Bayesian Model Selection for Structural Break Models*

Andrew T. Levin
Federal Reserve Board

Jeremy M. Piger
University of Oregon

May 2008

Abstract: We take a Bayesian approach to model selection in regression models with structural breaks in conditional mean and residual variance parameters. A novel feature of our approach is that it does not assume knowledge of the parameter subset that undergoes structural breaks, but instead conducts model selection jointly over the number of structural breaks and the subset of the parameter vector that changes at each break date. Simulation experiments demonstrate that conducting this joint model selection can be quite important in practice for the detection of structural breaks. We apply the proposed model selection procedure to characterize structural breaks in the parameters of an autoregressive model for post-war U.S. inflation. We find important changes in both residual variance and conditional mean parameters, the latter of which is revealed only upon conducting the joint model selection procedure developed here.

Keywords: Posterior Model Probability, Markov Chain Monte Carlo, Inflation Persistence

JEL Codes: C11, C22, C52, E31

* Levin: Federal Reserve Board, Stop 70, Washington, DC 20551 (Andrew.levin@frb.org); Piger: Dept. of Economics, 1285 University of Oregon, Eugene, OR 97403-1285 (jpiger@uoregon.edu). We thank Mike McCracken, Bruce Mizrach, James Morley, Jonathan Wright, Eric Zivot, and seminar participants at the University of Oregon, University of Houston, University of Washington, 2005 SEA meetings, 2007 ISF meetings, and 2008 SNDE meetings for helpful comments. The views expressed in this paper are solely the responsibility of the authors, and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System.

1. Introduction

Regression models with parameters that undergo structural change have become a staple of the applied time-series econometrician's toolkit. One reason for the popularity of such models is the substantial evidence for parameter instability in regressions involving key economic variables, particularly macroeconomic and financial variables, measured over the post-war sample period. For example, there is overwhelming evidence of parameter breaks in autoregressive models of U.S. real output, particularly in the residual variance parameter (Kim and Nelson, 1999 and McConnell and Perez-Quiros, 2000). Likewise, Garcia and Perron (1996), Rapach and Wohar (2005) and Levin and Piger (2002) find important shifts in the intercept parameter of autoregressive models for interest rates and inflation in G7 countries. Indeed, Stock and Watson (1996) document instability in the parameters of a univariate autoregression for a "significant fraction" of the 76 U.S. macroeconomic time-series they study over the post-war period.

In many cases, the nature of structural breaks, in terms of their number and timing, is not known ex-ante, and a large literature has emerged focusing on testing for the existence of structural breaks, where the date of the potential structural break is unknown. For example, Andrews (1993), Andrews and Ploberger (1994) and Diebold and Chen (1996) develop tests of a model with no structural breaks against the alternative of a model with a single structural break. Bai and Perron (1998) and Bai (1999) develop sequential testing procedures designed to reveal the number of, perhaps multiple, structural breaks. Wang and Zivot (2000) discuss Bayesian estimation of a time-series model with multiple structural breaks, as well as present approaches, based on Bayesian model comparison, to determine the number of structural breaks.

In developing these approaches to determine the number of structural breaks, the literature has taken the subset of parameters that change at each break date as given. That is, there has been little attention paid to the model selection question of which parameters change at each break date. This is a significant omission, as there are reasons to believe that it may be important to conduct model selection jointly over the number of structural breaks and the parameters that change at each break date. Perhaps most importantly, evidence for a break may be revealed only if the subset of parameters that undergo structural breaks is correctly specified. For example, in evaluating the evidence for parameter breaks, suppose the researcher has no a priori knowledge of which parameters are likely to have undergone breaks, and thus allows all parameters to change at each break date, a common practice in testing for structural breaks. Such a procedure is likely to have low power to identify structural breaks if only a small subset of the parameter vector actually changes. Further, even if one is able to accurately determine the number of structural breaks, interpreting the economic meaning of the breaks may be aided by identifying which parameters break at each break point. For example, in time series models of macroeconomic variables, the economic interpretation of changes in the persistence of the series is often quite different from the economic interpretation of changes in the residual variance.

It is not difficult to find examples of regression models where careful attention to establishing the subset of the parameter vector that undergoes parameter change might yield important dividends. For example, a lively debate has emerged on the existence of shifts in conditional mean parameters of key equations for models of the U.S. macroeconomy, such as the Phillips Curve and the Federal Reserve's "reaction function".

