

# Bayesian Model Averaging for Multiple Structural Change Models

Jeremy M. Piger\*  
University of Oregon

October 2008

*Preliminary and Incomplete*

## Abstract

I consider the problem of accounting for model uncertainty in linear models with multiple structural changes of unknown timing in intercept, slope, and residual variance parameters. A feasible approach to Bayesian Model Averaging (BMA) is developed, where the model space encompasses variable selection, the number of structural changes, and the type of each structural change. This model space is too large to allow for direct evaluation of each potential model in all but the simplest cases. Instead, posterior model probabilities are estimated by extending the MC<sup>3</sup> algorithm of Madigan and York (1995) to the multiple structural change model. Direct application of the MC<sup>3</sup> algorithm is complicated by time-consuming computations necessary for marginal likelihood calculations. I instead apply the MC<sup>3</sup> algorithm to an augmented model space that includes the timing of structural change, conditional on which the structural change model collapses to a linear regression model for which the marginal likelihood is computed relatively quickly. The timing of structural change is then integrated out to arrive at posterior model probabilities for the model space of interest. The approach is applied to investigate structural changes in the dynamics of U.S. post-war inflation.

*JEL Classification System:* C11, C22, C52

*Keywords:* posterior model probability, Markov chain Monte Carlo, model selection, structural break, changepoint

---

\* Department of Economics, 1285 University of Oregon, Eugene, OR 97403, email: jpiger@uoregon.edu

# 1 Introduction

There are a host of economic and political developments that could be expected to alter the behavior of, and relationship among, economic variables. As a result, substantial attention has been given to developing and applying econometric models that incorporate parameter changes, where the timing of the change is not necessarily known. This paper focuses on an important class of such models, namely linear models with discrete structural changes in intercept, slope, and residual variance parameters. This class of models, referred to as simply “structural change models” in the following, is a common choice in applied work across a number of fields.<sup>1</sup>

The potential for structural changes introduces significant dimensions of model uncertainty regarding their number (how many changes?) and type (which parameters change?). This uncertainty is in addition to the standard uncertainty regarding which explanatory variables should be included in the model. If the researcher does not have *a priori* knowledge of the number and type of structural changes, a common situation in practice, working with structural change models can quickly become a showcase in model uncertainty. In particular, for even small numbers of structural changes and potential explanatory variables, the space of potential models to consider becomes enormous.

The existing literature has devoted significant attention to the model selection question of the number of structural changes. From the classical framework, Andrews (1993) and Andrews and Ploberger (1994) developed hypothesis tests of the null hypothesis of no structural change vs. the alternative of a single change, while Bai and Perron (1998) and Bai (1999) develop sequential testing procedures designed to reveal the number of, perhaps multiple, structural changes. From the Bayesian framework, Inclán (1993) and Wang and Zivot (2000) use posterior odds ratios to compare models with different numbers of breaks. However, far less attention has been given

---

<sup>1</sup>Recent examples of the application of the structural change model include studies of U.S. trend productivity growth (Hansen, 2001), the predictive power of the yield curve (Estrella, Rodrigues, and Schich, 2003), the dynamics of international unemployment rates (Pappell, Murray and Ghiblawi, 2000; Summers, 2004), the dynamics of real interest rates and inflation (Perron, 1990; Wang and Zivot, 2000; Rapach and Wohar, 2005; Clark, 2006), the volatility of macroeconomic and financial time series (Inclán, 1993; McConnell and Perez-Quiros, 2000; Stock and Watson, 2002; Sensier and van Dijk, 2004), the evaluation of crime reduction programs (Piehl, Cooper, Braga, and Kennedy, 2003), the conduct of U.S. monetary policy (Duffy and Engle-Warnick, 2006), the integration of world equity markets (Bekaert, Harvey and Lumsdaine, 2002), the “Phillips Curve” relationship (Stock and Watson, 1999; Clark and McCracken, 2006), the effects of sick leave compensation on the amount of sick leave taken (Henrekson and Persson, 2004), and the effects of the “War on Terror” on transnational terrorism (Enders and Sandler, 2005)

to other dimensions of model uncertainty in structural change models. Inoue and Rossi (2008) present procedures based on sequential hypothesis tests designed to provide evidence on the type of structural change. Their approach provides a subset of parameters that contain the stable parameters with a pre-specified probability. Hultblad and Karlsson (2008) use a Bayesian approach to jointly consider uncertainty regarding lag length selection and the number of structural changes in autoregressive models. This provides techniques to incorporate uncertainty regarding both variable selection and the number of structural changes. Finally, Levin and Piger (2006) use Bayesian posterior odds ratios to compare alternative structural change models that can differ over the inclusion of explanatory variables, the number of structural changes, and the type of each structural change. However, the Levin and Piger (2006) procedure is based on direct Bayesian comparison of all potential models, and is thus only feasible for lightly parameterized models with a small number of structural changes.

Meanwhile, the focus of the existing literature has been on selection of a particular structural change model, upon which inferences are then based. However, this practice ignores uncertainty regarding the model itself, which can have dramatic consequences on inferences about quantities of interest (e.g. Leamer, 1978; Hodges, 1987; Moulton, 1991; Draper, 1995; Kass and Raftery, 1995; Raftery, 1996; Fernández Ley and Steel, 2001). From a Bayesian perspective, incorporating model uncertainty is conceptually straightforward. In particular, an additional, discrete, parameter is defined that lies in the model space, and the posterior mass function for this parameter then provides posterior probabilities that each model is the true model. Posterior distributions for objects of interest are then averaged across alternative models, using these posterior model probabilities as weights. This procedure, known as Bayesian Model Averaging (BMA), allows for model uncertainty to be incorporated into inference regarding objects of interest.<sup>2</sup> Unfortunately, the direct averaging of all possible models will not be feasible when the model space is large, as is the case for all but the simplest structural change models.

This paper develops a feasible approach to conduct BMA for structural change models, where the model space encompasses variable selection, the number of structural changes, and the type of each structural change. The approach exploits the fact that when the timing of the structural changes

---

<sup>2</sup>For an introduction to BMA and a review of related literature, see Hoeting, Madigan, Raftery and Volinsky (1999).

is known, the model collapses to a linear regression model with known regressors and, potentially, heteroskedasticity of a known form. If we define an augmented model space that includes a choice for the timing of the structural changes, alternative models are then simply heteroskedastic linear regression models with different sets of regressors. I then rely on existing techniques developed to conduct feasible BMA for linear regression models. In particular, I apply the Markov Chain Monte Carlo Model Composition (MC<sup>3</sup>) algorithm of Madigan and York (1995), which was implemented by Raftery, Madigan and Hoeting (1997) to conduct BMA in linear regression models. This algorithm uses a Metropolis-Hastings algorithm to generate samples from the model space, which are used to estimate posterior model probabilities. Finally, once the posterior model probabilities for the augmented model space are estimated, the timing of structural change is then integrated out to arrive at posterior model probabilities for the model space of interest.

