A Time Series Model of Multiple Structural Changes in Level, Trend and Variance

Jiahui WANG, Eric ZIVOT

Jiahui Wang is a Research Scientist, Data Analysis Products Division, MathSoft, Inc., Seattle, WA 98109. Eric Zivot is an Assistant Professor, Department of Economics, University of Washington, Seattle, WA 98195.

We consider a deterministically trending dynamic time series model in which multiple structural changes in level, trend and error variance are modeled explicitly and the number but not the timing of the changes are known. Estimation of the model is made possible by the use of the Gibbs sampler. The determination of the number of structural breaks and the form of structural change is considered as a problem of model selection and we compare the use of marginal likelihoods, posterior odds ratios and Schwarz' BIC model selection criterion to select the most appropriate model from the data. We evaluate the efficacy of the Bayesian approach using a small Monte Carlo experiment. As empirical examples, we investigate structural changes in the U.S. ex-post real interest rate and in a long time series of U.S. real GDP.

KEY WORDS: Multiple structural changes, Gibbs sampler, posterior odds ratio, BIC.

1. INTRODUCTION

Since the publication of the influential papers by Rappoport and Reichlin (1989) and Perron (1989), which provide evidence that many macroeconomic time series might best be modeled as stationary around a broken trend, the detection of structural change in the trend function of a time series has captured the attention of econometricians and applied researchers. Much of the subsequent research has focused on testing the unit root hypothesis in the presence of one time structural change where the date of structural change may or may not be known. Contributions in this area include Christiano (1992), Banerjee, Lumsdaine and Stock (1992), Zivot and Andrews (1992), Perron (1997) and Perron and Vogelsang (1998). Perron (1994) and Maddala and Kim (1996a) provide useful summaries. In related work, Vogelsang (1997) develops tests for a change in trend that are robust to whether the data are $I(0)$ or $I(1)$ thereby extending the results of Andrews (1993) to some models with trending data. Empirically, the unit root hypothesis has been rejected in favor of a broken trend model with one change for numerous series. Most notably, using various techniques and tests, the unit root hypothesis has been rejected for many international output series by Banerjee et al. (1992), Raj (1992), Perron (1992), De Haan and Zelhorst (1993), Zelhorst and De Haan (1995), Cheung and Chinn (1996), Ben-David and Papell (1995) and Perron (1997). Unit roots have been rejected in favor of a single trend break model for several inflation series by Evans and Lewis (1995) and Culver and Papell (1997), for real exchange rates by Edison and Fisher (1991), Perron and Vogelsang (1992) and Culver and Papell (1995), and for real interest rates by Perron (1990). Overall, there is a large body of evidence to suggest that the trend function of many macroeconomic time series can be modeled as deterministic with at least one structural change.

A natural extension of the literature on testing for unit roots in the presence of structural change involves allowing for more than one possible break date under the alternative broken trend stationary model. Indeed, for many macroeconomic time series for which the possibility of structural change is entertained the assumption of at most one break date is unrealistic and restrictive. For example, trend breaks are often motivated by "big events" like wars, oil price shocks, financial crisis or changes in political or institutional regimes and most long time series contain several such events. To this end, Lumsdaine and Papell (1997) extend the Zivot-Andrews (1992) testing procedure to allow for up to two possible endogenous breaks and they find more evidence against the unit root hypothesis than Zivot and Andrews, but less than Perron (1997). Ben-David, Lumsdaine and Papell (1997) find further evidence for at least two structural breaks for three quarters of the per capita real GDP series collected by Maddison (1995). In addition, Papell (1998) finds evidence of multiple breaks in numerous European real exchange rates, Kanas (1998) finds evidence for up to six breaks in ERM exchange rates and Garcia and Perron (1996) find evidence for two breaks in U.S. real interest rates. Indeed, there is a growing body of results that support trend stationary models with multiple breaks for many macroeconomic and financial time series.

In addition to changes in level and trend, changes in variance are often found in economic and financial data. For example, Schwert (1990) finds that the stock market volatility is higher during and after the 1987 crash, compared with other periods. Inclán (1993), Inclán and Tiao (1994) and Chen and Gupta (1997) detect multiple changes in variance for various series of stock returns. Lamoureux and Lastrapes (1990) suggest that the empirical persistence of volatility captured by GARCH models might be caused by structural changes in variance and this view has been supported by Wilson, Aggarwal and Inclán (1996) and Fong (1997). Engel and Hakkio (1996) find that EMS exchange rates have higher volatility dur-

???

ing the periods of alignment and Kim and Engel (1997) find multiple changes in variance in real exchange rates associated with historically significant monetary events. Finally, Kim and Nelson (1998) find evidence of variance changes in postwar business cycles.

Much of the evidence for multiple structural breaks in macroeconomic and financial time series has come from the results of unit root tests that allow for structural change, statistical tests for a single break or tests for parameter constancy and not from statistical techniques that are designed to estimate multiple break models. If interest is in determining the number and / or type of breaks, efficiently estimating break dates or constructing confidence intervals for specific breaks then methods specifically designed for estimating and testing models with multiple structural change are required. Broadly, multiple structural change models can be categorized into time varying parameter models, exogenous multiple break models, endogenous switching models (e.g., Markov switching models or self exciting threshold models) and outlier models. In this paper we focus on a model of exogenous multiple breaks in level, trend and variance of a dynamic time series.

There is a large existing literature on models that allow for multiple structural changes and for classical and Bayesian methods for analyzing these models. Some recent surveys are given by Zacks (1983), Broemeling and Tsurumi (1987), Krishnaiah and Miao (1988), Maddala and Kim (1996b) and Csörgó and Horváth (1997). Most of the methods described in these papers, however, do not apply to dynamic time series models.

Recently, general consistency and asymptotic distribution results have been derived for the classical estimation of dynamic linear models with multiple exogenous breaks using the least squares (LS) principle by Bai (1997a, 1997b) and Bai and Perron (1998)[1]. These methods, while quite general, are not straightforward to apply in practice and the bulk of the results are not applicable to trending data without substantial modification. In addition, their methods ignore possible structural change in the error variance. Instead of modifying the classical methods of Bai et al. (1998) for trending data and variance changes, we take a different approach based on Bayesian methods. The advantage of the Bayesian approach to the analysis of structural change models has long been acknowledged. For example, Raftery (1994) remarks that Bayesian analysis in the context of structural change models is technically simpler than classical methods, allows finite sample inferences which are optimal given the framework and allows for non-nested model comparisons. Furthermore, inference from the Bayesian approach is the same for non-trending and trending data. Finally, in the context of economic time series data, a researcher may or may not have some prior knowledge about the timing, form and maximum number of structural changes and a Bayesian approach can explicitly take this knowledge or lack of knowledge into consideration.

In this paper we start with a deterministically trending dynamic time series model in which multiple structural changes in level, trend and error variance are modeled explicitly and the number but not the timing of the changes are known. Our model is an extension of the models used by Carlin, Gelfand and Smith (1992), Inclán (1993) and Stephens (1994) to the case of deterministically trending dynamic models with heteroskedasticity. As in Carlin et al. (1992) and Stephens (1994), estimation of the model is made possible by the use of the Gibbs sampler. We consider the determination of the number of structural breaks and the form of the breaks as a problem of model selection and we compare the use of marginal likelihoods, posterior odds ratios and Schwarz' BIC model selection criterion to select the most appropriate model from the data.

The remainder of the paper is organized as follows. In section 2, we present the time series model of multiple structural changes and establish some notation. In section 3, we describe Bayesian inference using the Gibbs sampler in the multiple break model. We present the conditional densities required for the Gibbs sampling algorithm, we review how to obtain estimates of posterior moments of the parameters using the output of the Gibbs sampler and we discuss the issue of model selection. In section 4 we evaluate the efficacy of the Bayesian approach using a small Monte Carlo experiment. In section 5, we give some empirical applications of our methods. We first investigate structural changes in the U.S. ex-post real interest rate and then we consider structural changes in a long time series of U.S. real GDP. Section 6 contains our concluding remarks and suggestions for future work. Regarding notation, we use $f(\cdot)$ to denote a probability density or mass function, $f_0(\cdot)$ to denote a prior density or mass function and $f(\cdot|\cdot)$ to denote a conditional density (mass) function.

