on a particular approach that complements the sequential importance sampler introduced in the next section.

We start by noting that we have access to the predictive distribution for the states, $\{p(s_{T+h}|Y_T,\Theta,\lambda,\beta_\lambda) : h=1,\ldots,H\}$ at each iteration of the MCMC algorithm described in the preceding section by combining $p(s_T|Y_T,\Theta,\lambda,\beta_\lambda)$ with the transition function of the chain. Thus, draws of $(s_{T+1},\ldots,s_{T+H})$ can be easily obtained.[17] We emphasize that these draws are conditional on Θ (which includes draws up to M). Note also that we require only a single draw of $(s_{T+1},\ldots,s_{T+H})$ at each iteration of the MCMC algorithm. Given these draws of the regimes, we can then take draws $\{\theta_{s_{T+h}} : h=1,\ldots,H\}$. If the regime at time $T+H$ is less or equal to M then the elements of Θ are used and no additional random draws are required. If the regime number goes above M then we can use (4.2) to generate new values of $\theta_{s_{T+h}}$.

We stress that this approach will provide us with a predictive density that satisfies Bayes' rule. That is, it correctly combines information in the data with the probabilistic structure implied by the model using the rules of conditional probability.

[16] Note that, in the case of the autoregressive model, X_{t+1} is known and this density can be immediately calculated. For out-of-sample forecasting for $h>1$, X_{t+h} is produced by iterating on the known values at T using the sequence of autoregressive coefficients.

[17] For some values of h regime numbers greater than M might have non-zero probability. At this occurrence the Negative Binomial distribution implied by β_λ can be used to obtain an appropriate regime duration draw.

5.2 Sequential Importance Sampling

Models such as ours are often used for real time forecasting. As new data comes in, we want to update our forecasts. Of course, we could simply re-run our MCMC algorithm with a data set augmented by this new data and then calculate predictive densities as described in the previous section. However, in some cases, it is desirable to update forecasts without re-running the MCMC algorithm. In this section, we briefly describe how sequential importance sampling methods [these are a popular type of particle filter methods, see, e.g., Doucet, Godsill and Andrieu (2000) or Liu and Chen (1998)] can be used with our model to achieve this goal.

Let $z_t = (\theta_t', S_t)$ denote the unobserved states and, for simplicity, suppress conditioning arguments (all the p.d.f.s below are conditional on the model parameters $(\lambda, V, \eta, \beta_\lambda)$). The sequential importance sampler (SIS), is designed for models (such as ours) which can be written in terms of $p(y_t|z_t)$ and $p(z_t|z_{t-1})$. Based on a sample of size T, posterior and predictive features depend upon $p(Z_T|Y_T)$ which can be obtained using our MCMC algorithm where $Z_i = (z_1, \ldots, z_i)'$. Now suppose a $(T+1)^{st}$ observation becomes available and, thus, posterior and predictive features depend on $p(Z_{T+1}|Y_{T+1})$ which, by Bayes' rule, can be written as:

$$p(Z_{T+1}|Y_{T+1}) = p(Z_T|Y_T) \frac{p(y_{T+1}|z_{T+1}) p(z_{T+1}|z_T)}{p(y_{T+1}|Y_T)}.$$

Since $p(Z_T|Y_T)$ and $p(y_{T+1}|Y_T)$ are impossible to directly evaluate simulation methods are required. The likelihood, $p(y_{T+1}|z_{T+1})$, is evaluated using (5.1).

We can, of course, use the MCMC algorithm described above to evaluate properties of $p(Z_{T+1}|Y_{T+1})$. However, an alternative is to use importance sampling. If we take an importance function of the form:

$$\pi(z_{T+1}|Z_T, Y_{T+1}),$$

then the importance sampling weights become:

$$w_{T+1} = \frac{p(Z_T|Y_T) p(y_{T+1}|z_{T+1}) p(z_{T+1}|z_T)}{\pi(z_{T+1}|Z_T, Y_{T+1})}.$$

If a $(T+2)^{nd}$ observation becomes available and, hence, posterior and predictive features depend on $p(Z_{T+2}|Y_{T+2})$ which can also be handled using importance sampling. But note that the importance sampling weights now become:

$$w_{T+2} = w_{T+1} \frac{p\left(y_{T+2}|z_{T+2}\right) p\left(z_{T+2}|z_{T+1}\right)}{\pi\left(z_{T+2}|Z_{T+1}, Y_{T+2}\right)}.$$

In other words, we can recursively update the importance sampling weights rather than evaluating them anew, reducing computational effort. In general, importance sampling weights for $T+h$, are given by:

$$w_{T+h} = w_{T+h-1} \frac{p\left(y_{T+h}|z_{T+h}\right) p\left(z_{T+h}|z_{T+h-1}\right)}{\pi\left(z_{T+h}|Z_{T+h-1}, Y_{T+h}\right)}.$$

Computationally, such an approach can be very efficient. That is, instead of running a (computationally) demanding posterior simulation algorithm $h+1$ times (i.e. using data through period $T+i$ for $i = 0, 1, .., h$), the main posterior simulation algorithm is run only once, and then SIS allows for the fast and efficient updating as new information arises.

To be precise, if we begin using our MCMC algorithm for data through period T, we obtain $r = 1, .., R$ draws. These can be interpreted as importance sampling draws, each with an equal weight of $\frac{1}{R}$. In period $T+h$, using SIS we have R draws, each with weights $w_{T+h}^{(r)}$. As with any importance sampler, posterior properties of any feature of interest can be found by taking a weighted average of drawn features of interest using the normalized weights:

$$\widetilde{w}_{T+h}^{(r)} = \frac{w_{T+h}^{(r)}}{\sum_{r=1}^{R} w_{T+h}^{(r)}}.$$

Furthermore, the predictive likelihood[18] for H observations can estimated from the SIS output as:

$$p(y_{T+1}, \widehat{.., y_{T+H}}|Y_T) = \frac{\sum_{r=1}^{R} w_{T+H}^{(r)}}{R}.$$

The predictive likelihood for a single observation can be estimated using:

$$p(\widehat{y_{T+h}}|Y_T) = \sum_{r=1}^{R} p(y_{T+h}|Y_T, z_{T+h-1}^{(r)}, \lambda^{(r)}, V^{(r)}, \eta^{(r)}, \beta_{\lambda}^{(r)}) \widetilde{w}_{T+h-1}^{(r)}.$$

[18] As Poirier (1995), Chapter 8 discusses there is a lot of controversy in the frequentist literature about the meaning of predictive likelihood. In our Bayesian context the interpretation is clear: it is the predictive distribution evaluated at the observed value for y_{T+h}. Equivalently, it can be interpreted as a marginal likelihood, treating the posterior at time T as a prior.