The remainder of the paper is organized as follows. Section 2 defines the class of Bayesian structural change models that will be considered, and establishes aspects of model uncertainty. Section 3 discusses BMA in the context of the structural change model, while Section 4 develops the feasible BMA procedures for the structural change model. Section 5 presents an empirical application of the proposed BMA procedures. In particular, I revisit a large literature investigating structural changes in the univariate dynamics of post-war U.S. inflation. Section 6 concludes.

## 2 The Multiple Structural Change Model

### 2.1 Model Specification and Model Uncertainty

This section defines the class of multiple structural change models to be considered and discusses dimensions of model uncertainty. Suppose we have  $T$  observations through time on a scalar dependent variable, collected in the vector  $Y$ , and on  $k$  candidate explanatory variables, collected in the  $T \times k$  matrix  $X$ . To begin, consider the linear regression model with constant parameters:

$$Y = \alpha \iota_T + X_j \beta_j + \varepsilon, \tag{1}$$

where  $X_j$  is a  $T \times k_j$  matrix holding a subset of the regressors in  $X$ , so that  $k_j \leq k$ ,  $\iota_T$  is a  $T$  vector of 1's,  $\beta_j \in \mathbb{R}^{k_j}$  is a vector of slope coefficients, and  $\varepsilon$  is a  $T$  vector of i.i.d. normal random variables with zero mean and variance  $h^{-1} \in \mathbb{R}_+$ . I assume the only source of model uncertainty

in (1) is with regards to variable selection, so that alternative models differ only by the choice of  $X_j$ . A common choice is to admit all possible combinations of the regressors in  $X$  as potential  $X_j$ , in which case there are  $2^k$  potential models.

The structural change model augments equation (1) to allow for  $m_j$  structural changes that occur at the changepoints  $\tau_{m_j} = (\tau^1, \tau^2, \dots, \tau^{m_j})'$ , with  $\tau^i$  a positive integer. To allow for specification choice regarding the type of structural change, I write this model as follows:

$$Y = \alpha \iota_T + X_j \beta_j + (Z_j^1 \bullet D(\tau^1)) \gamma_j^1 + \dots + (Z_j^{m_j} \bullet D(\tau^{m_j})) \gamma_j^{m_j} + \varepsilon, \quad (2)$$

where  $\bullet$  indicates element by element multiplication,  $Z_j^i$  is a  $T \times k_j^i$  subset of  $Z_j = [\iota_T : X_j]$ , so that  $k_j^i \leq (k_j + 1)$ ,  $D(\tau^i)$  is a  $T$  vector of dummy variables with first  $\tau^i$  elements equal to 0 and remaining elements equal to 1, and  $\gamma_j^i \in \mathbb{R}^{k_j^i}$  is a vector of coefficients capturing structural changes in the intercept and slope parameters at the  $i^{th}$  break date. Equation (2) is written compactly as:

$$Y = \Xi_j \theta_j + \varepsilon, \quad (3)$$

where  $\Xi_j = [\iota_T, X_j, (Z_j^1 \bullet D(\tau^1)), \dots, (Z_j^{m_j} \bullet D(\tau^{m_j}))]$  and  $\theta_j = (\alpha, \beta_j', \gamma_j^{1'}, \dots, \gamma_j^{m_j'})'$ .

In addition to structural changes in conditional mean parameters, the vector of regression disturbances is allowed to experience structural changes in variance. This is modeled as:

$$\varepsilon \sim N(0, I_T \bullet H^{-1}), \quad (4)$$

where  $H$  is the  $T$  vector given by:

$$H = h^0 \iota_T + D_j \delta_j. \quad (5)$$

In (5),  $D_j$  is a  $T \times r_j$  matrix holding a subset of the variables in  $[D(\tau^1), D(\tau^2), \dots, D(\tau^{m_j})]$ , so that  $r_j \leq m_j$ , while  $\delta_j \in \mathbb{R}^{r_j}$  is a vector of coefficients with  $i^{th}$  element:

$$\delta_{j,i} = h^i - h^{i-1}, \quad i = 1, \dots, r_j, \quad (6)$$

where  $h^i \in \mathbb{R}_+$ . In this formulation, there are  $r_j$  structural changes in disturbance variance, and  $(h^i)^{-1}$  is the disturbance variance in the time period between the  $i^{\text{th}}$  and  $(i+1)^{\text{th}}$  variance change. For notational simplicity, collect the  $r_j$  variance parameters into the vector  $h_{r_j} = (h^1, h^2, \dots, h^{r_j})'$ .

As is common in the recent literature on structural change models, the changepoints are assumed unknown and treated as additional parameters to be estimated. In the following, I restrict the changepoints to be a minimum of  $b$  time periods apart:

$$(\tau^i - \tau^{i-1}) \geq b, \quad i = 1, \dots, m_j + 1, \quad (7)$$

where  $\tau^0 = 0$  and  $\tau^{m_j+1} = T$ .

The  $j^{\text{th}}$  structural change model requires a number of specification choices, which introduce significant dimensions of model uncertainty over the linear regression model with constant coefficients. First, as with (1), the regressors allowed to enter the model must be chosen, which is again determined by a choice of  $X_j$ . Second, the number of structural changes,  $m_j$ , must be set, where  $m_j$  can take on integer values from 0 to a maximum of  $m^*$ . Third, we must specify which of the conditional mean parameters ( $\alpha$  and each element of  $\beta_j$ ) are allowed to change at each of the  $m_j$  changepoints. This is alternatively stated as a choice of which of the variables in  $X_j$  is allowed to have a changed relationship with  $Y$  at each changepoint, and is determined by choices for the  $Z_j^i$  matrices. I denote this choice as  $R_j = [Z_j^1, \dots, Z_j^{m_j}]$ . Finally, we must specify at which of the changepoints the disturbance variance is allowed to change. This is given by a choice of  $D_j$  in (5), which holds the relevant dummy variables for capturing structural changes in variance.

Note that it is possible for  $Z_j^i$  or  $D_j$  to be the empty set, in which case certain parameters capturing structural changes drop out of the model. However, the choice of  $Z_j^i$  and  $D_j$  must be consistent with  $m_j$  structural changes. For example, if  $Z_j^i$  is the empty set, so that there are no structural changes in conditional mean parameters at the  $i^{\text{th}}$  changepoint, then it must be the case that  $D_j$  includes  $D(\tau^i)$ , so that there is a residual variance change at  $\tau^i$ . Otherwise, the model would not have  $m_j$  structural changes, but  $m_j - 1$ .