2. A TIME SERIES MODEL WITH MULTIPLE STRUCTURAL CHANGES

Our approach to modeling multiple structural changes is based on extending the single break switching regression models used by Ferreira (1977), Chin Choy and Broemeling (1980), Smith (1980), Booth and Smith (1982), Holbert (1982), Broemeling et al. (1987), Zivot and Phillips (1995) and DeJong (1996), to allow for multiple breaks in the regression parameters as well as in the variances of the errors. The model is a segmented deterministically trending and heteroskedastic autoregressive model

$$y_t = a_t + b_t t + \phi_1 y_{t-1} + \cdots + \phi_r y_{t-r} + s_t u_t, \qquad (1)$$

for $t = 1, 2, \ldots, T$, where $u_t|\Omega_t \sim$ *i.i.d.* $N(0, 1)$ and Ω_t denotes the information set at time t. We assume that the parameters a_t, b_t and s_t are subject to $m < T$ structural changes, m initially known, with break dates $k_1, \ldots, k_m$, $1 < k_1 < k_2 < \cdots < k_m \leq T$, so that the observations can be separated into $m + 1$ regimes. Let $\mathbf{k} = (k_1, k_2, \ldots, k_m)'$ denote the vector of break dates. For each regime $i(i =$

$1, 2, \ldots, m+1)$, the parameters a_t, b_t and s_t are given by

$$a_t = \alpha_i, b_t = \beta_i, s_t = \sigma_i \geq 0$$

for $k_{i-1} \leq t < k_i$ with $k_0 = 1$ and $k_{m+1} = T + 1$. The model (1) is a *partial structural change* model since the autoregressive parameters are assumed to be constant across regimes[2].

The model (1) nests many types of multiple structural change models of interest for empirical work using economic time series. For example, if the roots of the autoregressive polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_r z^r = 0$ lie outside the complex unit circle and there are changes only in a_t then (1) reduces to an innovation outlier level shift model; if the roots of $\phi(z) = 0$ lie outside the unit circle and there are only changes in b_t then (1) becomes an innovation outlier broken trend model; and if there are only changes in s_t then (1) becomes a group-wise heteroskedastic model. In spite of the apparent complexity of the model, it can be re-written in the form of a standard linear regression model with group-wise heteroskedasticity. Let I_A denote an indicator variable such that I_A is equal to one if the event A is true and zero otherwise. Then (1) can be re-written as

$$y_t = \sum_{i=1}^{m+1} I_{\{k_{i-1} \leq t < k_i\}}(\alpha_i + \beta_i t) + \sum_{j=1}^{r} y_{t-r}\phi_j + s_t u_t,$$

or, more succinctly, as

$$y_t = \mathbf{x}_t' \mathbf{B} + s_t u_t, \tag{2}$$

where

$$\mathbf{x}_t = \begin{bmatrix} I_{\{k_0 \leq t < k_1\}} \\ \vdots \\ I_{\{k_m \leq t < k_{m+1}\}} \\ t \cdot I_{\{k_0 \leq t < k_1\}} \\ \vdots \\ t \cdot I_{\{k_m \leq t < k_{m+1}\}} \\ y_{t-1} \\ \vdots \\ y_{t-r} \end{bmatrix}$$

and $\mathbf{B} = (\alpha_1, \ldots, \alpha_{m+1}, \beta_1, \ldots, \beta_{m+1}, \phi_1, \ldots, \phi_r)'$. Let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_{m+1})'$ and define $\boldsymbol{\theta} = (\mathbf{B}', \boldsymbol{\sigma}', \mathbf{k}')'$ as the vector of unknown parameters of (2), $\mathbf{Y}_0$ as the vector of r initial values of y_t, and $\mathbf{Y} = (y_1, \ldots, y_T)'$ as the vector of observed data. Given the normality of the errors u_t, the likelihood function of (2) takes the form

$$L(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{Y}_0) \propto (\prod_{t=1}^{T} s_t)^{-1} \exp\{-\frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - \mathbf{x}_t'\mathbf{B})^2}{s_t^2}\} \tag{3}$$

$$= |\mathbf{S}|^{-1} \exp\{-\frac{(\mathbf{Y} - \mathbf{X}\mathbf{B})'\mathbf{S}^{-2}(\mathbf{Y} - \mathbf{X}\mathbf{B})}{2}\}, \tag{4}$$

where $\mathbf{S}$ is a diagonal matrix with $(s_1, \ldots, s_T)$ on the diagonal and $\mathbf{X}$ is a $T \times (2m + 2 + r)$ matrix with t-th row given by $\mathbf{x}_t'$. For notational brevity, hereafter the conditioning on $\mathbf{Y}_0$ will be suppressed.

3. BAYESIAN INFERENCE

In a Bayesian context inference on the unknown parameters $\boldsymbol{\theta}$ is made from the joint posterior distribution of $\boldsymbol{\theta}$, which, by Bayes' theorem, is given by

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta})f_0(\boldsymbol{\theta})}{f(\mathbf{Y})} \propto L(\boldsymbol{\theta}|\mathbf{Y})f_0(\boldsymbol{\theta}).$$

That is, given a prior specification $f_0(\boldsymbol{\theta})$ our knowledge of $\boldsymbol{\theta}$ is updated using information in the likelihood function $L(\boldsymbol{\theta}|\mathbf{Y})$. If we are interested in a specific element of $\boldsymbol{\theta}$, say θ_i, then we require the marginal posterior of θ_i which is found by integrating the joint posterior with respect to the remaining elements of $\boldsymbol{\theta}$. Except for very simple problems with a few parameters the evaluation of the joint posterior and the computation of marginal posteriors is not analytically tractable. Additionally, brute force numerical integration is not a practical option as well. Recent advances in Markov chain Monte Carlo techniques, however, have allowed fast and efficient methods for the evaluation of high dimensional integrals in the context of Bayesian analysis and we utilize these methods for our model.

3.1 Prior Specification

With regard to the specification of the prior for the parameters in the multiple break context, we follow Inclan (1993) and Stephens (1994) and make the following assumptions regarding the general form of the prior $f_0(\boldsymbol{\theta})$. We assume that $\mathbf{k}$, $\mathbf{B}$ and $\boldsymbol{\sigma}$ are mutually independent and that the elements of $\boldsymbol{\sigma}$ are independent. Hence, the joint prior is of the form

$$f_0(\boldsymbol{\theta}) = f_0(\mathbf{k})f_0(\mathbf{B})\prod_{i=1}^{m+1} f_0(\sigma_i^2). \tag{5}$$

We specify proper priors for the blocks of parameters in (5). Regarding the location of the break dates, we adopt a diffuse prior such that $f_0(\mathbf{k})$ is discrete uniform over all ordered subsequences of $(2, 3, \ldots, T)$ of length m. This prior specification was also used by Inclán (1993) and Stephens (1994) and is common in the multiple break literature. For the regression parameters $\mathbf{B}$, we use the natural conjugate multivariate normal prior $N(\mathbf{B}_0, \Sigma_\mathbf{B})$, where $\mathbf{B}_0$ denotes the prior mean and $\Sigma_\mathbf{B}$ denotes the prior covariance matrix. Prior ignorance about the values of $\mathbf{B}$, for example, can be captured by specifying $\Sigma_\mathbf{B}$ as a diagonal matrix with large diagonal elements. Similarly, we use the natural conjugate inverted gamma prior $IG(v_0, \lambda_0)$ for the regime specific variances σ_i^2:

$$f_0(\sigma_i^2) \propto (\sigma_i^2)^{-(v_0-1)} \exp\left\{-\frac{\lambda_0}{\sigma_i^2}\right\}, \; i = 1, \ldots, m+1.$$

3.2 The Gibbs Sampler Algorithm

Given our prior specifications for the elements of $\boldsymbol{\theta}$ and the likelihood function (4), the posterior distribution of $\boldsymbol{\theta}$

$$f(\boldsymbol{\theta}|\mathbf{Y}) \propto f_0(\mathbf{k})f_0(\mathbf{B})\prod_{i=1}^{m+1} f_0(\sigma_i^2)L(\boldsymbol{\theta}|\mathbf{Y}) \qquad (6)$$

is not of a standard form and analysis based on numerical or Monte Carlo integration techniques is extremely difficult. However, sample draws from this posterior can be generated in a straightforward manner using the Gibbs sampler[3]. The idea of the Gibbs sampler is simple. Although the joint posterior distribution of the multivariate random vector $\boldsymbol{\theta}$, $f(\boldsymbol{\theta}|\mathbf{Y})$, may be unknown, samples of $\boldsymbol{\theta}$ can still be generated if the draws from the full conditional posterior distributions $f(\theta_i|\mathbf{Y},[\boldsymbol{\theta}-\theta_i])$, where $[\boldsymbol{\theta}-\theta_i]$ denotes the elements of the vector $\boldsymbol{\theta}$ except θ_i, can be easily generated. By iteratively generating samples from the complete set of conditional distributions, the draws of the random variables from these distributions form a Markov Chain and, under mild conditions, will converge to draws from the joint posterior distribution $f(\boldsymbol{\theta}|\mathbf{Y})$. In addition, the draws of a particular element of $\boldsymbol{\theta}$ will converge to draws from the marginal posterior distribution of that element[4]. It turns out that for the multiple break model we consider the conditional posterior distributions necessary for the Gibbs sampler algorithm are all of standard form and samples are easy to generate. Our algorithm is similar to Stephens (1994) method modified to handle lagged dependent data, partial structural change and group-wise heteroskedasticity.