On the one hand, Clarida, Gali and Gertler (2000), Cogley and Sargent (2001, 2005), and Boivin (1999) find important within-sample variation in their estimates of key conditional mean parameters. However, Sims (1999, 2001) and Sims and Zha (2006) argue that allowing for such changes does not provide a statistically superior fit over models with constant conditional mean parameters and structural change in covariance matrix parameters. Given the potentially large number of conditional mean parameters that may change in such models, the results could be quite sensitive to whether all conditional mean parameters are allowed to change, or only a subset. Indeed, Boivin (1999) notes that a key reason for discrepancies in statistical evidence for structural change observed in this literature is the differing number of parameters that are allowed to break in alternative model specifications.

In this paper we take a Bayesian approach to model selection in regression models with structural breaks in conditional mean and residual variance parameters. An important element of our approach is that it does not condition on the parameter subset that undergoes structural breaks, but instead conducts model selection jointly over the number of structural breaks and the subset of the parameter vector that changes at each break date. Specifically, we proceed by computing and comparing posterior model probabilities, where the space of potential models is expanded to include models that differ not just by the number of structural breaks, but also by which elements of the parameter vector are fixed and which elements change across individual break dates.

The Bayesian approach is well suited for the model selection problem studied in this paper, as this problem involves the comparison of non-nested models. For example, one may be interested in comparing a model in which there are two structural breaks, one in

the intercept parameter and one in the residual variance parameter, against a model in which there are two breaks in the residual variance parameter. The Bayesian approach proceeds by comparing posterior probabilities for various competing models, an approach for which non-nested models create no special considerations. Bayesian model selection for the number of structural breaks was developed and discussed by Wang and Zivot (2000).[1] Here we extend the Wang and Zivot framework to allow for model selection that encompasses the subset of parameters that undergoes structural breaks.

To evaluate the performance of our proposed model selection procedure, we conduct a series of simulation experiments in which we generate artificial data from regression models with varying numbers of structural breaks, and the structural breaks occur in a subset of the parameter vector. The results of these experiments suggest there are potentially sizeable gains to conducting model selection over the subset of parameters that undergo breaks rather than simply allowing all parameters to change. In particular, the likelihood of selecting the model with the correct number of structural breaks is substantially enhanced when model selection is expanded to include the subset of parameters that undergo breaks. Further, the simulation experiments suggest that the proposed Bayesian approach is relatively successful at identifying the correct subset of parameters that undergo change.

Finally, we apply our model selection approach to characterize possible structural breaks in conditional mean and residual variance parameters in an autoregression for the U.S. inflation rate. There is substantial ongoing debate about the existence of such breaks, with Cogley and Sargent (2001, 2005) arguing that the inflation process has

[1] Summers (2004) uses the Wang and Zivot (2000) methodology to model structural breaks in OECD unemployment rates.

undergone important changes in these conditional mean parameters, including the persistence of the process, while other authors, such as Stock (2001) and Pivetta and Reis (2007) argue that, while there is strong evidence of changes in residual variance parameters, changes in conditional mean parameters are much less obvious. Our results reveal several insights that contribute to this literature. First of all, the Bayesian model-selection procedures suggest there have been substantial changes in both conditional mean and residual variance parameters. Second, and importantly, the evidence for conditional mean parameters is revealed only when one conducts model selection over the subset of conditional mean parameters that undergo breaks, a result that demonstrates the empirical relevance of our proposed model selection procedure. Finally, the results suggest that evidence for intercept parameter shifts is substantial, but evidence for shifts in the persistence of the process is less so. Indeed, estimates of inflation persistence obtained by Bayesian model averaging of the various models under consideration suggests that inflation persistence has been roughly constant over the sample, albeit at a substantially lower level than estimates obtained assuming a constant parameter autoregression.

Section 2 lays out the empirical model of interest and describes the Bayesian approach to model selection. Section 3 details the results of simulation experiments designed to evaluate the performance of the Bayesian techniques. Section 4 presents an application of the Bayesian model selection procedures to modeling U.S. inflation rates. Section 5 concludes and offers some directions for future research.