The  $j^{\text{th}}$  structural change model, denoted  $M_j$ , is then determined by a choice of  $\{X_j, m_j, R_j, D_j\}$ , and the number of potential models is determined by the number of unique combinations of  $\{X_j, m_j, R_j, D_j\}$  considered. Throughout the remainder of this paper, I set  $m^*$  to

its largest possible value given  $T$  and  $b$ , which is  $m^* = \lfloor \frac{T}{b} \rfloor - 1$ , allow  $X_j$  to be any combination of  $X$ , allow  $Z_j^i$  to be any combination of  $Z_j$ , and allow  $D_j$  to hold any combination of the dummy variables in  $[D(\tau^1), D(\tau^2), \dots, D(\tau^{m_j})]$ . In this case, the number of potential structural change models is:

$$N = \sum_{m_j=0}^{\lfloor \frac{T}{b} \rfloor - 1} \left( \sum_{k_j=0}^k \left( \frac{k!}{k_j!(k-k_j)!} \right) (2^{k_j+2} - 1)^{m_j} \right). \quad (8)$$

The space of all  $N$  possible models is denoted by  $\mathcal{M}$ , where  $\mathcal{M} = \{M_j : j = 1, \dots, N\}$ .

## 2.2 Prior Specification

The Bayesian approach to parameter estimation and model comparison for alternative structural change models requires the specification of prior distributions for all model parameters. The joint prior density function for  $M_j$ , denoted  $p(\theta_j, h^0, h_{r_j}, \tau_{m_j} | M_j)$ , is usefully factored as follows:

$$p(\theta_j, h^0, h_{r_j}, \tau_{m_j} | M_j) = p(\tau_{m_j} | \theta_j, h^0, h_{r_j}, M_j) p(\theta_j, h^0 | h_{r_j}, M_j) p(h_{r_j} | M_j). \quad (9)$$

I specify proper prior density functions for each of the components of (9). For  $\tau_{m_j}$  I place equal prior probability on all possible locations of the changepoints. This is a commonly used prior in the literature investigating Bayesian multiple break models (e.g. Inclán, 1993; Stephens, 1994; Wang and Zivot, 2000). Denote the countable set of  $C_{m_j}$  possible values for  $\tau_{m_j}$  as  $\mathcal{S}_{m_j}$ . The probability mass function  $p(\tau_{m_j} | \theta_j, h^0, h_{r_j}, M_j)$  is then uniform:

$$p(\tau_{m_j} | \theta_j, h^0, h_{r_j}, M_j) = p(\tau_{m_j}) = \frac{1}{C_{m_j}}, \quad \tau_{m_j} \in \mathcal{S}_{m_j}, \quad (10)$$

where:

$$C_{m_j} = \frac{((T - 2b + 1) - (m_j - 1)(b - 1))!}{m_j! ((T - 2b + 1) - (m_j - 1)(b - 1) - m_j)!}. \quad (11)$$

I assume prior independence between each element of  $h_{r_j}$  and the other model parameters, so that  $p(\theta_j, h^0 | h_{r_j}, M_j) = p(\theta_j, h^0 | M_j)$ . For  $\theta_j$  and  $h^0$ , I use the natural conjugate Normal-Gamma prior density function:

$$\theta_j | h^0 \sim N(\underline{\mu}_j, (h^0)^{-1} \underline{V}_j), \quad (12)$$

$$h^0 \sim G(\underline{a}^0, \underline{b}^0). \quad (13)$$

The use of the natural conjugate prior will simplify calculations considerably in the case where there are no structural changes in variance. For the remaining residual variance parameters collected in  $h_{r_j}$ , I specify independent and identical Gamma prior density functions, so that  $p(h_{r_j} | M_j) = \prod_{i=1}^{r_j} p(h^i | M_j)$ , where:

$$h^i \sim G(\underline{a}, \underline{b}). \quad (14)$$

### 3 Bayesian Model Averaging for the Multiple Structural Change Model

The Bayesian approach to comparing alternative  $M_j$  is based on the posterior probability that  $M_j$  is the true model:

$$\Pr(M_j | Y) = \frac{f(Y | M_j) \Pr(M_j)}{\sum_{n=1}^N f(Y | M_n) \Pr(M_n)}, \quad M_j \in \mathcal{M}. \quad (15)$$

In (15),  $\Pr(M_j)$  is the prior probability that  $M_j$  is the true model, while  $f(Y | M_j)$  is the marginal, or integrated, likelihood function:

$$f(Y | M_j) = \int f(Y | \theta_j, h^0, h_{r_j}, \tau_{m_j}, M_j) p(\theta_j, h^0, h_{r_j}, \tau_{m_j} | M_j) d\theta_j dh^0 dh_{r_j} d\tau_{m_j}, \quad (16)$$

where:

$$f(Y | \theta_j, h^0, h_{r_j}, \tau_{m_j}, M_j) = \frac{|I_T \bullet H|^{\frac{1}{2}}}{(2\pi)^{\frac{T}{2}}} \exp \left[ -\frac{1}{2} (Y - \Xi_j \theta_j)' (I_T \bullet H) (Y - \Xi_j \theta_j) \right], \quad (17)$$

is the likelihood function for  $M_j$ .

Construction of  $\Pr(M_j|Y)$  requires specification of  $\Pr(M_j)$ , where  $\sum_{n=1}^N \Pr(M_n) = 1$ . Here I describe one strategy for setting  $\Pr(M_j)$  that will be used in the application presented in Section 6. To begin, factor  $\Pr(M_j)$  as follows:

$$\Pr(M_j) = \Pr(X_j, m_j, R_j, D_j) = \Pr(R_j, D_j | X_j, m_j) \Pr(X_j | m_j) \Pr(m_j). \quad (18)$$

For  $\Pr(m_j)$ , I assign equal probability to the model that contains no breaks ( $m_j = 0$ ) and the class of models with at least one break ( $m_j > 0$ ), and further assume equal probability across those models with at least one break:

$$\Pr(m_j) = \left\{ \begin{array}{ll} \frac{1}{2}, & m_j = 0 \\ \frac{1}{2(\lfloor \frac{T}{b} \rfloor - 1)}, & m_j = 1, 2, \dots, \lfloor \frac{T}{b} \rfloor - 1 \end{array} \right\} \quad (19)$$

For  $\Pr(X_j | m_j)$ , I assign equal probability to each of the  $2^k$  possible choices of  $X_j$ :

$$\Pr(X_j | m_j) = \Pr(X_j) = \frac{1}{2^k}. \quad (20)$$

Finally, for  $\Pr(R_j, D_j | X_j, m_j)$  I assign equal probability to each of the possible choices of  $R_j$  and  $D_j$  given  $X_j$  and  $m_j$ :

$$\Pr(R_j, D_j | X_j, m_j) = \frac{1}{(2^{k_j+2} - 1)^{m_j}}. \quad (21)$$

Once obtained, the posterior model probabilities can be used to provide statistical evidence regarding various objects of interest in a straightforward fashion. One approach for using  $\Pr(M_j|Y)$  is to find the model with highest posterior probability:

$$\widetilde{M}_j = \max_{M_j} [\Pr(M_j|Y)]. \quad (22)$$

and then make inference about objects of interest based on  $\widetilde{M}_j$  alone. However, this approach ignores information in models other than  $\widetilde{M}_j$ , and thus does not yield inferences that fully incorporate model uncertainty. Instead of basing inference on a single, highest probability model, we could