We now specify the conditional posteriors required for the Gibbs sampling algorithm. Let n_i denote the number of observations in regime i, $\mathbf{Y}^i$ denote the $n_i \times 1$ vector of y_t values in regime i and $\mathbf{X}^i$ denote the $n_i \times (2m+2+r)$ matrix of $\mathbf{x}_t'$ values in regime i. Consider first the conditional posterior of k_i, $i=1,\ldots,m$. Given that $1=k_0<\cdots<k_{i-1}<k_i<k_{i+1}<\cdots<k_{m+1}=T$ and the form of the joint prior (5), the sample space of the conditional posterior of k_i only depends on the neighboring break dates k_{i-1} and k_{i+1}. It follows that for $k_i \in [k_{i-1},k_{i+1}]$:

$$f(k_i|\mathbf{Y},[\boldsymbol{\theta}-k_i]) \propto f(k_i|\mathbf{Y},k_{i-1},k_{i+1},\mathbf{B},\boldsymbol{\sigma}), \qquad (7)$$

for $i=1,\ldots,m$, which is proportional to the likelihood function for $\boldsymbol{\theta}$ evaluated with a break at k_i only using data between k_{i-1} and k_{i+1}. The probabilities associated with each $k_i \in [k_{i-1},k_{i+1}]$ are easily computed and, hence, sample draws of k_i from $f(k_i|\mathbf{Y},[\boldsymbol{\theta}-k_i])$ can be generated as multinomial random variables with number of bins equal to the number of break dates between k_{i-1} and k_{i+1} and probabilities proportional to the likelihood function.

Next, consider the conditional posterior of $\mathbf{B}$. Given the normal prior for $\mathbf{B}$ and the form of (4), it is readily verified that

$$\mathbf{B}|\mathbf{Y},[\boldsymbol{\theta}-\mathbf{B}] \sim N(\Psi_{\mathbf{B}}(\Sigma_{\mathbf{B}}^{-1}\mathbf{B}_0+\mathbf{X}'\mathbf{S}^{-2}\mathbf{Y}),\Psi_{\mathbf{B}}), \quad (8)$$

where $\Psi_{\mathbf{B}}=(\Sigma_{\mathbf{B}}^{-1}+\mathbf{X}'\mathbf{S}^{-2}\mathbf{X})^{-1}$. Thus sample draws of $\mathbf{B}$ can be generated as a multivariate normal random vector.

Finally, consider the conditional posterior σ_i^2, $i=1,\ldots,m$. Given the inverted gamma prior for σ_i^2 and the form of (3), it can be verified that

$$\sigma_i^2|\mathbf{Y},[\boldsymbol{\theta}-\sigma_i^2] \sim IG(v_i,\lambda_i),\ i=1,\ldots,m \qquad (9)$$

where $v_i=v_0+n_i/2$, and $\lambda_i=\lambda_0+(\mathbf{Y}^i-\mathbf{X}^i\mathbf{B})'(\mathbf{Y}^i-\mathbf{X}^i\mathbf{B})/2$. Thus sample draws of σ_i^2 can be generated as inverted gamma random variables.

Given the full set of conditional distributions (7), (8) and (9), we now summarize the Gibbs sampling algorithm for generating sample draws from the joint posterior (6)[5]:

- Step 1: Specify starting values $\boldsymbol{\theta}^{(0)}$ and set the iteration number $j=1$, where
$$\boldsymbol{\theta}^{(0)}=(\mathbf{k}^{(0)'},\mathbf{B}^{(0)'},\boldsymbol{\sigma}^{(0)'})'.$$
- Step 2: Generate a draw for the first break date k_1 as a multinomial random variable on the sample space $[k_0^{(j-1)},k_2^{(j-1)}]$ from the conditional posterior $f(k_1^{(j)}|\mathbf{Y},k_0^{(j-1)},k_2^{(j-1)},\mathbf{B}^{(j-1)},\boldsymbol{\sigma}^{(j-1)})$.
- Step $i(i=3,\ldots,m+1)$: Generate a draw of the $(i-1)$-th break date $k_{i-1}^{(j)}$ from the conditional posterior $f(k_{i-1}^{(j)}|\mathbf{Y},k_{i-2}^{(j-1)},k_i^{(j-1)},\mathbf{B}^{(j-1)},\boldsymbol{\sigma}^{(j-1)})$ as a multinomial random variable on the sample space $[k_{i-2}^{(j-1)},k_i^{(j-1)}]$.
- Step $m+2$: Generate a draw of $\mathbf{B}^{(j)}$ from the conditional posterior $f(\mathbf{B}^{(j)}|\mathbf{Y},\mathbf{k}^{(j)},\boldsymbol{\sigma}^{(j-1)})$ given in (8) as a multivariate normal random vector.
- Step $m+3$: Generate a draw of $(\sigma_i^2)^{(j)}$ from the conditional posterior $f((\sigma_i^2)^{(j)}|\mathbf{Y},\mathbf{k}^{(j)},\mathbf{B}^{(j)})$ given in (9) as an inverted gamma random variable.
- Step $m+4$: Set $j=j+1$ and go to step 2.

There does not appear to be an agreed upon method to determine convergence of the Gibbs sampler[6]. A common informal graphical technique involves plotting $\theta_i^{(j)}$ or some function of $\theta_i^{(j)}$. Convergence is indicated if the trajectory exhibits the same qualitative behavior through iterations after an initial burn-in period. Similarly, the trajectory of ergodic averages can be evaluated and plotted and an asymptotic behavior over many successive iterations suggests convergence. For the examples in the Monte Carlo study in the next section, informal graphical diagnostics suggest convergence is achieved after an initial burn-in of about 300 iterations.

We note that convergence of the Gibbs sampler is improved if highly correlated parameters can be sampled jointly. In the above algorithm, the break dates k_i, $i=1,\ldots,m$, are likely to be highly correlated because the conditional posterior of k_i depends on k_{i-1} and k_{i+1}. This suggests that monitoring the autocorrelations in the Gibbs draws of the break dates may be helpful for ascertaining convergence of the sampler. In our algorithm, the k_i are sampled individually for computational convenience and this approach works well for our examples.

It is possible to sample the break dates jointly but the number of computations required for the Gibbs sampling algorithm quickly becomes prohibitive as the number of break dates increases.

3.3 Estimation of Posterior Moments

In practice there are two main ways to obtain N sample draws from the joint posterior $f(\boldsymbol{\theta}|\mathbf{Y})$ based on output from the Gibbs sampler. This first method, suggested by Gelfand and Smith (1990), is to save a sample draw after the convergence of the sampler has occurred, say after l iterations, then restart the sampler and obtain another draw after convergence is achieved. This process is repeated a total of N times to obtain N sample draws from $f(\boldsymbol{\theta}|\mathbf{Y})$ and lN total iterations are required. Draws obtained this way are independent if the chains are initialized independently. The second method, advocated by Geyer (1992), utilizes one long iteration of the Gibbs sampler. With this method, the sampler is iterated N more times after convergence for a total of $l+N$ iterations. The N samples drawn this way are correlated but are stationary and ergodic[7]. We use the second method for computational savings.

With a single iteration of the Gibbs sampler, under the conditions of convergence, the N sample draws of any element θ_i of $\boldsymbol{\theta}$ are correlated observations from a stationary and ergodic process whose distribution is given by the posterior marginal distribution of θ_i. Let $t(\theta_i)$ denote a real valued function of θ_i that is integrable with respect to the posterior marginal distribution of θ_i. Then, given the post-convergent N sample draws $(\theta_i^{(1)}, \theta_i^{(2)}, \ldots, \theta_i^{(N)})$ from the Gibbs sampler, by the ergodic theorem the posterior mean of $t(\theta_i)$ can be consistently estimated by

$$E[\widehat{t(\theta_i)}] = \frac{1}{N}\sum_{j=1}^{N} t(\theta_i^{(j)}). \qquad (10)$$

Further, if $t(\theta_i)^2$ is integrable with respect to the posterior distribution of θ_i then the asymptotic variance of $\sqrt{N}(E[\widehat{t(\theta_i)}] - E[t(\theta_i)])$ can be consistently estimated using the Newey-West estimator

$$\gamma_{i0} + 2\sum_{j=1}^{M_N}(1 - \frac{j}{M_N+1})\gamma_{ij}, \qquad (11)$$

where γ_{ij} is the j-th order sample autocovariance of $t(\theta_i)$ from the N Gibbs draws and M_N is a truncation parameter such that as $N \to \infty$, $M_N \to \infty$ and $M_N/N \to 0$. Typically, $M_N = 4\,(N/100)^{1/4}$.