2. Model Specification, Bayesian Estimation, and Model Selection

2.1 Model Specification

Consider the following time-series regression with m structural breaks:

$$\begin{aligned} y_t &= x_t'\beta^1 + \sigma^1\varepsilon_t \qquad && \tau_0 + 1 \le t \le \tau_1, \\ y_t &= x_t'\beta^2 + \sigma^2\varepsilon_t \qquad && \tau_1 + 1 \le t \le \tau_2, \\ &. && . \\ &. && . \\ y_t &= x_t'\beta^{m+1} + \sigma^{m+1}\varepsilon_t \qquad && \tau_m + 1 \le t \le \tau_{m+1}, \\ t &= 1,....,T\,, \end{aligned} \tag{1}$$

where y_t is a scalar dependent variable observed at time *t*, $x_t = \left(x_{1,t}, x_{2,t},..., x_{k,t}\right)'$ is a $(k \times 1)$ vector of exogenous or predetermined covariates, $\beta^i = \left(\beta_1^i, \beta_2^i,..., \beta_k^i\right)'$ is a $(k \times 1)$ vector of coefficients, $i = 1,..., m+1$, and $\varepsilon_t \sim i.i.d.N(0,1)$. The initial and final break dates, τ_0 and τ_{m+1} are equal to 0 and *T* respectively. The *m* remaining break dates are assumed unknown and treated as additional parameters to be estimated.

The model in (1) can be cast in matrix notation as follows. Define D_m as the $\left(T \times (m+1)\right)$ matrix whose (t,i) element is equal to one if $\tau_{i-1} < t \le \tau_i$ and zero otherwise. In other words, D_m is a matrix whose i^{th} column is a dummy variable indicating those time periods that are included in the regime beginning at $\tau_{i-1} + 1$ and ending at τ_i. Collect the conditional mean and residual variance parameters into the vectors $\beta = \left(\beta_1^1, \beta_1^2,..., \beta_1^{m+1}, \beta_2^1, \beta_2^2,..., \beta_2^{m+1},...., \beta_k^{m+1}\right)'$ and $\sigma = \left(\sigma^1, \sigma^2,..., \sigma^{m+1}\right)'$.

Finally, define $\tilde{y} = (y_1, y_2, ..., y_T)'$, $\tilde{x}_j = (x_{j,1}, x_{j,2}, ..., x_{j,T})'$, $j = 1, ..., k$, and $\tilde{s} = D_m \sigma$. We can then write equation (1) as:

$$\tilde{y} = X\beta + \tilde{u}, \tag{2}$$

where $X = [D_m \bullet \tilde{x}_1 \quad D_m \bullet \tilde{x}_2 \quad . \quad . \quad D_m \bullet \tilde{x}_k]$ is a $(T \times ((m+1)*k))$ matrix with $\bullet$ indicating element by element multiplication, and $\tilde{u} \sim N(0_T, I_T \bullet \tilde{s})$.

We want to consider models where not all parameters are allowed to change across each break date. Define the matrix R_{β_j}, $j = 1, ..., k$, as the $((m+1) \times q_{\beta_j})$ matrix that, through post-multiplication of D_m, sums the i^{th} and i+1th columns of D_m if $\beta_j^i = \beta_j^{i+1}$, with R_σ defined analogously. For example, for a model with constant residual variance parameter, the matrix R_σ is simply an $((m+1) \times 1)$ matrix of ones. Alternatively, for a model with residual variance parameter that is allowed to change across all $m+1$ break dates, R_σ is the $(m+1)$ identity matrix. In this notation, q_{β_j} and q_σ denote the number of regimes over which the coefficient on $\tilde{x}_j$ and the residual variance parameter are allowed to take on different values respectively. Thus, the total number of unique conditional mean parameters is $\sum_{j=1}^{k} q_{\beta_j}$ and the total number of unique residual variance parameters is q_σ. Note that R_{β_j} and R_σ matrices that restrict all parameters to be equal across any particular break date, that is that restrict $(\beta^i, \sigma^i)' = (\beta^{i+1}, \sigma^{i+1})'$, are inadmissible, as these restrictions imply that the number of structural breaks is not *m*, but *m-1*.