In theory, SIS could be used to draw the states for all time periods (i.e. we could set $T = 0$ in the preceding equations and simply have h index time). However, all sequential importance sampling algorithms have the property that the variance of the importance sampling weights increase over time (i.e. the algorithm becomes less efficient as h increases). If the variance of the importance sampling weights becomes too large, then the importance sampling estimates become dominated by only a few of the draws. One simple measure of the effective number of draws is

$$\widehat{R}_{T+h}^{eff} = \frac{R}{\frac{1}{R}\sum_{r=1}^{R}(w_{T+h}^{(r)})^2}$$

There are numerous methods [see, e.g., Doucet, Godsill and Andrieu (2000) or Liu and Chen (1998)] which attempt to minimize this problem by using a resampling technique when this effective number of draws moves below a threshold. In our case, we have the advantage that if the effective number of draws falls below the threshold value we can always switch back to our original MCMC algorithm.

An obvious choice for the importance function in our case is our hierarchical prior evaluated at the parameters values after observing the first T observations:

$$\pi\left(z_{T+h}|Z_{T+h-1}, Y_{T+h-1}\right) = p\left(z_{T+h}|z_{T+h-1}, \lambda, V, \eta, \beta_\lambda\right).$$

To simulate from this importance function we use the same approach as described for producing the predictive distributions.

5.3 Computation Issues in Context

Computation issues are important in the change-point literature. Unless the number of change-points is very small, the computational burden can be quite demanding. Bayesian approaches, such as ours and Chib (1998), which adopt a hierarchical prior for the change-points, will have a much lower computational burden than those which involve evaluating something (e.g. a likelihood or a posterior) at every possible set of change-points. Our MCMC approach largely draws on standard algorithms and, hence, programming costs are not large. Our approach will involve a computational burden greater than Chib (1998) since our transition probabilities depend on the duration spent in each regime (see equation 3.3 and surrounding discussion). 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), so the added computational burden only enters through draws of the states. With a modern personal computer, the computational burden of our approach is still trivial for data sets of the length typically used by macroeconomists.[19] Furthermore, methods such as sequential importance sampling can be used to lessen the computational burden in real time forecasting exercises. In our model the ability to use the sequential importance sampling produces a tremendous reduction in net computation because most of the draws required are produced during the MCMC algorithm and just need to be stored.

The approach of Pesaran, Pettenuzzo and Timmerman (2006) shares many of the computational advantages of Chib (1998). However, some of the benefits of assuming a constant transition probability within a regime (except at the end of the sample), are lost in their forecasting exercise since their assumed (out-of-sample) hierarchy is not conditionally conjugate and the sequential importance sampling approach is not available due to their imposition of a fixed number of change-points in sample. This contrasts to our hierarchy, which is conditionally conjugate.

In contrast to our approach, the influential non-Bayesian 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. This is true in our approach as well, since we allow for any number of breaks in 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 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. Of course, without restrictions on the time between change-points the minimum is achieved with a perfect fit. To get around this issue, the minimizer is found by imposing additional restrictions on the minimum time between change-points. Inference is then performed by using insights from the asymptotic distribution. Note that this implies that a different partition of the data with the same number of change-points but a slightly higher value for the least squares minimand receives no weight in the inference. A strength of the Bayesian approach is that it puts weight on this

[19] A recent paper by Giordani and Kohn (2006) develops computationally efficient methods for the model of Chib (1998) and a simple version of our model. The authors show how, in either of these models, break dates can be drawn in $O(T)$ operations by integrating out the states analytically. It is probable that these methods can be extended for the general version of our model and, thus, computation will be very efficient indeed. The working paper version of our paper has further discussion on these points.

different partition. This would be particularly important for forecasting. In an earlier working paper version, we discussed how a frequentist could use a hierarchical approach to obtain information on the probability of change-points at the global minimizer for the regime coefficients found using Bai and Perron's algorithm

Relative to the approach of Bai and Perron (1998, 2003), we would have the same $O(T^2)$ computational burden if we had assumed the standard constant transition probability matrix (and, we note this holds for a much wider class of models than Bai and Perron). As we have seen, many Bayesian approaches adopt this constant transition probability matrix, assuming a fixed number of regimes occur in-sample. Our model is somewhat more computationally demanding than this. However, a simplified version of our model with a constant transition matrix would not be more computationally demanding. Note that such a version of our model would allow for an unknown number of regimes in sample. In the case where the transition probability was the same across regimes, the sequential importance sampler would be available and the net computational difference for real time updating would favor our method in most cases.

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.[20] On the 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, 2005) present evidence of time variation

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

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 (2005) 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 as an annual rates) from 1947Q2 through 2005Q4. 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 1947Q4 through 2005Q4.

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).[21]

The appendix 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 the appendix we carry out a prior sensitivity analysis. The researcher interested in more objective prior elicitation could work with a prior based on a training sample.

6.1 Estimation 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 $t = 1, \ldots, T$) 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 substan-

[21] In this one break model, we restrict the prior for the change-points such that the change-point cannot occur in the first or last 5% of the sample.

tially over time, with a particularly dramatic drop occurring around 1984.[22] 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. The posterior of the breakpoint (not presented here), is quite tight and indicates the single break to be at or very near to 1984. With regards to the autoregressive coefficients, with both models the posterior means 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 for the most interesting feature: volatility.[23] This figure is smoother but otherwise similar to the comparable TVP result in Figure 1b, but differs quite substantially from the one break model result. With our model, the number of regimes in-sample can be estimated. Its posterior mean is 45.35 (with posterior standard deviation of 12.70). This lies somewhere between the one break and TVP models (where the latter, by definition, will have $T = 233$ regimes). Thus, we are finding evidence that a model between the one break and TVP models is most sensible (although the TVP model is more sensible than the one break

[22] 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.

[23] The prior sensitivity analysis done in the appendix presents more posterior results for our model.

model). We stress that we have found such evidence in the context of a model which could have allowed for very few breaks.

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 sum of the two autoregressive parameters from the TVP model, shows an increasing tendency in the persistence of inflation up to the late 1970s followed by a decreasing tendency. 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. The volatility of inflation shows a similar pattern, with a noted increase in the 1970s and early 1980s. These sensible results are found by both the TVP model and our model and are consistent with evidence presented in Cogley and Sargent (2001), although at odds with some of the evidence presented in Primiceri (2005). But it is worth noting that our model smooths out some of the erratic patterns produced by the TVP model. 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, persistence 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.

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.

6.2 Predictive Exercise

The previous section focusses on estimation for our model, the TVP model and a one break model. In order to compare these different models, we could calculate marginal likelihoods in order to construct Bayes' factors or posterior odds ratios using standard methods. For instance, the methods

of Chib (1995), Chib and Jeliazkov (2001) or Gelfand and Dey (1994) can be used to estimate the marginal likelihood in change-point models. More simply, information criteria such as the Schwarz criteria can be used to approximate marginal likelihoods. However, in these models (which are very parameter rich) marginal likelihoods can be sensitive to priors. Accordingly, we prefer to compare models using predictive criteria such as the predictive likelihood discussed in Sections 5.1 and 5.2. As discussed in Section 5.2, these can easily be calculated using sequential importance sampling.