instead average inference about objects of interest across alternative models, where averaging is with respect to  $\Pr(M_j|Y)$ . In particular, suppose we have an object of interest from the structural change model, denoted  $\Psi$ , which has a common interpretation across alternative models. We can construct a posterior density function for  $\Psi$  that incorporates model uncertainty as follows:

$$p(\Psi|Y) = \sum_{n=1}^N p(\Psi|Y, M_n) \Pr(M_n|Y), \quad (23)$$

where  $p(\Psi|Y, M_j)$  is the posterior density for  $\Psi$  from model  $M_j$ . This approach to incorporating model uncertainty is called Bayesian Model Averaging (BMA). As an example of BMA in the context of a structural change model, consider the posterior probability that the true model contains  $m_j$  structural changes. This is constructed as follows:

$$\Pr(m_j|Y) = \sum_{n=1}^N \Pr(m_j|Y, M_n) \Pr(M_n|Y), \quad (24)$$

where  $\Pr(m_j|Y, M_n) = \Pr(m_j|M_n)$  is 1 if  $m_n = m_j$  and is 0 otherwise. As another example, consider the posterior density for the coefficient on the first regressor in  $X$ , denoted  $x_1$ . Define  $\delta_t^*$  as the value of this coefficient in period  $t$ , where:

$$\delta_t^* = \beta_j^* + \gamma_j^{1*} D_t(\tau^1) + \dots + \gamma_j^{m_j^*} D_t(\tau^{m_j}). \quad (25)$$

In (25),  $\beta_j^*$  is equal to the appropriate element of  $\beta_j$  when  $x_1$  is included in  $X_j$  and is zero otherwise,  $\gamma_j^{i*}$  is equal to the appropriate element of  $\gamma_j^i$  when  $x_1$  is included in  $Z_j^i$  and is zero otherwise, and  $D_t(\tau^i)$  is the  $t^{\text{th}}$  element of  $D(\tau^i)$ . The BMA posterior density for  $\delta_t^*$  is then:

$$p(\delta_t^*|Y) = \sum_{n=1}^N p(\delta_t^*|Y, M_n) \Pr(M_n|Y) \quad (26)$$

## 4 A Feasible Approach to BMA for the Multiple Structural Change Model

While BMA for the multiple structural change model is conceptually straightforward, the potentially enormous model space makes direct calculation of each  $\Pr(M_j|Y)$  based on (15) practically

infeasible for all but the simplest cases. For example, even with a small number of potential regressors ( $k = 2$ ), a moderate sample size ( $T = 100$ ), and a minimum regime length of 10 time periods ( $b = 10$ ), there are  $N > 10^{11}$  alternative models to consider. Although exacerbated by the presence of structural breaks, this problem is not unique to multiple structural change models. Indeed, direct calculation of all posterior model probabilities is infeasible for the linear regression model in (1) once  $k$  becomes large. For linear regression models, a popular alternative approach is to sample the model space using a posterior simulator designed to obtain draws from the multinomial probability distribution given by the posterior model probabilities. A commonly-used example of this approach is the Markov Chain Monte Carlo Model Composition (MC<sup>3</sup>) algorithm of Madigan and York (1995), which uses a posterior simulator based on the Metropolis-Hastings algorithm. MC<sup>3</sup> was implemented by Raftery, Madigan and Hoeting (1997) for Bayesian Model Averaging in linear regression models, and has been used in a number of economic applications involving linear regression (e.g. Fernández, Ley and Steele, 2001a, 2001b).

In this section I design an MC<sup>3</sup> algorithm to conduct BMA for the multiple structural change model. One could apply an MC<sup>3</sup> algorithm to sample directly from the model space defined by  $\mathcal{M}$ . However, in order to calculate the Metropolis-Hastings acceptance probability, this algorithm would require computation of the marginal likelihood in (16) for each model, which is very computationally expensive. To sidestep this computational difficulty, I instead apply the MC<sup>3</sup> algorithm to an augmented model space that includes a choice for the timing of the structural changes, conditional on which the structural change model collapses to a linear regression model for which the marginal likelihood is computed relatively quickly. This added element of the augmented model space is then integrated out to arrive at estimates for  $\Pr(M_j|Y)$ .

In particular, consider a model defined by a particular structural break model,  $M_j$ , augmented with a specific location for the break dates,  $\tau_{m_j}$ . Inspection of (3)-(5) reveals that this model is simply a Normal linear regression model with known regressor matrix,  $\Xi_j$ , and, potentially, a disturbance variance that changes at known dates. In the case where there are no structural changes in disturbance variance, the marginal likelihood can be computed analytically:

$$f(Y|M_j, \tau_{m_j}) = \frac{\Gamma(\frac{b^0+T}{2})(\underline{a}^0 \underline{b}^0)^{\frac{b^0}{2}}}{\Gamma(\frac{b^0}{2})\pi^{\frac{T}{2}}} \left( \frac{\left| \left( \underline{V}_j^{-1} + \Xi_j' \Xi_j \right)^{-1} \right|}{|\underline{V}_j|} \right)^{\frac{1}{2}} (d_j)^{\left( -\frac{b^0+T}{2} \right)}, \quad (27)$$

where:

$$d_j = \underline{a}^0 \underline{b}^0 + \left( Y - \Xi_j \hat{\theta}_j \right)' \left( Y - \Xi_j \hat{\theta}_j \right) + \left( \hat{\theta}_j - \underline{\mu}_j \right)' \left[ \underline{V}_j + \left( \Xi_j' \Xi_j \right)^{-1} \right]^{-1} \left( \hat{\theta}_j - \underline{\mu}_j \right), \quad (28)$$

and  $\hat{\theta}_j$  is the ordinary least squares estimate of  $\theta_j$ . In the case where there are structural changes in disturbance variance, the marginal likelihood no longer takes a convenient analytical form. However, there are a number of approximating techniques that can be used to provide very accurate estimates of the marginal likelihood for such a model, with little computational expense. Here, I use the approach of Chib(1995), which is based on simulations from the Gibbs Sampler applied to the linear regression model. For the linear regression model with structural changes in disturbance variance of known timing, and the parameter priors discussed above, Chib's approach can be implemented using only a single Gibbs run. In untabulated Monte Carlo experiments using a linear regression model with 10 regressors, 5 breaks in disturbance variance, and a sample size of 200, I found that the procedure produced estimates of the log marginal likelihood across 1000 separate Gibbs runs that were never more than 0.08% apart, even when using a very small number of total Gibbs simulations (100 simulations following 100 burn-in simulations).<sup>3</sup> Such a small number of Gibbs simulations can be produced in approximately one-100<sup>th</sup> of a second at current computing speeds, making use of the Chib approach as a step in the MC<sup>3</sup> sampler feasible. Details of the Chib (1995) technique are discussed in the appendix.