Define the vector β_R as the matrix β reduced to contain only the $\sum_{j=1}^{k} q_{\beta_j}$ unique elements of β, and σ_R as the vector σ reduced to contain only the q_σ unique elements of σ. The model in (2) can then be rewritten to incorporate cross-regime parameter equality restrictions as follows:

$$\tilde{y} = X_R \beta_R + \tilde{u}_R, \tag{3}$$

where $X_R = \left[\left(D_m R_{\beta_1}\right) \bullet \tilde{x}_1 \quad \left(D_m R_{\beta_2}\right) \bullet \tilde{x}_2 \quad . \quad . \quad \left(D_m R_{\beta_k}\right) \bullet \tilde{x}_k \right]$, and $\tilde{u}_R \sim N(0_T, I_T \bullet \tilde{s}_R)$ with $\tilde{s}_R = \left(D_m R_\sigma\right)\sigma_R$.

2.2 Prior Specification

In this paper we focus on Bayesian estimation of the model in (3). We begin with specification of prior density functions for the model parameters, conditional on values for *m* and $R = \left\{R_{\beta_1},...,R_{\beta_k},R_\sigma\right\}$. Partition the parameters into three blocks, given by β_R, σ_R, and $\tau = \left(\tau_1,\tau_2,....,\tau_m\right)'$. We assume prior independence of β_R, σ_R, and τ, as well as prior independence of the elements of σ_R. The joint prior is then given by:

$$f\left(\beta_R,\tau,\sigma_R\right) = f(\beta_R) f(\tau) \prod_{i=1}^{q_\sigma} f(\sigma^i). \tag{4}$$

We specify proper priors for each parameter block. We define $f\left(\beta_R\right)$ as a multivariate Gaussian random variable with mean vector C and variance-covariance matrix Σ. For each σ^i, $i = 1,...,q_\sigma$, we specify $f\left(\sigma^i\right)$ as an inverted gamma density function with parameters ν and δ. Finally, our prior for τ is a uniform distribution

over the space of allowable break dates. A set of break dates is allowable if they meet the following set of restrictions:

$$\left(\tau_i - \tau_j\right) \geq b, \ \forall i > j \,, \tag{5}$$

where $b \geq 1$. The constraint (5) requires that a regime have minimum length *b*.

2.3 Bayesian Estimation via the Gibbs Sampler

For given values of *m* and *R*, and the prior density functions defined above, Bayesian estimation of the model in (3) can proceed via the Gibbs Sampler (Gelfand and Smith, 1990). The Gibbs sampler for a time-series regression model with multiple structural breaks is described in Wang and Zivot (2000).[2] We follow closely the Wang and Zivot (2000) algorithm, modified to allow for subsets of parameters that break across some, but not all, break dates. The details of this algorithm are presented in the appendix.

2.4 Model Selection

In practice, the number of breaks, *m*, as well as the subsets of parameters that change across each break date, defined by *R*, are unknown. In the following, a model is determined by values of *m* and *R*, and is denoted as $M(m, R)$. The problem of choosing *m* and *R* is then cast in terms of model comparisons across alternative $M\left(m, R\right)$.

The standard Bayesian approach to model comparison is to compute posterior probabilities of alternative models. In particular, the posterior probability of $M(m, R)$ is given by:

[2] See also Stephens (1994).

$$P(M(m,R)\mid \tilde{y}) = \frac{f(\tilde{y}\mid M(m,R))P(M(m,R))}{f(\tilde{y})}. \qquad (6)$$

In (6), $P(M(m,R))$ is the researcher's prior probability on $M(m,R)$, while $f(\tilde{y}\mid M(m,R))$ is the marginal likelihood, or likelihood function integrated free of model parameters. Finally, $f(\tilde{y})$ is an integrating constant that can be recovered, given $P(M(m,R))$ and $f(\tilde{y}\mid M(m,R))$, from the constraint that $\sum_m \sum_R P(M(m,R)\mid \tilde{y}) = 1$.

Obtaining $P(M(m,R)\mid \tilde{y})$ requires knowledge of $f(\tilde{y}\mid M(m,R))$ and $P(M(m,R))$. To obtain the marginal likelihood, we use the approach of Chib (1995), which provides a simulation consistent estimate of $f(\tilde{y}\mid M(m,R))$ based on the output of full and reduced runs of the Gibbs sampler described in Section 2.3. We have considered alternative approaches to estimating $f(\tilde{y}\mid M(m,R))$, such as those based on importance sampling, and have found that the approach of Chib (1995) performed best in the simulation evidence presented in Section 3.