Table 1 presents predictive likelihoods for a period of two years at the end of our sample. That is, we use data through period 2003Q4 to calculate predictive distributions for each quarter through 2005Q4 and then evaluate the predictive at the observed outcome. It can be seen that, in terms of overall forecast performance over these 8 observations, our model does substantially outperform the one break model and (less substantially) the TVP model.

Table 1: Joint Predictive Likelihoods for 2004/2005			
	Our Model	TVP	One Break
GDP Growth	9.77×10^{-7}	6.06×10^{-7}	4.93×10^{-7}
Inflation	3.69×10^{-6}	3.31×10^{-6}	1.75×10^{-6}

Table 1 presents results relating to joint performance over two years. In a real time forecasting exercise, one might also be interested in forecasting performance one quarter at a time (where each quarter new data is used to update the predictive density). To illustrate how this can be done, we carry out a pseudo real-time forecasting exercise for 2004 and 2005. That is, beginning in 2003Q4 we construct the predictive distribution for 2004Q1, then use data through 2004Q1 to predict 2004Q2, etc.. A simple summary of forecasting performance involves seeing where the actual outcome lies in these one-period-ahead predictive distributions. In particular, we can calculate where the actual outcome lies in the predictive cumulative distribution function. Table 2 presents this information for our three models and two data series. An informal examination of Table 2 suggests all three of our models are predicting GDP growth fairly well. None of the outcomes are too far out in the tails of the predictive distributions. This is unsurprising since 2004-2005 were years of stable GDP growth with little evidence of structural change. More formal metrics of predictive performance can be developed by noting that, if a model is correct and there is no parameter uncertainty, then probabilities such as those in Table 2 should be drawn from the Uniform [0,1] distribution [see, e.g., Diebold, Gunther and Tay (1998)]. Using

the numbers in Table 2, the standard Chi-squared statistics for testing for Uniformity [e.g., Wonnacott and Wonnacott (1990), pages 550-551] are 9.5, 9.5 and 12.0 for our model, the TVP and the one break model, respectively. This provides some evidence that our model and the TVP are forecasting comparably to one another and the one break model is doing slightly worse. Note that the frequentist 0.05 critical value is 16.9 so we cannot reject the hypothesis of Uniformity for any of our models.

For inflation (which was somewhat more erratic in the 2004-2005 period), the Chi-squared statistics for testing Uniformity of the numbers in Table 2 are 12.0, 9.5 and 17.0 for our model, the TVP and the one break model, respectively. Hence, for inflation we are finding the TVP model forecasts slightly better than the other models. Furthermore, the frequentist hypothesis that the one break model is correct can be rejected at the 5% level of significance.

Table 2: Predictive Probability of Being Less than Actual Outcome

	GDP growth			Inflation		
	Our Model	TVP	One Break	Our Model	TVP	One Break
2004Q1	0.581	0.579	0.485	0.966	0.967	0.942
2004Q2	0.525	0.456	0.582	0.935	0.935	0.768
2004Q3	0.601	0.636	0.635	0.125	0.096	0.035
2004Q4	0.514	0.499	0.535	0.732	0.711	0.737
2005Q1	0.628	0.649	0.663	0.852	0.830	0.781
2005Q2	0.533	0.510	0.557	0.503	0.471	0.400
2005Q3	0.726	0.732	0.733	0.840	0.822	0.760
2005Q4	0.186	0.179	0.225	0.805	0.755	0.653

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 six 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.

Bai, J. and Perron, P., 2003, Computation and analysis of multiple structural change models, *Journal of Applied Econometrics* 18, 1-22.

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., 1995, Marginal likelihood from the Gibbs Sampler, *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.

Chib, S. and Jeliazkov, I., 2001. Marginal likelihood from the Metropolis-Hastings output, *Journal of the American Statistical Association*, 96, 270-281.

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., 2005, Drifts and volatilities: monetary policies and outcomes on Post WWII US, *Review of Economic Dynamics*, 8, 262-302.

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

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

Doucet, A., Godsill, S. and Andrieu, C., 2000, On sequential Monte Carlo sampling methods for Bayesian filtering, *Statistics and Computing*, 10, 197-208.

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., 2006, Optimally testing general breaking processes in linear time series models, *Review of Economic Studies*, forthcoming.

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.

Gelfand, A. and Dey, D., 1994, Bayesian model choice: Asymptotics and exact calculations, *Journal of the Royal Statistical Society Series B*, 56, 501-514.

Giordani, P. and Kohn, R., 2006, Efficient Bayesian inference for multiple change-point and mixture innovation model, manuscript.

Hamilton, J., 1994, *Time Series Analysis.* Princeton: Princeton University Press.

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/.

Liu, J. and Chen, R., 1998, Sequential Monte Carlo methods for dynamic systems, *Journal of the American Statistical Association,* 93, 1032–1044.

Maheu, J. and Gordon, S., 2005, 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.

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

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., 2006, Forecasting time series subject to multiple structural breaks, *Review of Economic Studies*, forthcoming.

Poirier, D., 1995, *Intermediate Statistics and Econometrics: A Comparative Approach.* Cambridge: The MIT Press.

Primiceri, G., 2005, Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies*, 72, 821-852.

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

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.*

Wonnacott, T. and Wonnacott, W., 1990, *Introductory Statistics for Business and Economics (Fourth edition).* New York: John Wiley and Sons.

Appendix: Properties of the Prior and Prior Sensitivity Analysis

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 as well as carry out a prior sensitivity analysis. 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)]. One strategy commonly-pursued in the related literature [see, e.g., Cogley and Sargent (2001, 2005)] 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.

Figures A1 through A3 plot 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 A1 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 A2 and A3 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.