The MC<sup>3</sup> algorithm is then designed to simulate a chain of models from the augmented model space defined by  $M_j \in \mathcal{M}$  and  $\tau_{m_j} \in \mathcal{S}_{m_j}$ , where the limiting distribution of the chain is the multinomial distribution given by  $\Pr(M_j, \tau_{m_j}|Y)$ , the joint posterior probability that  $M_j$  is the true model and  $\tau_{m_j}$  is the true location of the changepoints. Denote the  $g^{th}$  simulated model from this chain as  $(M_q^{(g)}, \tau_{m_q}^{(g)})$ . That is,  $M_q^{(g)}$  is one of the structural break models,  $M_1, M_2, \dots, M_N$ ,

---

<sup>3</sup>These results are consistent with the Monte Carlo simulations of Bos (2002), which show the Chib (1995) approach to provide accurate estimates of the marginal likelihood for the linear regression model.

and  $\tau_{m_q}^{(g)}$  is one of the  $C_{m_q}$  possible values for  $\tau_{m_q}$ . The  $(g + 1)^{st}$  simulation is then generated via the following steps:

1. Choose a candidate model, denoted  $(M_s^{(*)}, \tau_{m_s}^{(*)})$  with equal probability from a set of neighborhood models to  $(M_q^{(g)}, \tau_{m_q}^{(g)})$ . The set of neighborhood models is defined as:
  - (a) The current model,  $(M_q^{(g)}, \tau_{m_q}^{(g)})$ .
  - (b) All models that add one explanatory variable to the current model, so that  $k_s = k_q + 1$ .
  - (c) All models that remove one explanatory variable from the current model, so that  $k_s = k_q - 1$ . Note that such models will also involve removing any regressors in  $R_q$  related to structural changes in the coefficient on the deleted explanatory variable.
  - (d) All models that add a single structural change for an existing explanatory variable, that is an explanatory variable included in  $X_q$ . This can be achieved by adding a coefficient break at an existing changepoint, so that  $m_s = m_q$ , or by adding a coefficient break at a new changepoint, in which case  $m_s = m_{q+1}$ .
  - (e) All models that remove a single structural change for an existing explanatory variable.
  - (f) All models that add a single structural change in disturbance variance. The new structural change can be either at an existing changepoint, so that  $m_s = m_q$ , or at a new changepoint, in which case  $m_s = m_{q+1}$ .
  - (g) All models that remove a single structural change in disturbance variance.
2. Compute the Metropolis-Hastings acceptance probability:

$$\alpha(g, *) = \min \left( \frac{f(Y|M_s^{(*)}, \tau_{m_s}^{(*)}) C_{m_q} \Pr(M_s) B_g}{f(Y|M_q^{(g)}, \tau_{m_q}^{(g)}) C_{m_s} \Pr(M_q) B_*}, 1 \right) \quad (29)$$

where  $B_g$  is the number of neighborhood models to the  $g^{th}$  simulated model and  $B_*$  is the number of neighborhood models to the candidate model. In (29), the prior probabilities for the structural break model,  $\Pr(M_s^{(*)})$  and  $\Pr(M_q^{(g)})$ , are given in (18). As discussed above, the marginal likelihoods,  $f(Y|M_s^{(*)}, \tau_{m_s}^{(*)})$  and  $f(Y|M_q^{(g)}, \tau_{m_q}^{(g)})$  can be computed using (27) when

the corresponding structural break model,  $M_s^{(*)}$  or  $M_q^{(g)}$ , does not contain structural changes in disturbance variance. Again, when the structural break model does contain structural changes in variance, the marginal likelihoods are computed using the approach in Chib(1995).

3. Set the  $(g + 1)^{st}$  simulated model equal to the candidate model with probability  $\alpha$  and equal to the  $g^{th}$  simulated model with probability  $1 - \alpha$ .

The algorithm is initialized with an arbitrary initial model.  $G_1 + G_0$  simulations are then conducted, with the first  $G_0$  discarded to ensure convergence.

To estimate posterior model probabilities for the alternative structural break models, we could simply count the relative frequency that each model is visited by the MC<sup>3</sup> algorithm. In other words, an estimate of  $\Pr(M_j|Y)$  is given by:

$$\frac{1}{G_1} \sum_{g=G_0+1}^{G_1} I(g), \quad (30)$$

where  $I(g)$  is an indicator function that equals one if  $M_q^{(g)} = M_j$  and is zero otherwise. Note that this approach implicitly gives zero posterior probability to any model that is not visited by the sampler. This fact suggests an alternative estimate of  $\Pr(M_j|Y)$ , which is to calculate posterior model probabilities directly, where the set of possible models is reduced to the set of models that were actually visited by the sampler. In other words, an estimate of  $\Pr(M_j|Y)$  is given by:

$$\frac{\sum_{\tau_{m_j} \in \mathcal{S}_{m_j}} f(Y|M_j, \tau_{m_j}) \Pr(M_j) \frac{1}{C_{m_j}} I(M_j, \tau_{m_j})}{\sum_{M_s \in \mathcal{M}, \tau_{m_s} \in \mathcal{S}_{m_s}} f(Y|M_s, \tau_{m_s}) \Pr(M_s) \frac{1}{C_{m_s}} I(M_s, \tau_{m_s})}, \quad (31)$$

where  $I(M_s, \tau_{m_s})$  is an indicator function that equals one if  $(M_s, \tau_{m_s})$  is a model that is visited by the MC<sup>3</sup> sampler and is zero otherwise. As the set of models visited by the sampler will have far fewer elements than the set of all possible models, these calculations will be feasible. Further, provided the estimates of  $f(Y|M_j, \tau_{m_j})$  are accurate, the posterior model probability estimates based on (31) are more accurate than those based on (30), as probabilities for the reduced set of models are computed directly rather than based on the relative frequency of model visits. When posterior model probabilities are based on (31), it is clear that the MC<sup>3</sup> algorithm is simply providing an

mechanism to seek out high probability models. This use of the MC<sup>3</sup> algorithm is discussed in detail in Fernandez, Ley and Steele (2001a, 2001b).

The two approaches to estimate  $\Pr(M_j|Y)$  also provides a useful diagnostic tool to check the convergence of the MC<sup>3</sup> sampler. In particular, provided that the MC<sup>3</sup> sampler has converged to taking draws from its ergodic distribution, and provided that enough draws are taken to accurately characterize relative model probabilities, the estimates based on (30) should be close to those based on (31). Indeed, Fernandez, Ley and Steele (2001b) suggest calculating the correlation between the two posterior model probability estimates as a metric to evaluate convergence of the MC<sup>3</sup> sampler.

## 5 Application: Structural Changes in the U.S. Post-War Inflation Rate

A substantial recent literature is devoted to evaluating the evidence for parameter change in time-series models for the post-war U.S. inflation rate. In particular, Cogley and Sargent (2001) argue that the persistence of shocks to the U.S. inflation rate have varied considerably over the sample period, being lower prior to the “great inflation” and after the Volcker disinflation, and higher between these episodes. The Cogley and Sargent results are consistent with earlier work by Evans and Wachtel (1993) and Barsky (1987), which documented variation in inflation persistence. However, the Cogley and Sargent results have been challenged by Pivetta and Reis (2007) and Stock (2001). In particular, these authors argue that evidence for shifts in persistence is not statistically significant, particularly once one allows for shifts in the residual variance of the model for the inflation rate. A large subsequent literature has investigated the evidence for changes in inflation persistence using a variety of models of parameter instability, and has provided mixed results.<sup>4</sup> The stakes in this debate are quite high, as the stylized facts regarding inflation are key metrics often used to evaluate the plausibility of structural macroeconomic models.