Table 1. Evaluating Bayes Factors

$2\ln B_{ij}$	B_{ij}	*Evidence against H_0*
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
>10	>150	Very Strong

3.4 Model Selection

In a Bayesian context, competing hypotheses or models are compared using the posterior probabilities of the models[8]. An advantage of using posterior probabilities is that the competing models need not be nested. Let M_i and M_j denote two competing models and let $\Pr(M_i)$ and $\Pr(M_j)$ denote the prior probabilities associated with these models. The comparison of the competing models M_i and M_j after observing the data $\mathbf{Y}$ is summarized by the posterior odds ratio

$$POR_{ij} = \frac{\Pr(M_i|\mathbf{Y})}{\Pr(M_j|\mathbf{Y})} = \frac{\Pr(M_i)}{\Pr(M_j)} \cdot \frac{f(\mathbf{Y}|M_i)}{f(\mathbf{Y}|M_j)},$$

where $f(\mathbf{Y}|M_i)$ and $f(\mathbf{Y}|M_j)$ denote the marginal likelihoods of models i and j, respectively. The first term in POR_{ij} is the prior odds ratio and the second term is the Bayes factor, B_{ij}, or ratio of marginal likelihoods. If each hypothesis is deemed equally likely then the prior odds is unity and POR_{ij} reduces to the Bayes factor.

Regardless of the value of the prior odds of model M_i versus M_j, the Bayes factor is the ratio of the posterior odds to its prior odds and so provides useful evidence for evaluating the plausibility of one model versus another. Kass and Raftery (1995) provide a rule of thumb for interpreting the magnitude of a Bayes factor using the transformation $2\ln B_{ij}$ to put it on the same scale as the likelihood ratio. This rule of thumb is reproduced in Table 1.

The computation of the marginal likelihoods under competing models is required for the construction of Bayes factors and posterior odds and proper priors are required for the parameters of the competing model to avoid ambiguities. Let $\boldsymbol{\theta}$ denote the parameter vector associated with a particular model M and $f_0(\boldsymbol{\theta})$ the prior over these parameters. Then the marginal likelihood is defined as

$$f(\mathbf{Y}|M) = \int_{\Theta} L(\boldsymbol{\theta}|\mathbf{Y}) f_0(\boldsymbol{\theta}) d\,\boldsymbol{\theta},$$

which is just the integrating constant for the joint posterior under model M. Given that the prior $f_0(\boldsymbol{\theta})$ is proper, the marginal likelihood can be interpreted as the expectation of the likelihood with respect to the prior; i.e., $f(\mathbf{Y}|M) = E_{f_0}[L(\boldsymbol{\theta}|\mathbf{Y})]$. In general, this expectation can be consistently estimated using Monte Carlo integration with importance sampling

$$f(\mathbf{Y}|M) = E_{f_0}[L(\boldsymbol{\theta}|\mathbf{Y})] = \frac{1}{N}\frac{\sum_{k=1}^{N} L(\boldsymbol{\theta}|\mathbf{Y}) w^k}{\sum_{k=1}^{N} w^k}, \qquad (12)$$

where $w^k = f_0(\boldsymbol{\theta}^k)/q(\boldsymbol{\theta}^k), k = 1, \ldots, N$ are the importance weights and $q(\boldsymbol{\theta})$ is the importance function, see Gelfand and Dey (1994). Newton and Raftery (1994) suggest using the posterior density $f(\boldsymbol{\theta}|\mathbf{Y})$ as the importance function since samples from $f(\boldsymbol{\theta}|\mathbf{Y})$ arise directly from the Gibbs sampler. In this case, (12) simplifies consider-

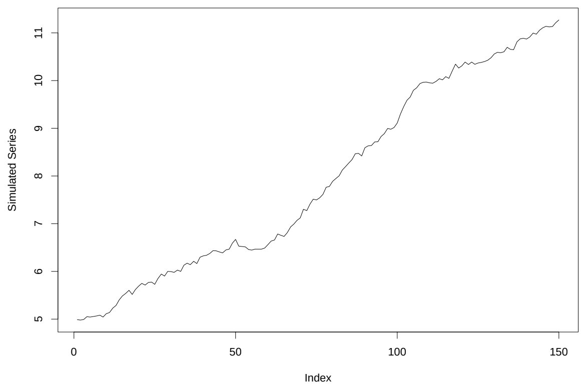

Figure 1. A Simulated Series from Design I

ably and reduces to the harmonic mean of the likelihood

$$\left[\frac{1}{N}\sum_{k=1}^{N}\frac{1}{L(\boldsymbol{\theta}^{(k)}|\mathbf{Y})}\right]^{-1},$$

where $\boldsymbol{\theta}^{(k)}$, $k = 1, \ldots, N$, are sample draws from the Gibbs sampler[9].

3.5 Determining the number of structural breaks

The discussion of the multiple break model thus far has assumed that the number of breaks is known to be m. In practice the number of breaks is generally not known. In the case of an unknown number of breaks the determination of the number of breaks can be treated as a model selection problem and model choice can be made using Bayes factors or posterior odds.

Inclán (1993) considers the use of posterior odds to determine the number of changes in variance of a time series and we consider an extension of her methodology to our setup. Let M_i denote the model with $m = i$ breaks and let $\Pr(M_i)$ denote the prior probability. Inclán determines the priors over the different models by supposing that at each point in time the probability of observing a break is described by an independent Bernoulli process with probability $p \in [0, 1]$ of observing a break[10]. Then the total number of breaks, m, follows a binomial$(T-1, p)$ distribution from which it can be deduced that $\Pr(M_i) = {}_{T-1}C_i \cdot p^i(1-p)^{T-1-i}$. Accordingly, the prior odds for a model with i breaks versus a model with j breaks $(i > j)$ is given by

$$\frac{\Pr(M_i)}{\Pr(M_j)} = \left(\frac{p}{1-p}\right)^{i-j}\frac{j!\cdot(T-1-j)!}{i!\cdot(T-1-i)!}.$$

If it is thought *a priori* that breaks are not very likely to occur then a sensible choice is to set $p = 1/T$. In this case, the prior odds of one break versus zero breaks is unity.

Inclán also considers treating p as a hyperparameter of the model and specifies a beta prior with parameters a and b. With this specification, the prior odds for a model with j breaks versus a model with $j-1$ breaks is shown in the appendix to be[11]

$$\frac{\Pr(M_j)}{\Pr(M_{j-1})} = \frac{T-j+1}{j}\frac{a+j-1}{b+T-j}.$$

If it is thought that p takes a small value *a priori* then $a = 1$ or 2 and $b = T$ are sensible prior parameters.

An alternative approach to determine the number of breaks is to follow Yao (1988) and Liu, Wu and Zidek (1997) and use the Schwarz Bayesian Information Criterion (BIC) defined as

$$BIC(m) = -2\ln L(\widehat{\boldsymbol{\theta}}_{MLE}|\mathbf{Y}) + q\ln(T),$$

where $L(\widehat{\boldsymbol{\theta}}_{MLE}|\mathbf{Y})$ denotes the likelihood function (3) evaluated at $\widehat{\boldsymbol{\theta}}_{MLE}$ and q denotes the total number of estimated parameters in a model with m breaks[12]. The Schwarz criterion indicates that the model with the highest posterior probability is the one that minimizes the BIC and $BIC(j) - BIC(i)$ can be viewed as a rough approximation to $2\ln B_{ij}$, see Kass et al. (1995). Yao (1988) has shown that minimizing $BIC(m)$ is a consistent criterion for determining the number of structural changes in a normal sequence of random variables with an unknown number of shifts in mean and Liu et al. (1997) extend his result to segmented regression models with exogenous regressors. Instead of using $\widehat{\theta}_{MLE}$, we compute $BIC(m)$ using the posterior modes of k_i for $i = 1, \ldots, m$ and the posterior means of the remaining parameters based on the output of the Gibbs sampler.

4. APPLICATION TO SIMULATED DATA

To examine the performance of our Gibbs sampling algorithm we conduct a small Monte Carlo study of the approach. We consider two designs. Design I represents a structural change model in the trend, while design II represents a structural change model in the mean and variance. In both designs, there are two breaks. These two types of models are common in empirical analyses of structural change in macroeconomic and financial time series. For example, modeling of trending series like GDP or per capita GDP typically follows design I as in Ben-David et al. (1995) and Ben-David et al. (1997), while modeling of exchange rates and interest rates usually follows design II as in Perron et al. (1992) and Garcia et al. (1996).

In the Monte Carlo study, we concentrate on two aspects of the Bayesian approach: (1) the model selection issue when the number of structural changes is unknown; and (2) the estimation performance when the number of structural changes is correctly specified. let M_j denote the model with j breaks. For model selection, we compare the performance of four criteria: (i) marginal likelihoods using the decision rules in Table 1; (ii) POR with $p = 2/T$ (POR1); (iii) POR with $p \sim Beta(2, T)$ (POR2); (iv) BIC. For estimation when the number of structural changes is correctly specified, we report estimates from a single Gibbs sampler run as well as Monte Carlo means of the estimates.