To specify the prior model probability, we use a flat prior over the number of structural breaks up to a pre-specified maximum, denoted m^*, as well as a flat prior over the different permutations of *R* considered for a given value of *m*, denoted by N_m. That is:

$$P(M(m,R)) = \frac{1}{m^*+1} * \frac{1}{N_m}. \qquad (7)$$

The first term in (7) divides the probability space equally among the $m^* + 1$ potential break models, $m = 0,1,...m^*$, while the second term divides the probability space for a given *m* equally among the N_m potential models.

Given $P(M(m,R)|\tilde{y})$, selection of *m* and *R* proceeds in two steps. In the first step, we choose the number of structural breaks as the value of *m* that solves:

$$\max_{(m=0,1,..,m^*)} \left(P(M(m)|\tilde{y})\right) = \max_{(m=0,1,..,m^*)} \left(\sum_R P(M(m,R)|\tilde{y}) \right). \tag{8}$$

Given a choice for *m*, denoted $\tilde{m}$, in the second step we choose *R* as that value of *R* that solves:

$$\max_{(R)} P(M(\tilde{m},R)|\tilde{y}). \tag{9}$$

An alternative, commonly used approach to model selection is based on information criteria such as the Schwarz Information Criterion (SIC). The SIC was shown to be a consistent criterion for selecting the number of structural breaks in a linear regression with exogenous regressors by Liu et al. (1997), and was shown to perform well at selecting the number of structural breaks in dynamic models by Wang and Zivot (2000). Here we will also consider model selection based on the SIC. In particular, the SIC for $M(m,R)$, denoted $SIC(m,R)$, is given by:

$$SIC(m,R) = -2\ln L\left(\hat{\beta}_R, \hat{\sigma}_R, \hat{\tau}|\tilde{y}\right) + \left(m + q_\sigma + \sum_{j=1}^{k} q_{\beta_j} \right) \ln(T), \tag{10}$$

where hats indicate the maximum likelihood estimates. Due to the substantial computational burden involved with obtaining maximum likelihood estimates for certain structural beak models, we follow Wang and Zivot (2000) and instead evaluate (10) at the median of the posterior distribution for each parameter. A value for *m* and *R* is then chosen as the solution to the following problem:

$$\min_{(m=0,1,..,m^*;R)} SIC\left(M(m,R)\,|\,\tilde{y}\right). \tag{11}$$

2.5 Model Averaging

In many cases, the objective is not to select a particular $M(m,R)$ from the set of possible structural break models, but instead to draw inference on a particular subset of the parameter space that has a common interpretation across models with different *m* and *R*. A distinct advantage of the Bayesian approach is that it allows the researcher to obtain a posterior distribution for this subset of parameters of interest, without conditioning on a particular model.

Specifically, suppose one is interested in a subset of the parameter space, denoted θ^*. The posterior density for θ^*, conditional on only the observed data, is given as:

$$p(\theta^*\,|\,\tilde{y}) = \sum_m \sum_R f\left(\theta^*\,|\,M(m,R),\tilde{y}\right) * P\left(M(m,R)\,|\,\tilde{y}\right), \tag{12}$$

where $f\left(\theta^*\,|\,M(m,R),\tilde{y}\right)$ is the posterior density of θ^* conditional on a particular model, and can be sampled using the Gibbs Sampler as described in section 2.3, while $P\left(M(m,R)\,|\,\tilde{y}\right)$ is the posterior model probability computed as in section 2.4. Operationally, a draw is obtained from $p(\theta^*\,|\,\tilde{y})$ by obtaining a draw from

$f\left(\theta^* \mid M\left(m,R\right),\tilde{y}\right)$ for all *m* and *R*, and then forming a weighted average of these draws, where the weights are given by $P\left(M\left(m,R\right)\mid \tilde{y}\right)$.