In a sense, by presenting results for the TVP and one break models, we have already carried out a prior sensitivity analysis. That is, the TVP model can be considered as an extreme case where we use a dogmatic prior which imposes a break every time period and the one break model a dogmatic prior which imposes a precise number of breaks. However, in this appendix we provide some additional evidence on prior sensitivity with regards to the most important feature of our model: the prior on regime duration. In this regard, the prior for β is of most importance and it is characterized by two parameters, $\underline{\xi}_1$ and $\underline{\xi}_2$. In the empirical results section, we have set $\underline{\xi}_1 = \underline{\xi}_2 = 12$, values which yield the prior over durations in Figure A.1. Equation 3.5 and the discussion following it describes the prior. For present purposes, perhaps the most important things are to note that the prior mean of d_m is

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

if $\underline{\xi}_1 > 1$ and the properties of the Gamma distribution can be used to understand the other aspect of the prior. We consider an extremely wide range of prior hyperparameter values by allowing for $\underline{\xi}_1 = 1, 12, 100$ and $\underline{\xi}_2 = 1, 12, 100$. Table A.1 relates to GDP growth and presents the posterior properties of a key feature in our model: the number of in-sample regimes. Note that our base prior, used in the empirical work, provides an estimate of 45.35 regimes in-sample (with posterior standard deviation of 12.70). This is between the $T = 233$ regimes implied by the TVP model and the 2 regimes implied by the one break model. As we tighten the prior towards having short durations (i.e. towards the TVP model), by increasing $\underline{\xi}_1$ or decreasing $\underline{\xi}_2$, the number of regimes in sample does indeed increase (although this has relatively little effect on figures such as Figure 3 or on the predictive features). Tightening the prior in the direction of the one-break model (by

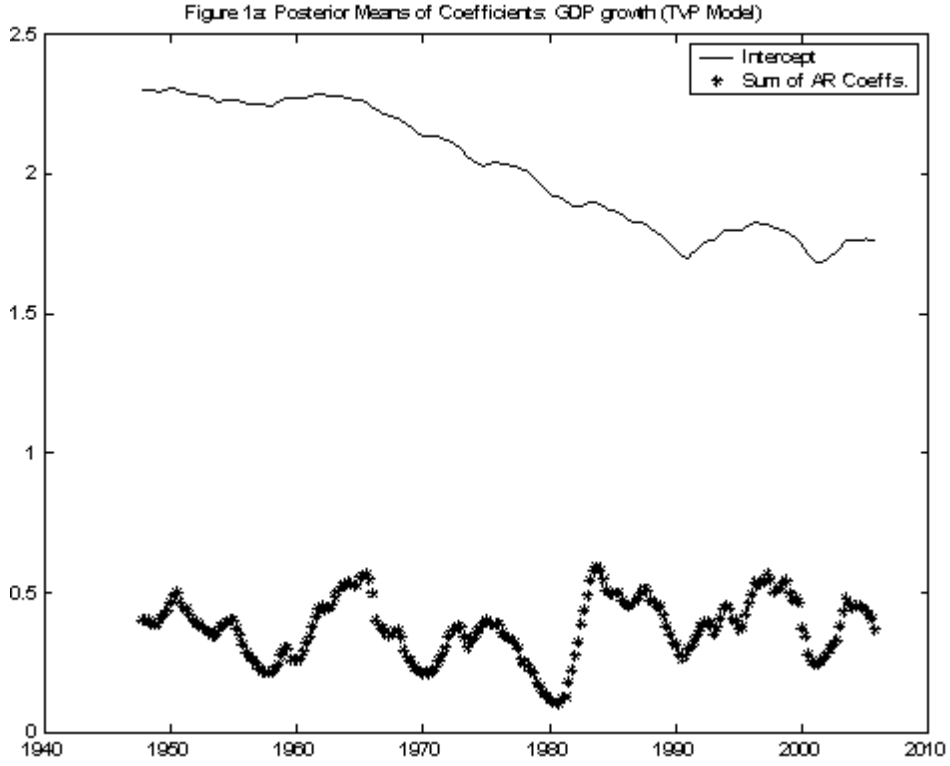
decreasing $\underline{\xi}_1$ or increasing $\underline{\xi}_2$) the number of regimes in sample does indeed decrease, but never to fewer than 12. That is, even if we set prior where the mean duration of a regime is very close to one, there is enough information in the data to pull the posterior in the direction of many regime changes.

Table A.1: Prior Sensitivity Analysis For GDP Growth			
Posterior Mean of # Regimes			
	$\underline{\xi}_2 = 1$	$\underline{\xi}_2 = 12$	$\underline{\xi}_2 = 100$
$\underline{\xi}_1 = 1$	73.32	12.48	12.15
$\underline{\xi}_1 = 12$	74.16	45.35	12.17
$\underline{\xi}_1 = 100$	78.29	76.23	17.38

Table A.2 repeats the analysis for inflation. Similar patterns appears. Note that, as expected, the inflation data is closer to coming from a TVP model than the GDP growth data. With our base prior, the number of in-sample regimes is estimated to be 124.00 (with posterior standard deviation of 32.64). Very large changes in the prior offer little change in this story. Even when the prior pulls strongly towards model with few breaks, the posterior still indicates very many breaks in-sample (i.e. the posterior mean number of regimes is at least 86.96). When the prior is shifted towards the TVP model, the number of in-sample breaks increases slightly, but the basic findings of our model are unaffected.

Table A.2: Prior Sensitivity Analysis For Inflation			
Posterior Mean of # Regimes (Post. St. Dev. in parentheses)			
	$\underline{\xi}_2 = 1$	$\underline{\xi}_2 = 12$	$\underline{\xi}_2 = 100$
$\underline{\xi}_1 = 1$	189.25	90.32	86.96
$\underline{\xi}_1 = 12$	189.80	124.00	87.96
$\underline{\xi}_1 = 100$	190.62	186.85	88.92

Figure 1a: Posterior Means of Coefficients: GDP growth (TVP Model)



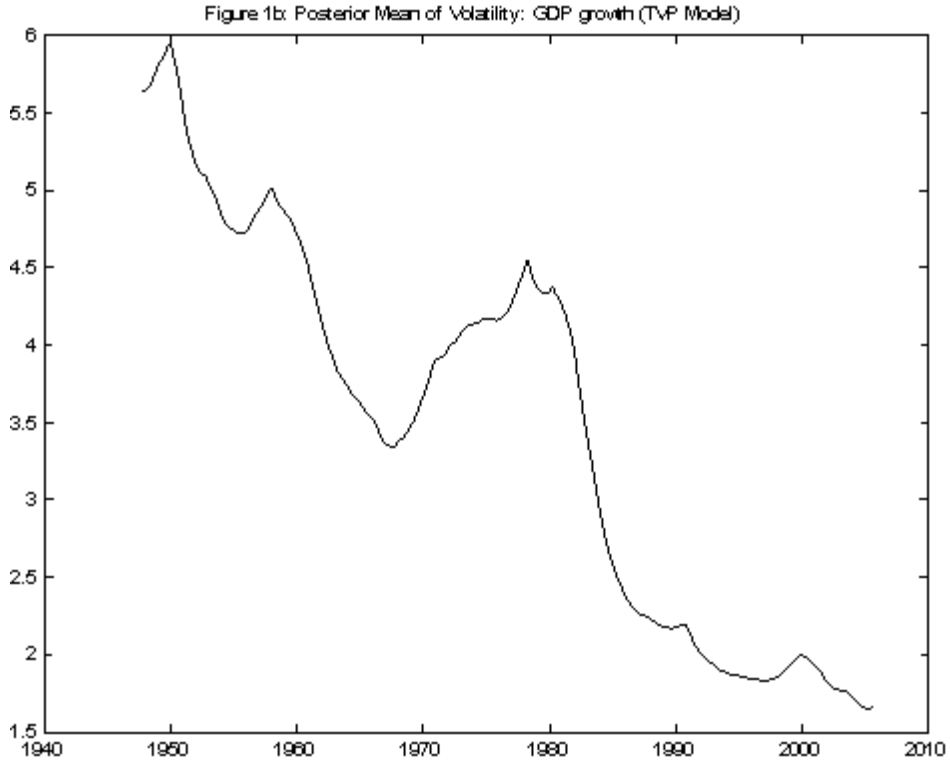


Figure 1b: Posterior Mean of Volatility: GDP growth (TVP Model)
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1940
1950
1960
1970
1980
1990
2000
2010