Here we apply the BMA procedures described above to evaluate the evidence for the number and type of structural changes in the parameters of an autoregressive process fit to the post-war U.S. inflation rate. We measure inflation as the quarterly percentage change in the U.S. GDP

---

<sup>4</sup>See, for example, Stock and Watson, 2007; Levin and Piger, 2002; Benati, ??; Giordani and Kohn, 2008; and Murray, Nikolsko-Rzhevskyy, and Papell, 2008.

Deflator, sampled from the first quarter of 1952 through the second quarter of 2008. To begin, consider the  $k_j^{th}$ -order autoregressive model with constant parameters:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_{k_j} y_{t-k_j} + \epsilon_t, \quad (32)$$

where  $y_t$  is the inflation rate measured for quarter  $t$ , and  $\epsilon_t \sim i.i.d.N(0, h^{-1})$ . A key quantity of interest in the study of inflation dynamics is the persistence of inflation, or the extent to which an innovation,  $\epsilon_t$ , has long-lived effects on the level of the inflation rate. A commonly used measure of persistence is the sum of the autoregressive coefficients, denoted  $\rho = \sum_{i=1}^{k_j} \phi_i$ . To measure  $\rho$  directly, we rewrite (32) using the Dickey-Fuller transformation:

$$y_t = \alpha + \rho y_{t-1} + \gamma_1 \Delta y_{t-1} + \dots + \gamma_{k-1} \Delta y_{t-(k_j-1)} + \epsilon_t, \quad (33)$$

where the  $\gamma_h$  are linear combinations of the  $\phi_i$ .

We use the BMA procedures developed above to consider both uncertainty about variable selection and uncertainty about the number and type of structural changes in (33). In the autoregressive model, uncertainty about variable selection is related to which autoregressive lags should be included. Suppose we consider a maximum of  $k^*$  autoregressive lags. Then, in the notation of the model in (2), we have the matrix of potential regressors,  $X = (y_{-1}, y_{-2}, \dots, y_{-k^*})$ , where  $y_{-k}$  is the vector holding the  $k^{th}$  lags of inflation.

The  $j^{th}$  inflation structural change model will then involve choices of  $X_j$ ,  $m_j$ ,  $R_j$ , and  $D_j$ , which in turn will define the matrix of regressors  $\Xi_j$  in (3). We parameterize the parameter prior distributions as follows. In terms of the conditional mean parameters, I set all elements of  $\underline{\mu}_j$  equal to zero, with the exception of the element corresponding to  $\rho$  when  $y_{-1}$  is included in the model, which is set equal to one. This centers the prior for  $\rho$  on a unit root process when  $y_{-1}$  is included in the model. The prior variance for conditional mean parameters is set as follows:

$$\underline{V}_j = \begin{pmatrix} c & 0 \\ 0 & \underline{V}_j^* \end{pmatrix}.$$

It is well known that posterior model probabilities are sensitive to the amount of prior information incorporated for parameters in alternative models. In particular, if a prior distribution is very

diffuse for a particular parameter, models that do not include this parameter will tend to be preferred over those that do include this parameter. Here I follow a strategy that is consistent with Fernandez, Ley and Steele (2001), which involves using very diffuse priors for parameters that are in all models, and more informative priors for parameters that are in some models, but not others. In particular, the element  $c$  in  $\underline{V}_j$  captures the prior variance on the intercept parameter,  $\alpha$ , which is assumed to be in all models considered. As such, I set  $c = 100,000$ , which suggests very little prior information regarding  $\alpha$ . The matrix  $\underline{V}_j^*$  captures the prior variance-covariance matrix of the remaining conditional mean parameters in (3), which are not in each possible model. To set  $\underline{V}_j^*$  I use the "g-prior" specification:

$$\underline{V}_j^* = [g_j \Xi_j' \Xi_j]^{-1} \quad (34)$$

The g-prior simplifies the specification of  $\underline{V}_j^*$  to the choice of a single value,  $g_j$ . The g-prior was introduced in Zellner (1986), and its use was investigated in the context of BMA for linear regression models by Fernandez, Ley and Steel (2001). Here I set  $g_j = \frac{1}{T}$ , where  $T$  is the sample size, which is based on the "unit information prior" of Kass and Wasserman (1995), and was one of the priors investigated in Fernandez, Ley and Steel (2001). Future versions will investigate the sensitivity of results to alternative specifications of  $g$ .

I specify the prior densities for the precision parameters,  $h^0, h^1, \dots, h^{r_j}$  as follows. The first-regime precision parameter,  $h^0$ , is in all models considered, and as such I specify a prior distribution that is quite diffuse, setting  $\underline{a}^0 = 1$  and  $\underline{b}^0 = 1 \times 10^{-10}$ . This yields a Gamma density with mean of 1 and variance  $1 \times 10^{10}$ . For the remaining precision parameters, I set  $\underline{a} = 1$  and  $\underline{b} = 0.5$ , which yields a Gamma density with mean of 1 and variance of 2. Finally, I set the minimum regime length to  $b = 8$  quarters, and the maximum lag length to  $k^* = 4$ . The prior probability of alternative models,  $\Pr(M_j|Y)$ , is set as described in Section 3.

The MC<sup>3</sup> algorithm described in Section 4 is run for  $G_0 = 1$  million simulations to ensure convergence, after which an additional  $G_1 = 2$  million simulations are collected on which to base inference.<sup>5</sup> I perform a number of checks to evaluate convergence of the sampler. First, the correlation between the probabilities based on sampler visit rates (based on (30)) and those based

---

<sup>5</sup>These simulations required roughly 8 hours - more details on computation time to come...

Table 1: Posterior Inclusion Probabilities for Potential Lags  
in U.S. Inflation Autoregression

Variable	Posterior Inclusion Probability
$y_{-1}$	1.000
$\Delta y_{-1}$	1.000
$\Delta y_{-2}$	0.939
$\Delta y_{-3}$	0.253

on direct calculation of posterior probabilities for the set of visited models (based on (31)) is greater than 0.99. Second, results based on a shorter run of  $G_1 = 500,000$  simulations after only  $G_0 = 100,000$  burn-in drawings were very similar to those obtained from the longer simulation. Third, results obtained from two runs, one in which the initial model used to start the sampler is the model with no breaks, and one in which the initial model is a model with the maximum number of breaks in all parameters, were also very similar.