2.6 Dimensionality of the Model Space

A practical problem with the model selection procedure outlined above is the proliferating dimensionality of the model space. For example, consider a simple AR(1) model with three parameters, namely an intercept, an autoregressive parameter, and the residual variance parameter. These parameters yield seven potential combinations of parameters that can go undergo a structural break at any particular break point. If the model is allowed to have *m* breaks, and one wants to compare all possible combinations of parameter change for this *m* break model, we have $N_m = 7^m$ potential models to consider. The model space increases even further if we compare across alternative values of *m.*

That being said, given modern computing speeds, it is quite feasible to conduct model selection over all potential models for moderately parameterized regressions, such as univariate autoregressions, and for moderate numbers of breaks. Second, for more highly parameterized models, such as a vector autoregression, model selection might be conducted over parameter blocks, which would reduce the dimensionality of the model space considerably. Finally, although considering a large number of potential models has a measurable cost, the benefit might also be substantial. Indeed, we will demonstrate in the following sections that basing inference regarding structural breaks on a model with breaks in all parameters can lead to very misleading results, even for models with a small numbers of parameters.

3. Simulation Evidence

In this section we describe results from simulation experiments conducted to evaluate the performance of the model selection procedures detailed above. For each simulation experiment, we generate artificial time series from a data generating process that is a member of the following class of models:

$$
\begin{aligned}
&y_t = \beta_0^1 + \beta_1^1 x_t + \sigma^1 \varepsilon_t && \tau_0 + 1 \le t \le \tau_1, \\
&y_t = \beta_0^2 + \beta_1^2 x_t + \sigma^2 \varepsilon_t && \tau_1 + 1 \le t \le \tau_2, \\
&. && . \\
&. && . \\
&y_t = \beta_0^{m+1} + \beta_1^{m+1} x_t + \sigma^{m+1} \varepsilon_t && \tau_m + 1 \le t \le \tau_{m+1}, \\
&t = 1,....,T ,
\end{aligned}
\tag{13}
$$

where $\varepsilon_t \sim i.i.d\ N(0,1)$, x_t is scalar and follows a first-order autoregressive process with standard normal innovations and autoregressive parameter ρ, and the sample size is T = 200. The alternative data generating processes considered differ according to the number of structural breaks, the subset of the parameter vector that is allowed to break at each break point, the size of the structural breaks, and the autoregressive parameter ρ. For each simulated time series, we then select a particular model from (13), where model selection is conducted over all possible variations for m, R_{β_0}, R_{β_1} and R_σ. The maximum number of structural breaks considered for model selection is m^*=2, yielding a total of 57 models to compare for each generated time series, one corresponding to the $m = 0$ case, seven corresponding to the $m = 1$ case, and 49 corresponding to the $m = 2$ case. For each $M(m, R)$, we compute the posterior probability of the model, $P(M(m, R))$, as well as the SIC, $SIC(M(m, R))$, and select a model as described in Section 2.4.

As a means of comparison, we also consider the ability of procedures that do not conduct model selection over *R* to select the correct number of structural breaks. For these procedures, only three potential models are compared, one for each value of *m* under consideration. The model for each *m* sets R_{β_0}, R_{β_1} and R_{σ} equal to the $m+1$ identify matrix, thus allowing each of β_0, β_1 and σ to break at each of the *m* break dates. For the posterior odds based procedure, a value for *m* is chosen as the *m* that yields the highest posterior model probability, where we again specify a uniform prior over *m.* For the SIC procedure, m is chosen as the *m* that yields the minimum SIC. We refer to these procedures that do not conduct model selection over *R* as the "baseline posterior odds" and "baseline SIC" procedures in the following, while the procedures that do conduct model selection over *R* are referred to as the "preferred posterior odds" and "preferred SIC" procedures.

We specify the following parameter prior distributions for use in the simulation experiments. The prior mean and variance-covariance matrix for β_R, given by C and Σ, are set equal to a vector of zeros and the identify matrix respectively, implying that each element of β_R has a standard normal prior distribution and is independent of all other elements of β_R. The prior for each residual variance parameter, σ_i, is inverted gamma with parameters $\nu = 2.001$ and $\delta = 1$. Finally, the minimum regime length considered is set to $b = 12$, which is 6% of the sample size. Results for each model are based on 10,000 draws from the Gibbs Sampler after an initial 5000 "burn-in" draws to obtain convergence. All results reported below are based on 500 generated time series.