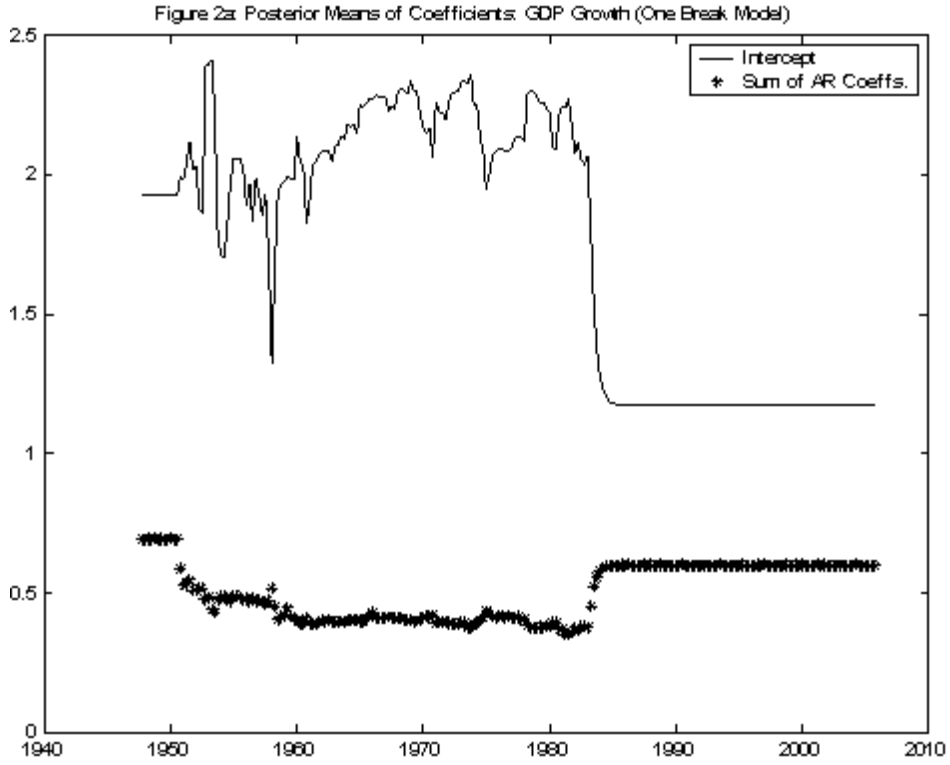


Figure 2a: Posterior Means of Coefficients: GDP Growth (One Break Model)
Intercept
Sum of AR Coeffs.
2.5
2
1.5
1
0.5
0
1940
1950
1960
1970
1980
1990
2000
2010

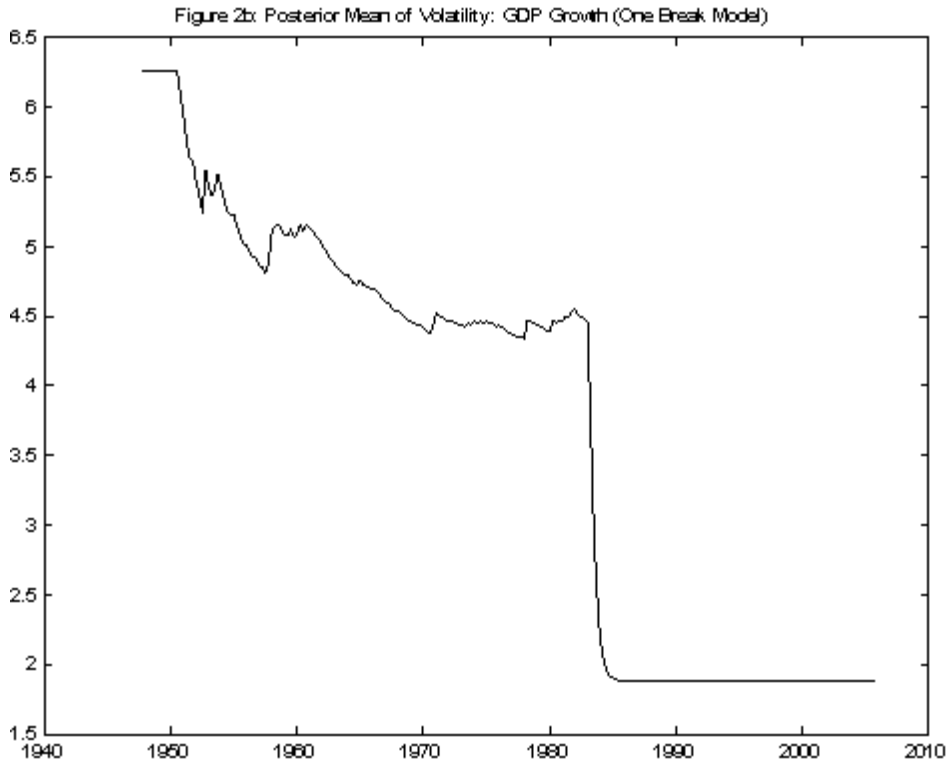


Figure 2b: Posterior Mean of Volatility: GDP Growth (One Break Model)
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1940
1950
1960
1970
1980
1990
2000
2010