Table 2. Design I – Model Selection

Criterion	$m=0$	$m=1$	$m=2$	$m=3$
POR1	0%	0%	56%	44%
POR2	0%	0%	56%	44%
BIC	2%	0%	95%	3%

4.1 Design I: Structural Changes in Trend

In design I, the data are generated according to

$$y_t = a_t + b_t t + \phi y_{t-1} + \sigma u_t,\ t = 1, 2, \ldots, 150, \quad (13)$$

where the notation follows equation (1) with $\phi = 0.7$, $\sigma = 0.05$, and

$$\begin{aligned} a_t &= \alpha_1 = 1.5,\ b_t = \beta_1 = 0.01, && \text{for} \quad 0 < t \leq 50, \\ a_t &= \alpha_2 = 0.8,\ b_t = \beta_2 = 0.02, && \text{for} \quad 50 < t \leq 100, \\ a_t &= \alpha_3 = 1.9,\ b_t = \beta_3 = 0.01, && \text{for} \quad 101 < t \leq 150, \end{aligned}$$

so that there are two structural changes in the trend function with $\mathbf{k} = (51, 101)'$ and no changes in variance. This data generating process (DGP) is intended to mimic the behavior of GDP series in many industrial countries.

Figure 1 gives a simulated series from this DGP. Notice that even though the series exhibits two possible structural changes, it is not completely clear when the changes exactly occurred.

The Gibbs sampling algorithm presented in section 3 is employed for the estimation of models with no change in variance for $m =$ 0, 1, 2, and 3 break points. To represent our prior ignorance over the parameters of the DGP, diffuse priors are used for $\mathbf{B} = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \phi)'$ and σ such that $\mathbf{B}_0 = 0$, $v_0 = 1.001$, $\lambda_0 = 0.001$, and $\Sigma_\mathbf{B}$ is set to a diagonal matrix with elements 1000 on the diagonal. For a model with m structural change(s), the starting value of $\mathbf{k}$ is set at the m (approximate) equi-distant points between 1 and 150. Then the starting values of $\mathbf{B}$ and σ^2 are computed as in a standard linear model. After running the Gibbs sampler for 300 iterations, we save the next 2000 draws for inference. Finally this procedure is replicated 100 times.

Out of the 100 replications, the percentage of each model being chosen by a certain criterion is recorded in Table 2. Using Kass and Raftery's rules for evaluating Bayes factors, there is always "very strong" evidence in favor of M_2 against M_1 and M_0. Comparing M_3 against M_2 there is "strong evidence" in favor of M_3 17% of the time and there is "very strong" evidence in favor of M_3 only 5% of the time. The two PORs choose the right number of structural changes, where M_i is chosen if $POR_{ij} > 1$, only a little more than 50% of the time, while BIC chooses the right number 95% of the time.

For M_2, Table 3 reports the posterior mean estimates (under column "Mean") computed using (10) and standard deviations of the estimates (under column "SD") computed using (11) for the series plotted in Figure 1, together with Monte Carlo means of the estimates and Monte Carlo standard deviations for 100 replications. The estimates for the series in Figure 1 are generally close to the Monte Carlo statistics and to the true values. In particular, the estimated standard deviations are very close to Monte Carlo standard deviations.

Table 3. Design I – Estimation Results when $m = 2$

Parameter	*Mean*	*SD*	*MC Mean*	*MC SD*
α_1	1.370	0.224	1.645	0.231
α_2	0.771	0.145	0.877	0.147
α_3	1.826	0.305	2.082	0.292
β_1	0.010	0.002	0.011	0.002
β_2	0.018	0.003	0.022	0.003
β_3	0.008	0.002	0.011	0.002
ϕ	0.726	0.046	0.671	0.048
σ	0.051	0.003	0.050	0.003

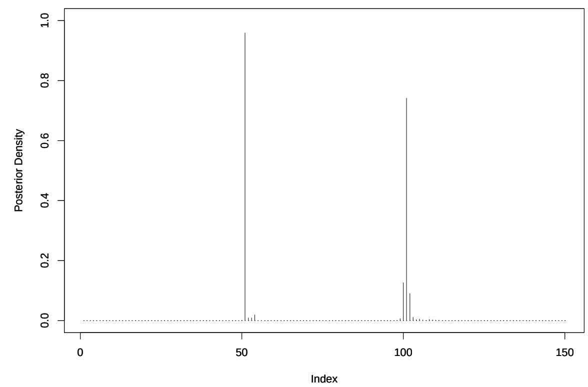

Figure 2. Estimated Changepoints – An Example from Design I

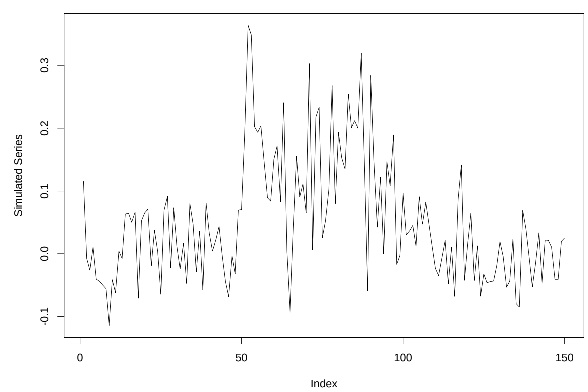

Figure 3. A Simulated Series from Design II

The Gibbs sampler also produces the posterior mass functions for each estimated change point. For the series in Figure 1, the posterior mass functions of the change points for M_2 is plotted in Figure 2. The posterior mass functions of each change point has a mode at the true break date.

4.2 Design II: Structural Changes in Mean and Variance

In design II, the data are generated according to

$$y_t = a_t + \phi y_{t-1} + s_t u_t,\ \text{for } t = 1, 2, \ldots, 150, \quad (14)$$

Table 4. Design II – Model Selection

Criterion	$m = 0$	$m = 1$	$m = 2$	$m = 3$
POR1	0%	0%	38%	62%
POR2	0%	0%	38%	62%
BIC	0%	0%	98%	2%

Table 5. Design II – Estimation Results when $m = 2$

Parameter	Mean	SD	MC Mean	MC SD
α_1	0.008	0.008	0.000	0.008
α_2	0.115	0.019	0.107	0.018
α_3	0.002	0.007	0.000	0.008
ϕ	0.179	0.087	0.175	0.096
σ_1	0.055	0.007	0.051	0.005
σ_2	0.103	0.011	0.099	0.009
σ_3	0.049	0.005	0.051	0.005

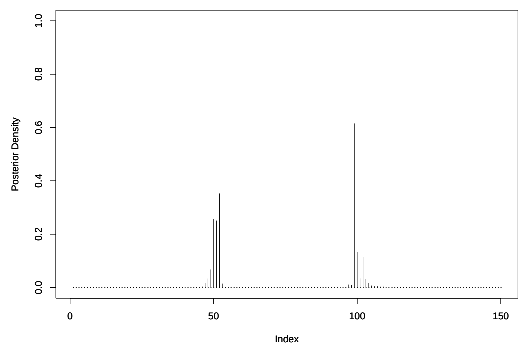

Figure 4. Estimated Changepoints – An Example from Design II

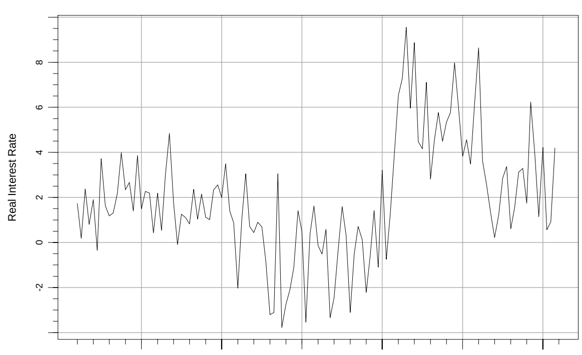

Figure 5. U.S. Real Interest Rate

where the notation follows equation (1) with $\phi = 0.2$, and

$$\begin{aligned} a_t = \alpha_1 = 0.0,\ s_t = \sigma_1 = 0.05, &\quad \text{for } 0 < t \le 50, \\ a_t = \alpha_2 = 0.1,\ s_t = \sigma_2 = 0.10, &\quad \text{for } 50 < t \le 100, \\ a_t = \alpha_3 = 0.0,\ s_t = \sigma_3 = 0.05, &\quad \text{for } 101 < t \le 150, \end{aligned}$$

so that there are two structural changes in the mean and variance with $\mathbf{k} = (51, 101)'$. This DGP is designed to mimic the behavior of the U.S. real interest rates.

Table 6. Parameter Estimates for U.S. Real Interest Rate

Parameters	Mean	Std. Error
α_1	1.813	0.374
α_2	-0.383	0.396
α_3	5.623	0.954
α_4	2.439	0.544
ϕ_1	0.064	0.114
ϕ_2	-0.050	0.094
σ_1	1.185	0.172
σ_2	1.908	0.252
σ_3	1.886	0.361
σ_4	1.616	0.352

Figure 3 plots a simulated series from this DGP. The center of the data seem to have a higher mean and variance but it is not exactly clear when the changes occurred.