Table 1 begins by giving the posterior inclusion probabilities for each of the potential explanatory variables in  $X$ . That is, for each particular variable in  $X$ , Table 1 reports the sum of the posterior probability of all models that contain this element. As can be seen from the table, all models visited by the sampler included  $y_{-1}$  and  $y_{-2}$ , so that the posterior probability that there is at least AR(2) dynamics in the inflation rate is unity. There is substantial evidence of higher order dynamics as well. In particular, the posterior inclusion probability for  $\Delta y_{-3}$  is 0.94. The inclusion probability for  $\Delta y_{-4}$ , which corresponds to AR(4) dynamics, is much lower, but still substantial.

Table 2 presents the posterior probabilities for alternative values of the number of structural breaks,  $m_j$ , defined as in (24). The results demonstrate that there is overwhelming evidence for structural breaks, and that two is the preferred number of structural breaks, receiving 72% of the posterior probability. There is some non-trivial uncertainty regarding the true number of structural breaks, with one and three breaks receiving 17% and 10% posterior probability respectively.

Next we move to evaluating the specific type of structural change models that are preferred. Table 3 gives the posterior probability for alternative numbers of structural breaks in the individual parameters of the model. In particular, for the model intercept ( $\alpha$ ), the measure of persistence ( $\rho$ ), and the disturbance precision ( $h$ ), Table 3 reports the sum of the posterior probability of all models that contain 0, 1, 2, ...,  $m^*$  structural changes in that parameter. These results suggest substantial

Table 2: Posterior Probability of Alternative Numbers of Structural Breaks in U.S. Inflation Autoregression

Number of Breaks	Posterior Probability
0	0.000
1	0.174
2	0.723
3	0.102
4	0.001
>4	0.000

Table 3: Posterior Probability of Structural Breaks in Individual Parameters of U.S. Inflation Autoregression

Parameter	Number of Breaks	Posterior Probability
$\alpha$	0	0.912
	1	0.080
	2	0.009
	>2	0.000
$\rho$	0	0.016
	1	0.220
	2	0.763
	>2	0.000
$h$	0	0.000
	1	0.666
	2	0.237
	3	0.096
	>3	0.000

evidence of changes in inflation persistence and disturbance variance. In particular, there is 98% posterior probability of at least one break in  $\rho$ , and 100% posterior probability of at least one break in disturbance variance. There is less evidence of changes in the regression intercept, with 91% posterior probability given to models with no break in  $\alpha$ .

Table 4 shows posterior model probabilities for the top 10 posterior probability structural change models that were visited by the sampler. There are at least two salient points that stand out in Table 4. First, there is substantial uncertainty regarding the identity of the true model, with several models receiving non-trivial posterior weight. This implies that inference that is based on only one model will substantially understate uncertainty related to the choice of model, and argues for the

Table 4: Ten Highest Posterior Probability Structural Change Models for U.S. Inflation Autoregression

Posterior Probability	Included Regressors	First Break	Second Break	Third Break
0.255	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}$	$\rho$	$\rho, h$	–
0.128	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}$	$\rho, h$	$\rho, h$	–
0.101	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}$	$\rho, h$	–	–
0.088	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, \Delta y_{t-3}$	$\rho$	$\rho, h$	–
0.045	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}$	$h$	$\rho, h$	$\rho, h$
0.028	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, \Delta y_{t-3}$	$\rho, h$	$\rho, h$	–
0.026	$y_{t-1}, \Delta y_{t-1},$	$\rho$	$\rho, h$	–
0.021	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, \Delta y_{t-3}$	$\rho$	$\rho, h$	–
0.021	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, \Delta y_{t-3}$	$\rho, h$	–	–
0.020	$y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, \Delta y_{t-3}$	$h$	$\rho, h$	$\rho, h$

use of BMA. Second, all of the top 10 models are partial structural change models that are not represented by the baseline specification often used in the literature in which all parameters are allowed to change at each changepoint.

Finally, Figures 1-2 plot the median of the BMA posterior distribution for the sum of the autoregressive coefficients and the standard deviation of the disturbance variance at each point in the sample, constructed as in (26). Figure 1 shows a pattern for the persistence measure that is consistent with Cogley and Sargent (2001), in which persistence rises during the 1960's, remains high in the 1970's, and falls in the early 1980's. The disturbance standard deviation is high during the first half of the sample, before falling dramatically in the early 1980's.

## 6 Conclusion

Coming soon....

Figure 1: Bayesian Model Averaged Posterior Distribution of Sum of AR Coefficients for U.S. Inflation Autoregression

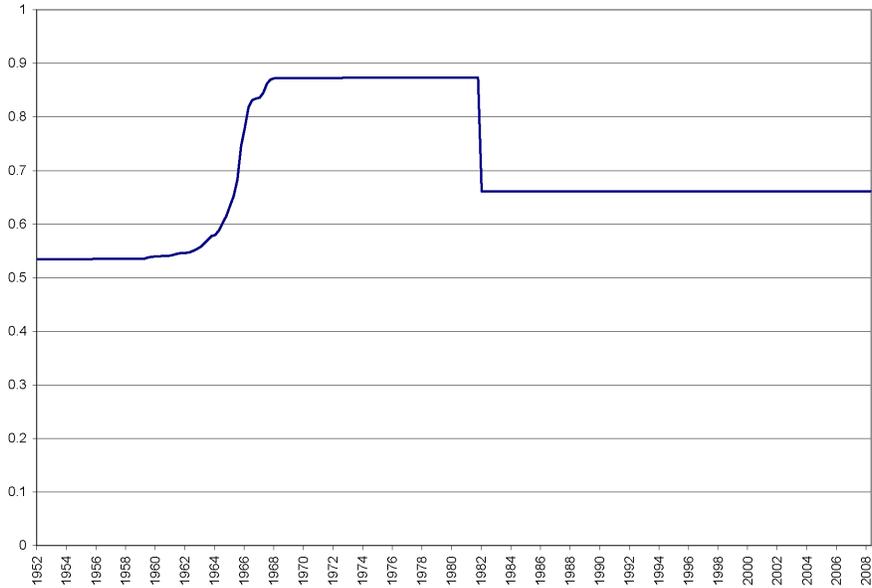
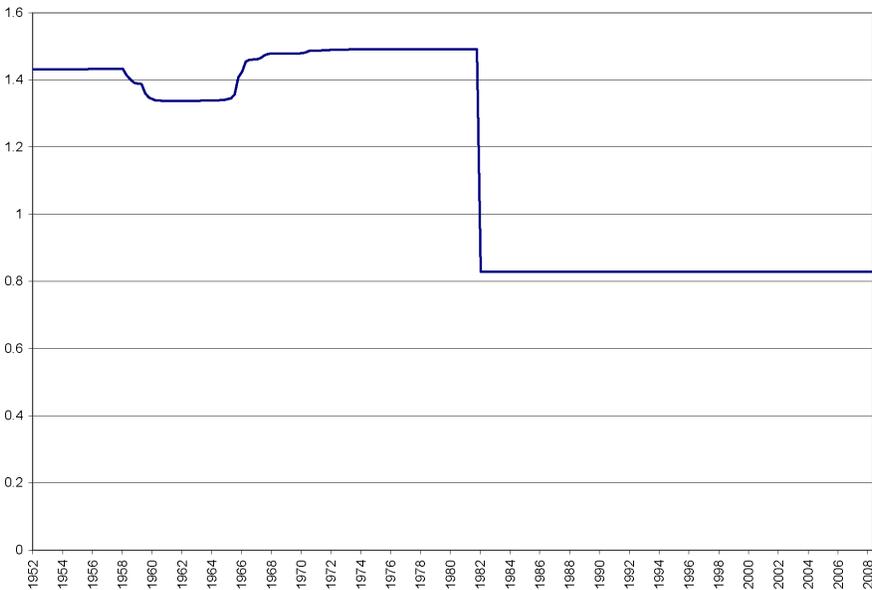