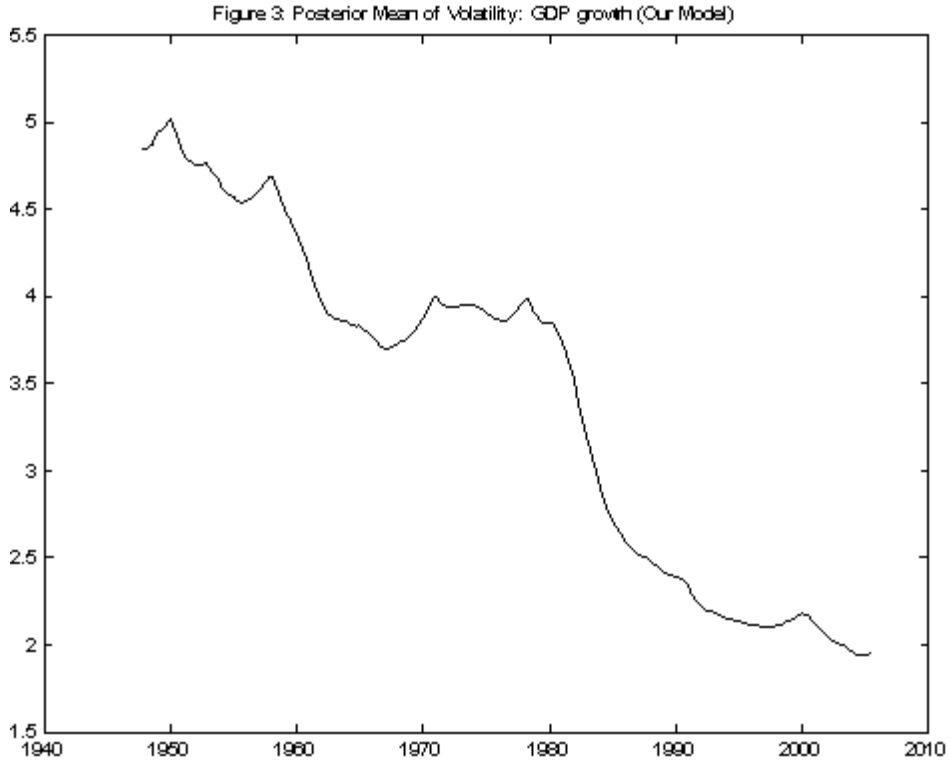


Figure 3: Posterior Mean of Volatility: GDP growth (Our Model)
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1940
1950
1960
1970
1980
1990
2000
2010

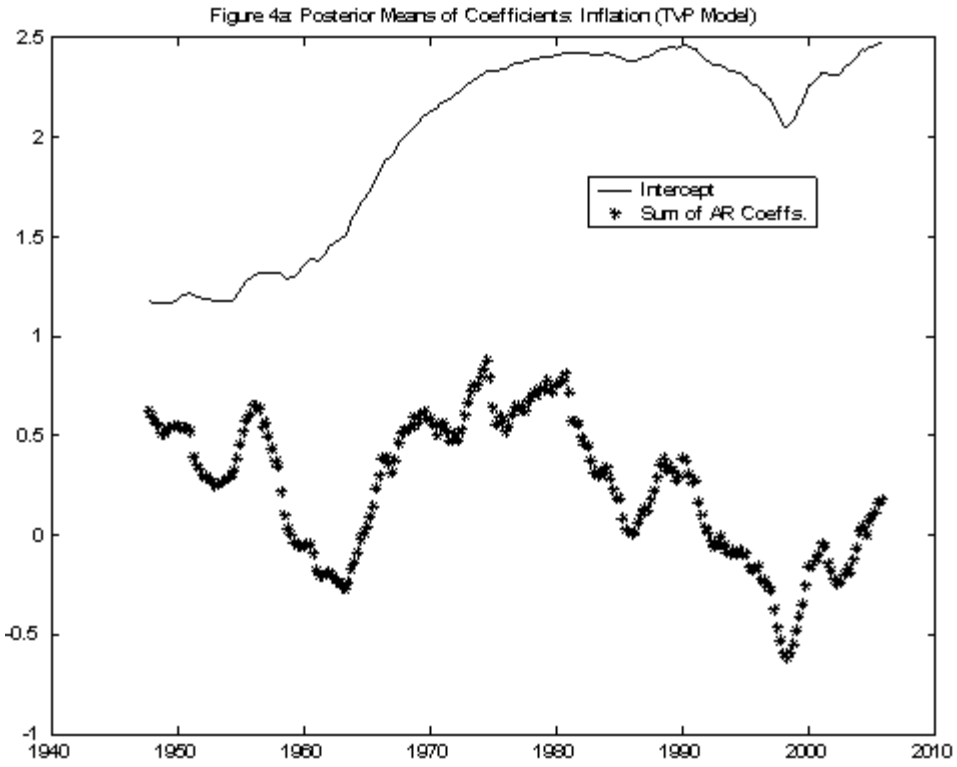


Figure 4a: Posterior Means of Coefficients: Inflation (TVP Model)
Intercept
Sum of AR Coeffs.
2.5
2
1.5
1
0.5
0
-0.5
-1
1940
1950
1960
1970
1980
1990
2000
2010

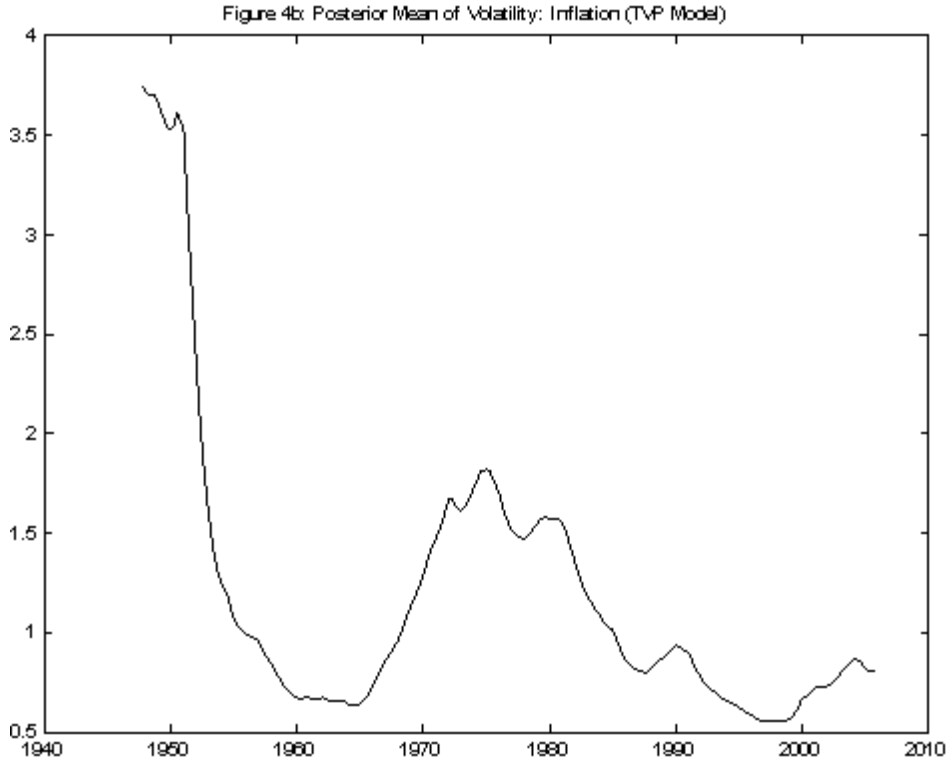
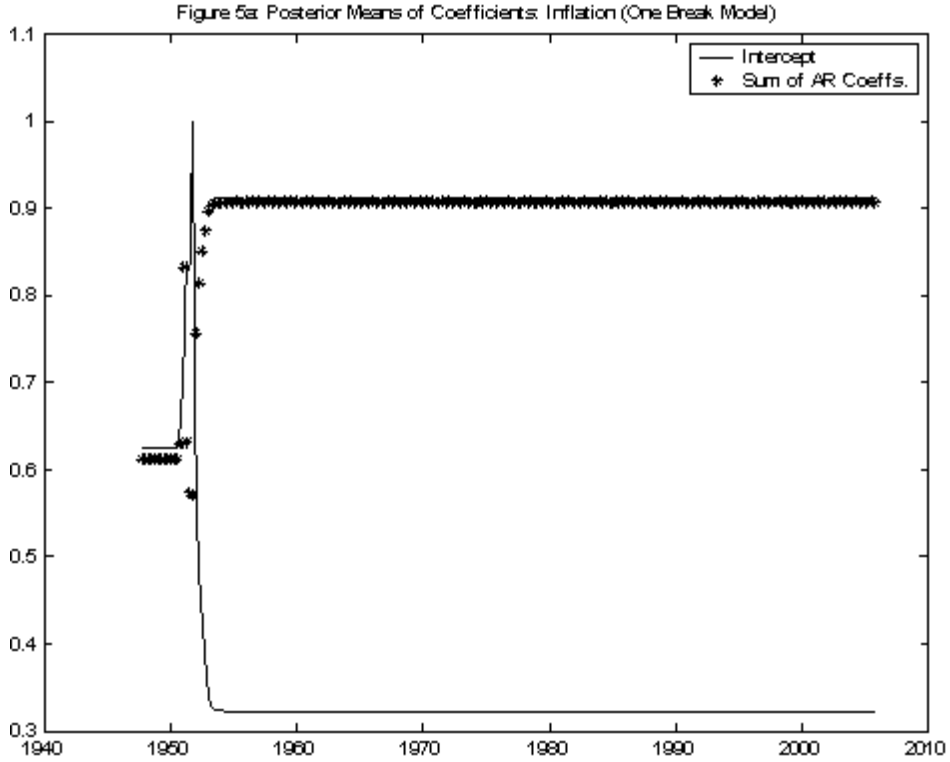