The Gibbs sampler is employed for the estimation of models with no change in trend with $m = 0$, 1, 2, and 3 break points. Diffuse priors are used for $\mathbf{B} = (\alpha_1, \alpha_2, \alpha_3, \phi)'$ and σ_i, $i = 1, 2, 3$, such that $\mathbf{B}_0 = 0$, $v_0 = 1.001$, $\lambda_0 = 0.001$, and $\Sigma_{\mathbf{B}}$ is set to a diagonal matrix with elements 1000 on the diagonal. For a model with m structural change(s), the starting value of $\mathbf{k}$ is set at the m (approximate) equi-distant points between 1 and 150. Then the starting values of $\mathbf{B}$ and σ_i^2, $i = 1, 2, 3$, are computed as in a standard linear model with group-wise heteroskedasticity. Again, after running the Gibbs sampler for 300 iterations, we save the next 2000 draws for inference. This procedure is also replicated 100 times.

Out of the 100 replications, the percentage of each model being chosen by a certain criterion is recorded in Table 4. This time the two POR select the right number of changes less than 40% of the time, while BIC selects the right number more than 95% of the time. Using Kass and Raftery's rule for Bayes factors, there is always "very strong" evidence in favor of M_2 versus M_1 and M_0. Comparing M_3 with M_2, there is "strong" evidence in favor of M_3 in 17% of the simulations and but there is "very strong" evidence in only 4% of the simulations.

For M_2, Table 5 summarizes the posterior estimates for the series plotted in Figure 3. The estimates for the series in Figure 3 are generally close to Monte Carlo statistics and to the true values.

For M_2, the posterior mass function for each estimated change point for the series in Figure 3 is plotted in Figure 4. For this particular series, the modes of posterior mass functions of the change points are not at the true break dates. However the mass functions give very narrow ranges which cover the true break dates.

5. APPLICATIONS TO EMPIRICAL DATA

5.1 Structural Changes in U.S. Real Interest Rate

In this section we analyze structural changes in a measure of the U.S. real interest rate. We use the monthly data on inflation and the Treasury bill rate described in Mishkin (1990) converted to quarterly observations by extracting the end-of-quarter figures from the monthly

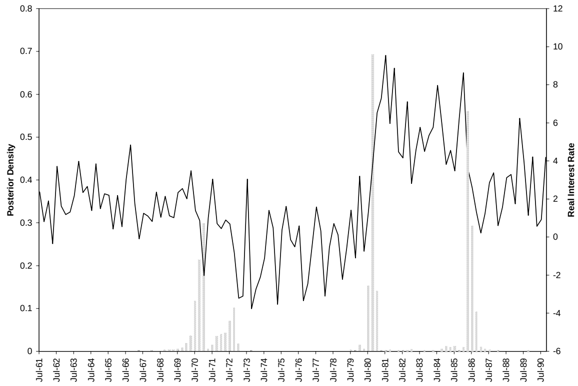

Figure 6. Posterior Probability Mass of Change Points

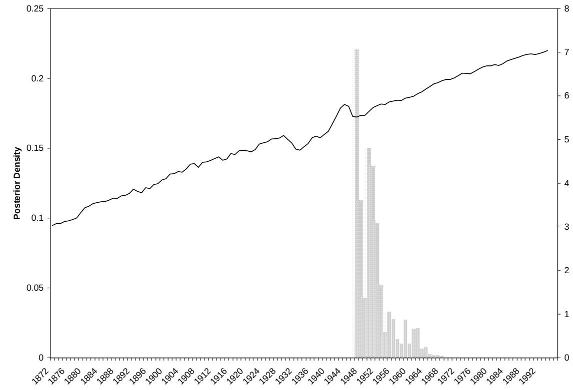

Figure 8. Posterior Probability Mass of the Change Point

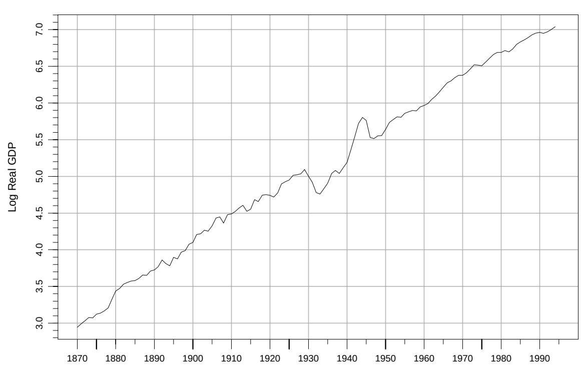

Figure 7. U.S. Real GDP

data. This data is also analyzed by Garcia et al. (1996) in a Markov switching analysis of multiple structural changes. A plot of the data is shown in Figure 5. Visually, there appears to be mean changes in the early part of 1970, the early part of 1980 and the latter part of 1980 and that the volatility of real interest rates appears smaller prior to 1970. To determine the number and type of structural changes and to be comparable to the analysis in Garcia et al. (1996), we estimate structural change models with two autoregressive terms allowing for breaks in the mean and variance with $m = 2$, 3 and 4 breaks. We adopt the prior distributions used for design II in the Monte Carlo section. After running the Gibbs sampler for 500 iterations, we save the next 2000 draws for inference.

The log marginal likelihood and BIC values for each model are:

$$\begin{aligned}
\ln f(Y|m=2) &= -240.69, \quad BIC(m=2) = 1058.16,\\
\ln f(Y|m=3) &= -232.26, \quad BIC(m=3) = 1053.08,\\
\ln f(Y|m=4) &= -231.15, \quad BIC(m=4) = 1077.26
\end{aligned}$$

Computing Bayes factors and evaluating them using the rules in Table 1 shows that there is “very strong” evidence in favor of three versus two breaks but only weak evidence in favor of four versus three breaks. In addition, minimizing the BIC gives the three break model.

The estimates of the parameters for the model with three structural changes are given in Table 6 and the posterior probability mass functions for the change points are shown in Figure 6. The posterior modes for the breaks are 1970.3, 1980.2 and 1985.4. Interestingly, our results indicate that the real interest rate first dropped just after the recession of 1970 and at about the time wage and price controls were implemented. The real rate rose after the huge rise in the real price of oil and the start of the Volker recession and then fell in the middle of 1986 when the real price of oil fell dramatically. Since the autoregressive parameters are essentially zero the estimates of the intercepts represent estimates of the ex-ante real interest rate. Our estimates show that real interest rates were slightly less than 2% prior to 1971, not significantly different from 0% during most of the 1970s, jumped to over 5% in the early 1980s and dropped to about 2.5% afterwards. In addition, our estimates show that the volatility of real interest rates was lower in the pre-1970 period. Our results for the mean shifts are qualitatively similar to those of Garcia and Perron but our results for the variance shifts differ. The difference in results for the variance can be attributed to Garcia and Perron’s restriction that both the mean and variance must change at the same time.

5.2 Structural Changes in U.S. Real GDP

We investigate the evidence of structural changes in U.S. real GDP using annual data over the years 1870 to 1994 taken from Maddison (1995). The logarithm of the series is plotted in Figure 7. Visually there appears to be a dip during the Great Depression, then faster growth during World War II, and after the war the growth rate appears to have gone back to the pre-Great-Depression trend. Ben-David et al. (1997) reject the null of a unit root

Table 7. Parameter Estimates for U.S. Real GDP with Change in Variance

Parameters	*Mean*	*Std. Error*
α_1	0.589	0.159
α_2	0.622	0.180
β_1	0.006	0.002
β_2	0.006	0.002
ϕ_1	1.130	0.090
ϕ_2	-0.312	0.091
σ_1	0.062	0.006
σ_2	0.023	0.003

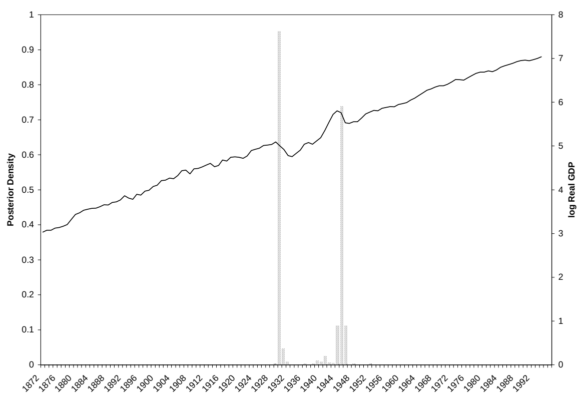

Figure 9. Posterior Probability Mass of the Change Point

Table 8.Parameter Estimates for U.S. Real GDP without Change in Variance

Parameters	*Mean*	*Std. Error*
α_1	0.965	0.224
α_2	-0.253	0.610
α_3	0.967	0.258
β_1	0.011	0.003
β_2	0.028	0.007
β_3	0.010	0.002
ϕ_1	0.902	0.102
ϕ_2	-0.212	0.121
σ	0.044	0.004

in this data at the 5% level using an extension of the Zivot et al. (1992) unit root test that allows for two endogenous breaks in level and trend.