Figure 2: Bayesian Model Averaged Posterior Distribution of Disturbance Std. Deviation for U.S. Inflation Autoregression



## References

- Andrews, D.W.K., 1993. Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61, 821-856.
- Andrews, D.W.K., Ploberger, W., 1994. Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative. *Journal of Econometrics* 70, 9-38.
- Bai, J., 1999. Likelihood Ratio Tests for Multiple Structural Changes. *Journal of Econometrics* 91, 299-323.
- Bai, J., Perron, P., 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66, 47-78.
- Bekaert, G., Harvey, C.R., Lumsdaine, R.L., 2002. Dating the Integration of World Equity Markets. *Journal of Financial Economics* 2, 203-247.
- Chib, S., 1995. Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* 90, 1313-1321.
- Chib, S., 1998. Estimation and Comparison of Multiple Change-Point Models. *Journal of Econometrics* 86, 221-241.
- Clark, T., 2006. Disaggregate Evidence on the Persistence of Consumer Price Inflation. *Journal of Applied Econometrics* 21, 563-587.
- Clark, T., McCracken, M., 2006. The Predictive Content of the Output Gap for Inflation: Resolving the In-Sample and Out-of-Sample Evidence. *Journal of Money, Credit and Banking* 38, 1127-1148.
- Cogley, T., Sargent, T., 2001. Evolving Post-World War II U.S. Inflation Dynamics. *NBER Macroeconomics Annual* 331-372.
- Cogley, T., Sargent, T., 2005. Drifts and Volatilities: Monetary Policies and Outcomes in the Post World War II U.S. *Review of Economic Dynamics* 8, 262-302.
- Draper, D., 1995. Assessment and Propagation of Model Uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B* 57, 45-97.
- Duffy, J., Engle-Warnick, J., 2006. Multiple Regimes in U.S. Monetary Policy? A Nonparametric Approach. *Journal of Money, Credit and Banking* 38, 1363-1377.
- Enders, W., Sandler, T., 2005. After 9/11: Is it All Different Now? *Journal of Conflict Resolution* 49, 259-277.
- Estrella, A., Rodrigues, A.P., Schich, S., 2003. How Stable is the Predictive Power of the Yield Curve? Evidence from Germany and the United States. *The Review of Economics and Statistics* 85, 629-644.
- Fernández, C., Ley, E., Steele, M.F.J., 2001. Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics* 100, 381-427.
- Giordani, P., Kohn, R., 2008. Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models. *Journal of Business and Economic Statistics* 26, 66-77.

- Hansen, B.E., 2001. The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity. *Journal of Economic Perspectives* 15, 117-128.
- Henrekson, M., Persson, M., 2004. The Effects on Sick Leave of Changes in the Sickness Insurance System. *Journal of Labor Economics* 22, 87-113.
- Hodges, J.S., 1987. Uncertainty, Policy Analysis, and Statistics. *Statistical Science* 2, 259-291.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian Model Averaging: A Tutorial (with Discussion). *Statistical Science* 14, 382-401. Correction: 15, 193-195.
- Inclán, C., 1993. Detection of Multiple Changes of Variance using Posterior Odds. *Journal of Business and Economic Statistics* 11, 289-300.
- Kass, R.E., Raftery, A.E., 1995. Bayes Factors. *Journal of the American Statistical Association* 90, 773-795.
- Leamer, E.E., 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. Wiley, New York.
- Levin, A., Piger, J., 2006. Bayesian Model Selection for Multiple Structural Break Models. mimeo.
- Madigan, D., York, J., 1995. Bayesian Graphical Models for Discrete Data. *International Statistical Review* 63, 215-232.
- McConnell, M.M., Quiros, G.P., 2000. Output fluctuations in the United States: What has changed since the early 1980s? *American Economic Review* 90, 1464-1476.
- Moulton, B.R., 1991. A Bayesian Approach to Regression Selection and Estimation with Application to a Price Index for Radio Services. *Journal of Econometrics* 49, 169-193.
- Murray, C.J., Nikolsko-Rzhevskyy, A., Papell, D.H., 2008. Inflation Persistence and the Taylor Principle. mimeo.
- Papell, D.H., Murray, C.J., Ghiblawi, H., 2000. The Structure of Unemployment. *The Review of Economics and Statistics* 82, 309-315.
- Perron, P., 1990. Testing for a Unit Root in a time Series with a Changing Mean. *Journal of Business and Economic Statistics* 8, 153-162.
- Piehl, A.M., Cooper, S.J., Braga, A.A., Kennedy, D.M., 2003. Testing for Structural Breaks in the Evaluation of Programs. *The Review of Economics and Statistics* 85, 550-558.
- Pivetta, F., Reis, R., 2007. The Persistence of Inflation in the United States. *Journal of Economic Dynamics and Control* 31, 1326-1358.
- Raftery, A.E., 1996. Approximate Bayes Factors and Accounting for Model Uncertainty in Generalised Linear Models," *Biometrika* 83, 251-266.
- Raftery, A.E., Madigan, D., Hoeting, J.A., 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 179-191.
- Rapach, D.E., Wohar, M.E., 2005. Regime Changes in International Real Interest Rates: Are They a Monetary Phenomenon? *Journal of Money, Credit, and Banking* 37, 887-906.

- Sensier, M., van Dijk, D., 2004. Testing for volatility changes in U.S. macroeconomic time series, *The Review of Economics and Statistics* 86, 833-839.
- Stock, J.H., 2001. Comment on Evolving Post-World War II U.S. Inflation Dynamics. *NBER Macroeconomics Annual* 379-387.
- Stock, J.H., Watson, M.W., 1999. Forecasting Inflation. *Journal of Monetary Economics* 44, 293-335.
- Stock, J.H., Watson, M.W., 2002. Has the business cycle changed and why? *NBER Macroeconomics Annual* 17, 159-218.
- Stock, J.H., Watson, M.W., 2007. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking* 39, 3-33.
- Summers, P.M., 2004. Bayesian Evidence on the Structure of Unemployment. *Economics Letters* 83, 299-306.
- Wang, J., Zivot, E., 2000. A Bayesian Time Series Model of Multiple Structural Changes in Level, Trend, and Variance. *Journal of Business and Economic Statistics* 18, 374-386.