Figure 4b: Posterior Mean of Volatility: Inflation (TVP Model)
4
3.5
3
2.5
2
1.5
1
0.5
1940
1950
1960
1970
1980
1990
2000
2010

Figure 5a: Posterior Means of Coefficients: Inflation (One Break Model)



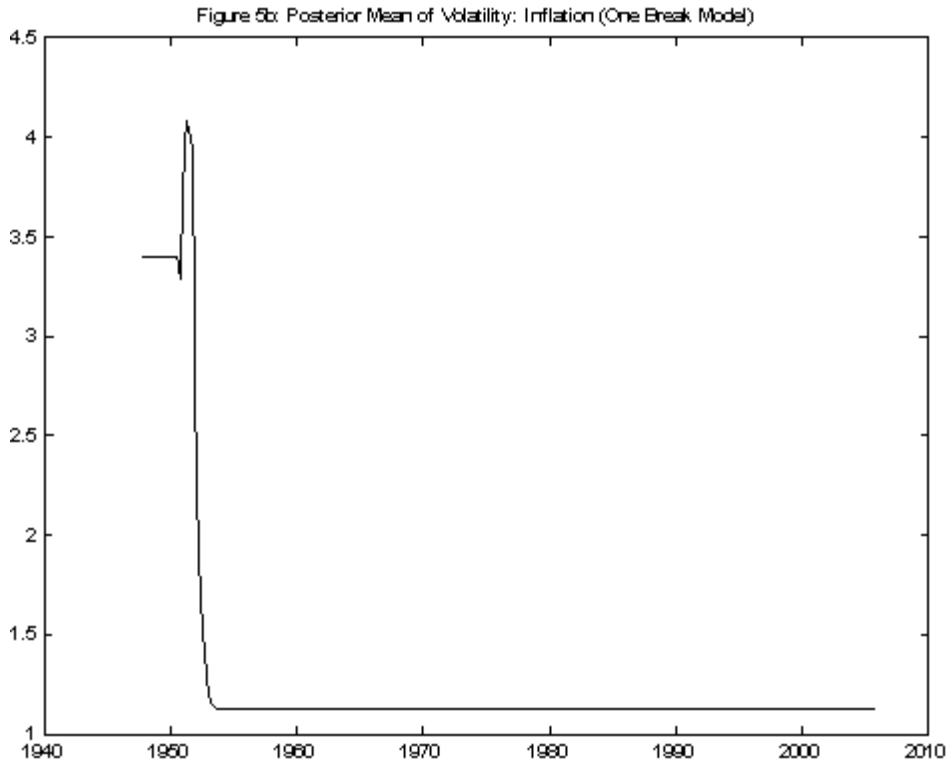


Figure 5b: Posterior Mean of Volatility: Inflation (One Break Model)
4.5
4
3.5
3
2.5
2
1.5
1
1940
1950
1960
1970
1980
1990
2000
2010

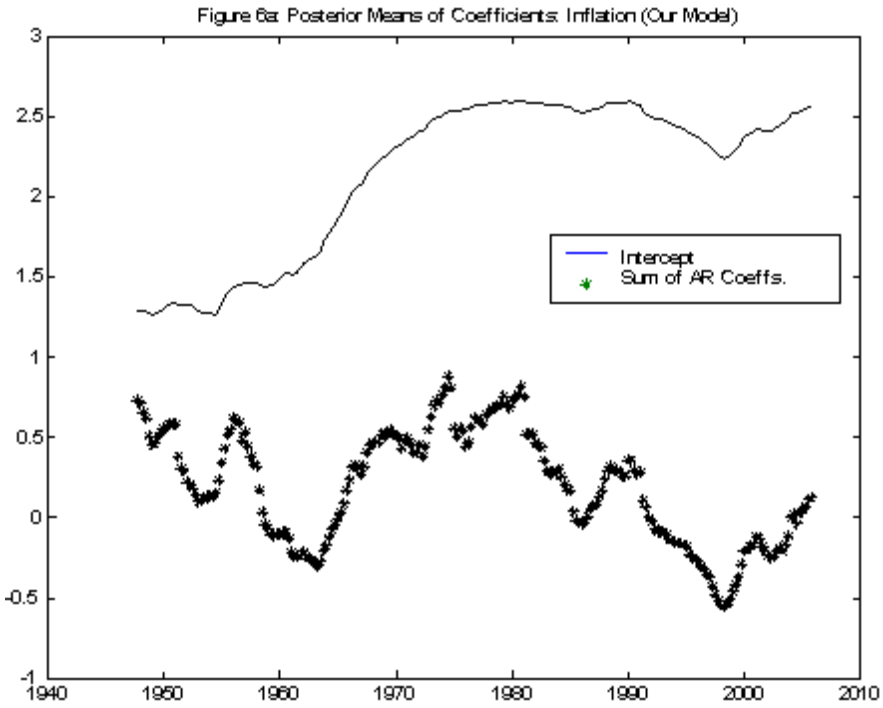


Figure 6a: Posterior Means of Coefficients: Inflation (Our Model)
Intercept
Sum of AR Coeffs.
3
2.5
2
1.5
1
0.5
0
-0.5
-1
1940
1950
1960
1970
1980
1990
2000
2010