Our visual inspection of the data suggests that there might be two structural changes in U.S. real GDP: one around 1930 and the other at the end of World War II. To determine the number and form of structural changes, we estimate three models with $m = 0$, 1 and 2 breaks. We fix the autoregressive lag at 2, which is common in the analysis of annual output data and also since higher order lags are not significant. We adopt the prior distributions used for design I in the Monte Carlo section. After running the Gibbs sampler for 300 iterations, we save the next 2000 draws for inference.

The logarithm of the marginal likelihood and BIC for each model are:

$$\ln f(Y|m=0) = 186.55, \quad BIC(m=0) = -698.08,$$
$$\ln f(Y|m=1) = 204.90, \quad BIC(m=1) = -732.98,$$
$$\ln f(Y|m=2) = 204.76, \quad BIC(m=2) = -693.50.$$

Clearly, the no structural change model is rejected by the data. Model choice based on Bayes factors and minimizing BIC favors the $m = 1$ model over the $m = 2$ model.

Given $m = 1$, the posterior probability of the change point is plotted in Figure 8 and the numerical moments of other parameters are reported in Table 7. From Figure 8, we see that the date of structural change occurred somewhere between 1947 and 1952, with the highest posterior probability being 0.22 at 1947. From the estimates in Table 7, the form of structural change appears to be a decrease in the variance term, with no significant changes in the level or time trend.

The above results seem to be contradictory with our visual inspection of the data that suggests two changes in trend. However, a model with two changes in trend implicitly assumes a constant variance over time. It appears that allowing the variance term s_t to be subject to structural changes as well the trend parameters effects the estimation of change points. To confirm this conjecture, we also estimate a model with two structural changes in the level and time trend while restricting the variance term s_t to be constant over time. The posterior probabilities of the two change points are plotted in Figure 9, and the numerical moments or the remaining parameters are reported in Table 8.

The two structural change points are estimated at 1930 and 1945 with very high posterior probabilities. The estimates in Table 8 are consistent with our visual inspection of the data: the pre-Great-Depression period and the post-World-War-II period share the same time trend, whereas the period between the Great Depression and World War II exhibits a drop in the level and a faster growth rate. For this model, the logarithm of the marginal likelihood is 201.68 and the BIC is -700.86. Hence, the data favor the one break model with a change in variance over the two break model with changes in trend and constant variance.

6. CONCLUSIONS

We developed a Bayesian approach for analyzing a dynamic time series model with multiple structural changes in level, trend and error variance based on the Gibbs sampler extending the approaches of Inclán (1993) and Stephens (1994). Our initial model is based on a fixed number of structural breaks and we treat the case of an unknown number of breaks as a model selection problem. Our Monte Carlo study demonstrated that for a fixed number of breaks our approach accurately locates the break dates and produces sharp estimates of the parameters of the model. Additionally, our Monte Carlo study revealed that when the number of breaks is unknown posterior odds are quite sensitive to the prior probabilities of models with different number of breaks whereas comparison of models based on Bayes factors using Kass and

Raftery's guidelines and the Schwarz BIC criterion accurately determined the number of breaks.

For future work, we want to compare our Bayesian methods with the classical methods of Bai et al. (1998). The Bayesian approach has the advantage of producing exact finite sample inference for all of the parameters of the model and in particular the break dates. We would also like to extend our univariate model to a multivariate model that can capture common break dates across series as in Bai, Lumsdaine and Stock (1997). We are also interested in the comparison of exogenous break models with endogenous break models; e.g. Markov switching models and self-exciting threshold models. The Bayesian framework with Gibbs sampling makes non-nested model comparison based on marginal likelihoods straightforward. Additionally, we want to see if it is possible to adapt the methods of Chib (1998) to the switching linear regression framework. The advantage of Chib's method is that it allows all of the break dates to be sampled simultaneously without a large increase in computations and so reduces the correlation between the sampled break dates in the Gibbs sampling algorithm. Finally, we would like to apply our methods to determine the empirical evidence for structural changes in international output series and compare our results with the results obtained by Ben-David et al. (1997) based on classical methods.

APPENDIX

Let N_T denote the total number of change points in a sample of size T and assume that the change points are $i.i.d.$ Bernoulli random variables with probability p. Let p $\sim$ beta(a, b) with density

$$f(p|a,b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)},\ 0 \le p \le 1,$$

where $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $\Gamma(c)$ is the gamma function such that $\Gamma(c) = (c-1)\Gamma(c-1)$. Then the prior density for N_T is

$$\begin{aligned}
\Pr(N_T) &= \int_0^1 f(N_T|p)f(p)dp \\
&= \int_0^1 ({}_TC_{N_T})p^{N_T}(1-p)^{T-N_T} \cdot \\
&\quad \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}dp \\
&= ({}_TC_{N_T})\frac{B(a+N_T, b+T-N_T)}{B(a,b)} \cdot \\
&\quad \int_0^1 \frac{p^{a+N_T-1}(1-p)^{b+T-N_T-1}}{B(a+N_T, b+T-N_T)}dp \\
&= ({}_TC_{N_T})\frac{B(a+N_T, b+T-N_T)}{B(a,b)} \\
&= \frac{1}{T-N_T}\frac{B(a+N_T, b+T-N_T)}{B(N_T+1, T-N_T)B(a,b)},
\end{aligned}$$

where we have used the relations

$${}_TC_n = \frac{T!}{n!(T-n)!} = \frac{1}{(T-n)B(n+1, T-n)},$$

and

$$\int_0^1 \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}dp = 1.$$

Then

$$\begin{aligned}
\frac{\Pr(N_T=j)}{\Pr(N_T=j-1)} &= \frac{T-j+1}{T-j}\frac{B(a+j, b+T-j)}{B(j+1, T-j)} \\
&\quad / \ \frac{B(a+j-1, b+T-j+1)}{B(j, T-j+1)} \\
&= \frac{T-j+1}{j}\frac{a+j-1}{b+T-j}.
\end{aligned}$$

NOTES

1. Related work for non-dynamic models is given in Liu, Wu and Zidek (1997).
2. The model can be modified without much difficulty to allow the autoregressive parameters to change as well. Also, under certain restrictions on a_t and b_t the trend function can become kinked at the break dates. Our parameterization does not impose these kinds of restrictions *a priori* but the model can be rewritten to incorporate them.
3. The Gibbs sampler has been used extensively in statistics and econometrics in recent years. For an introduction to Gibbs sampling and related Markov chain Monte Carlo methods see Casella and George (1992), Tanner (1993), Gelman, Carlin, Stern and Rubin (1995), Chib and Greenberg (1996) and Gamerman (1997).
4. For proofs of the convergence of the Gibbs sampler see Liu, Wong and Kong (1994, 1995), Roberts and Smith (1994) and Tierney (1994).
5. Gauss code implementing the Gibbs sampling algorithm is available upon request.
6. A comprehensive listing of convergence diagnostics for Markov Chain Monte Carlo methods can be found at `http://www.ensae.fr/crest/statistique/robert/McDiag`.
7. Some authors prefer to take every q-th draw after the convergence has occurred, which is described as "thinning the chain" to obtain approximately independent draws. However, the usual ways of picking q are ad hoc and the impact of thinning the chain may be undesirable as argued by Geyer (1992).
8. Stephens (1994) does not consider model selection issues in multiple break models.
9. Kass and Raftery (1995) note that although the harmonic mean is almost surely consistent it does not satisfy a central limit theorem and may exhibit unstable behavior. However, they argue that it often gives results that are accurate enough for interpretation on the log scale of Table 1. Chib (1995) describes an alternative method for calculating the marginal likelihood using the output of the Gibbs sampler.
10. Under this assumption all possible locations for the break dates $k_1 < k_2 < \cdots < k_m$ are equally probable with $f_0(\mathbf{k}|m) = 1/\,{}_{T-1}C_m$.
11. There is an error in the derivation of the prior odds under a beta prior for p in the Appendix of Inclán (1993).
12. Lubrano (1995) uses the Schwarz criterion to select models in a Bayesian analysis of cointegration with possible structural breaks.

REFERENCES

Andrews, D. W. K. (1993), "Tests for Parameter Instability and Structural Changes with Unknown Change Point", *Econometrica*, 61 (4), 821–856.

Bai, J. (1997a), “Estimation of a Change Point in Multiple Regression Models”, *Review of Economics and Statistics*, 79, 551–563.

——— (1997b), “Estimating Multiple Breaks One at a Time”, *Econometric Theory*, 13, 315–352.