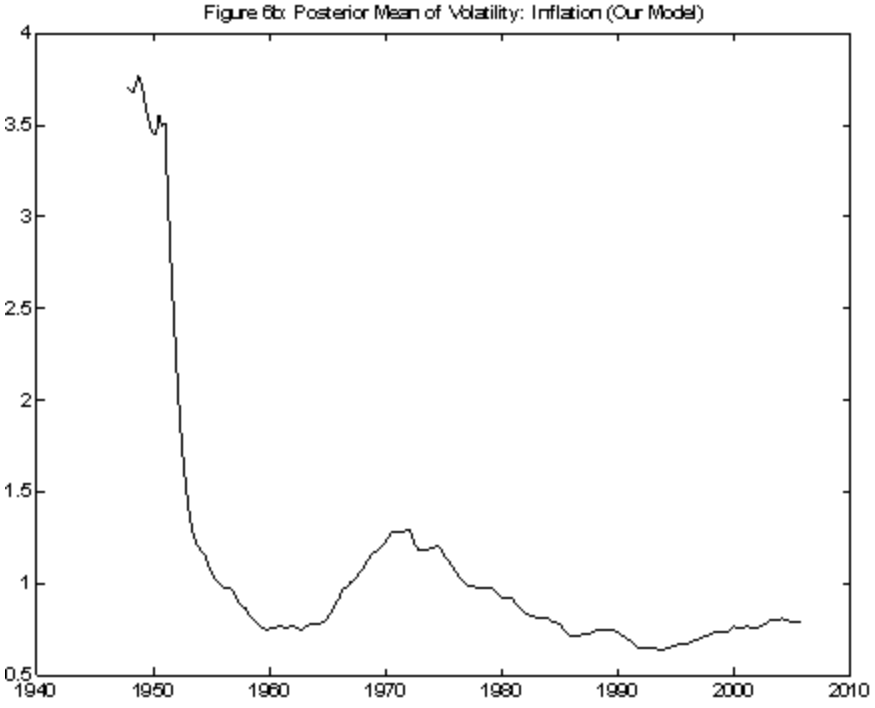


Figure 6b: Posterior Mean of Volatility: Inflation (Our Model)
4
3.5
3
2.5
2
1.5
1
0.5
1940
1950
1960
1970
1980
1990
2000
2010

Figure A1: Prior for Regime Duration

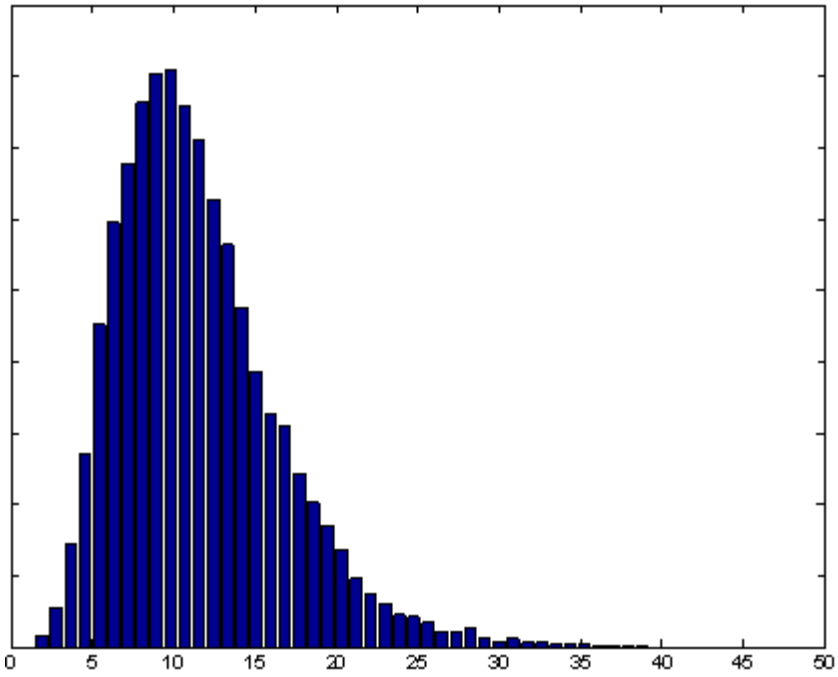


Figure A2: Prior for St. Dev. of S.V. Equation Innovation

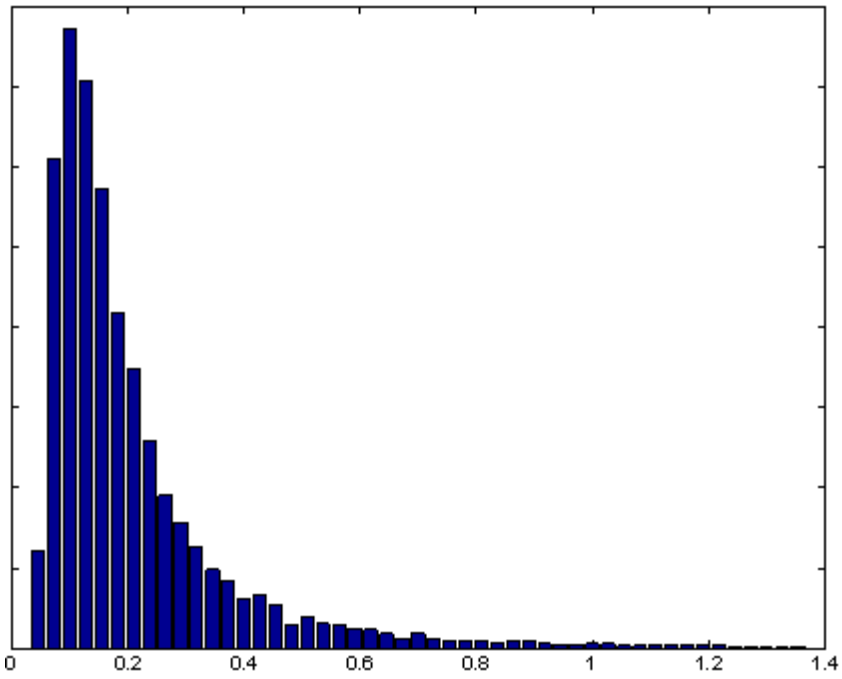


Figure A3: Prior for St. Dev. of AR(1) Equation Innovation