Bai, J., and Perron, P. (1998), “Testing for and Estimation of Multiple Structural Changes”, *Econometrica*, 66, 817–858.

Bai, J., Lumsdaine, R. L., and Stock, J. H. (1997), “Testing for and Dating Common Breaks in Multivariate Time Series”, *Review of Economic Studies*, forthcoming.

Banerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992), “Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence”, *Journal of Business and Economic Statistics*, 10 (3), 271–287.

Ben-David, D., and Papell, D. H. (1995), “The Great Wars, the Great Crash, and the Unit Root Hypothesis: Some New Evidence About an Old Stylized Fact”, *Journal of Monetary Economics*, 36, 453–475.

Ben-David, D., Lumsdaine, R. L., and Papell, D. H. (1997), “Unit Roots, Postwar Slowdowns and Long-Run Growth: Evidence from Two Structural Breaks”, NBER Working Paper No. 6397.

Booth, N. B., and Smith, A. F. M. (1982), “A Bayesian Approach to Retrospective Identification of Change-Points”, *Journal of Econometrics*, 19, 7-22.

Broemeling, L. D., and Tsurumi, H. (1987), *Econometrics and Structural Change*, Marcel Dekker, Inc.

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992), “Hierarchical Bayesian Analysis of Changepoint Problems”, *Applied Statistics*, 41 (2), 389–405.

Casella, G., and George, E. I. (1992), “Explaining the Gibbs Sampler”, *American Statistician*, 46, 167–174.

Cheung, Y. W., and Chinn, M. D. (1996), “Deterministic, Stochastic, and Segmented Trends in Aggregate Output: A Cross-Country Analysis”, *Oxford Economic Papers*, 48 (1), 134–162.

Chen, J., and Gupta, A. K. (1997), “Testing and Locating Variance Changepoints with Application to Stock Prices”, *Journal of the American Statistical Association*, 92 (438), 739–747.

Chib, S. (1995), “Marginal Likelihood From the Gibbs Output”, *Journal of the American Statistical Association*, 90 (432), 1313–1321.

——— (1998), “Estimation and Comparison of Multiple Change-Point Models”, *Journal of Econometrics*, 86, 221–241.

Chib, S., and Greenberg, E. (1996), “Markov Chain Monte Carlo Simulation Methods in Econometrics”, *Econometric Theory*, 12, 409–431.

Chin Choy, J., and Broemeling, L. D. (1980), “Some Bayesian Inferences for a Changing Linear Model”, *Technometrics*, 22, 71–78.

Christiano, L. J. (1992), “Searching for a Break in GNP”, *Journal of Business and Economic Statistics*, 10 (3), 237–256.

Csörgó, M., and Horváth, L. (1997), *Limit Theorems in Change-Point Analysis*, John Wiley & Sons, Inc.

Culver, S. E., and Papell, D. H. (1995), “Real Exchange Rates Under the Gold Standard: Can They Be Explained by the Trend Break Model”, *Journal of International Money and Finance*, 14 (4), 539–548.

——— (1997), “Is There a Unit Root in the Inflation Rate? Evidence from Sequential Break and Panel Data Models”, *Journal of Applied Econometrics*, 12 (4), 435–444.

De Haan, J., and Zelhorst, D. (1993), “Does Output Have a Unit Root? New International Evidence”, *Applied Economics*, 25, 953–960.

DeJong, D. N. (1996), “A Bayesian Search for Structural Breaks in U. S. GNP”, in *Advances in Econometrics*, Vol. 11, Part B, pp. 109–146. JAI Press Inc.

Edison, H. J., and Fisher, E. (1991), “A Long-Run View of the European Monetary System”, *Journal of International Money and Finance*, 10 (1), 53–70.

Engel, C., and Hakkio, C. S. (1996), “The Distribution of Exchange Rates in the EMS”, *International Journal of Finance and Economics*, 1, 55–67.

Evans, M. D., and Lewis, K. (1995), “Do Expected Shifts in Inflation Affect Estimates of the Long-Run Fisher Relation?”, *Journal of Finance*, 50 (1), 225–252.

Ferreira, P. E. (1977), “A Bayesian Analysis of a Switching Regression Model: A Known Number of Regime”, *Journal of the American Statistical Association*, 70, 370–374.

Fong, W-M. (1997), “Volatility Persistence and Switching ARCH in Japanese Stock Return”, *Financial Engineering and the Japanese Markets*, 4 (1), 37-57.

Gamerman, D. (1997), *Markov Chain Monte Carlo*, Chapman & Hall.

Garcia, R., and Perron, P. (1996), “An Analysis of the Real Interest Rate under Regime Shifts”, *Review of Economics and Statistics*, 78 (1), 111–125.

Gelfand, A. E., and Dey, D. K. (1994), “Bayesian Model Choice: Asymptotics and Exact Calculations”, *Journal of the Royal Statistical Society, Series B*, 56, 501–514.

Gelfand, A. E., and Smith, A. F. M. (1990), “Sampling-Based Approaches to Calculating Marginal Densities”, *Journal of the American Statistical Association*, 85 (410), 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, Chapman & Hall.

Geyer, C. J. (1992), “Practical Markov Chain Monte Carlo”, *Statistical Science*, 7 (4), 473–483.

Holbert, D. (1982), “A Bayesian Analysis of a Switching Linear Model”, *Journal of Econometrics*, 19, 77–87.

Inclán, C. (1993), “Detection of Multiple Changes of Variance Using Posterior Odds”, *Journal of Business and Economic Statistics*, 11 (3), 289–300.

Inclán, C., and Tiao, G. C. (1994), “Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance”, *Journal of the American Statistical Association*, 89 (427), 913–923.

Kanas, A. (1998), “Testing for a Unit Root in ERM Exchange Rates in the Presence of Structural Breaks: Evidence from the Bootstrap”, *Applied Economics*, 5, 407–410.

Kass, R. E., and Raftery, A. E. (1995), “Bayes Factors”, *Journal of the American Statistical Association*, 90 (430), 773–795.

Kim, C-J., and Engel, C. (1997), “The Long Run U.S. / U.K. Real Exchange Rate”, *Journal of Money, Credit and Banking*, forthcoming.

Kim, C-J., and Nelson, C. (1998), “A Test for Structural Change in Markov Switching Models: Has the U.S. Economy Become More Stable?”, unpublished manuscript, Department of Economics, University of Washington.

Krishnaiah, P. R., and Miao, B. Q. (1988), “Review About Estimation of Change Points”, in *Quality Control and Reliability*, Handbook of Statistics, Vol. 7, 375–402.

Lamoureux, C. G., and Lastrapes, W. D. (1990), “Persistence in Variance, Structural Changes and the GARCH Model”, *Journal of Business and Economic Statistics*, 8 (2), 225-234.

Liu, J., Wong, W., and Kong, A. (1994), “Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes”, *Biometrika*, 81, 27–40.

——— (1995), “Correlation Structure and Convergence Rate of the Gibbs Sampler with Various Scans”, *Journal of the Royal Statistical Society, Series B*, 57, 157–169.

Liu, J., Wu, S., and Zidek, J. V. (1997), “On Segmented Multivariate Regressions”, *Statistica Sinica*, 7, 497–525.

Lubrano, M. (1995), “Bayesian Tests for Co-integration in the Case of Structural Breaks: An Application to the Analysis of Wage Moderation in France”, *Recherches Economiques de Louvain*, 61 (4), 479–507.

Lumsdaine, R. L., and Papell, D. H. (1997), “Multiple Trend Breaks and the Unit Root Hypothesis”, *Review of Economics and Statistics*, 79, 212–218.

Maddala, G. S., and Kim, I-M. (1996a), “Structural Change and Unit Roots”, *Journal of Statistical Planning and Inference*, 49, 73–103.

——— (1996b), “Bayesian Detection of Structural Changes”, Chapter 30 in *Bayesian Analysis in Statistics and Econometrics* (eds. D. A. Berry, K. M. Chaloner, and J. K. Geweke), pp. 359–370. John Wiley & Sons, Inc.

Maddison, A. (1995), *Monitoring the World Economy: 1820–1992*, Development Centre of OECD.

Mishkin, F. S. (1990), “What Does the Term Structure Tell Us About Future Inflation?”, *Journal of Monetary Economics*, 25, 77–95.

Newton, M. A., and Raftery, A. E. (1994), “Approximate Bayesian Inference by the Weighted Likelihood Bootstrap”, *Journal of the Royal Statistical Society, Series B*, 56, 3-48.

International Evidence", *Oxford Economic Papers*, 47 (2), 357–362.

Zivot, E., and Andrews, D. W. K. (1992), "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis", *Journal of Business and Economic Statistics*, 10 (3), 251–270.

Zivot, E., and Phillips, P. C. B. (1995), "A Bayesian Analysis of Trend Determination", *Econometric Reviews*, 13 (3), 291–336.