In the first set of simulations, we evaluate the performance of the model selection procedures when the true model has no structural breaks. In particular, we generate time

series from (13) with $m = 0$. It is well known that the performance of frequentist-based tests for structural breaks is quite sensitive to the persistence of regressors, with tests being severely oversized for high levels of persistence (Diebold and Chen, 1996). To evaluate the sensitivity of the procedures developed here to persistent regressors we consider three possible calibrations of ρ, corresponding to low, moderate and high persistence, and given by $\rho = 0.3$, $\rho = 0.6$ and $\rho = 0.9$. For each value of ρ, we set $\beta_0 = \beta_1 = \sigma = 1$.

Table 1 records the proportion of simulations for which the indicated model selection procedure chose the $m = 0$ model. There are several items of particular interest in these results. First, the preferred posterior odds procedure performs quite well when the true data generating process has constant parameters, selecting an (incorrect) model that includes a structural break in only about 2% of the simulations. For the preferred SIC procedure these proportions are lower, but still above 90%. Second, the tendency for each of the model-selection procedures to falsely identify a structural break is largely unaffected by the persistence of the regressor in the simulated data. Finally, the performance of the preferred posterior odds procedure is similar to that for the baseline posterior odds procedure. That the preferred procedure does not improve on the baseline procedure in this case is not surprising, as there is no reason to expect that conducting model selection over both *m* and *R* would have an advantage in the case where the true model does not contain a structural break. However, there does not appear to be any noticeable disadvantage to using the preferred approach to model selection either.

In the second set of simulations, we evaluate the performance of the model selection procedures when the true model has a single structural break. In particular, we generate

data from the model in (13) with $m = 1$. We consider six potential data generating processes. In the first two, we generate data from a model with a break in β_0 only. We consider both small and large breaks, where these break sizes are calibrated by selecting values of β_0^1 and β_0^2 for which the baseline posterior odds procedure selects the correct value of *m* approximately 25% and 75% of the time respectively.[3] These values are given by $\{\beta_0^1 = 0, \beta_0^2 = 0.41\}$ for the small break case and $\{\beta_0^1 = 0, \beta_0^2 = 0.63\}$ for the large break case. For each case, we set $\beta_1 = \sigma = 1$. In the second set of data generating processes, we consider a model with a break in β_1 only. The break is again calibrated as being either small or large as described above, and parameterized as $\{\beta_1^1 = 0, \beta_1^2 = 0.32\}$ for the small break case and $\{\beta_1^1 = 0, \beta_1^2 = 0.53\}$ for the large break case. For each case, we set $\beta_0 = \sigma = 1$. Finally, we consider two cases for which there is a break in σ only. We again consider small or large breaks, parameterized as $\{\sigma^1 = 0.73, \sigma^2 = 1\}$ for the small break case, and $\{\sigma^1 = 0.63, \sigma^2 = 1\}$ for the large break case. For each case, we set $\beta_0 = \beta_1 = 1$. For all cases, we set the break date, τ_1, in the middle of the sample period, so that τ_1 = 100. The autoregressive parameter for the regressor, ρ, is set equal to 0.6.

Table 2 records the proportion of simulations for which the indicated model selection procedure chose the $m = 1$ model. For the preferred posterior odds and SIC procedures, the table also records the proportion of simulations for which the model selection procedure chose both *m* and *R* correctly. A primary result to emphasize from Table 2 is that conducting model selection over both *m* and *R* yields substantial improvements, in

[3] Specifically, the percentage of simulations for which the baseline Bayesian procedure selects the correct value for *m* is within one percentage point of 25% for the small break size and 75% for the large break size.

terms of the frequency with which the correct value of *m* is chosen, over the baseline procedures conducted over *m* only. For the small break case, where the baseline posterior odds procedure selects the correct value of *m* in 25% of the simulations, the preferred posterior odds procedure selects the correct value of *m* in well above 40% of the simulations in all cases. This improvement is even more substantial for SIC-based model selection. In this case, the baseline procedure selects the correct value of *m* in 20% or less of the simulations, while the preferred procedure selects the correct value of *m* in around 50% of the simulations. Not surprisingly, the improvements generated by the preferred procedure become smaller when we consider larger breaks. However, the improvement is still substantial, on the order of 10-15 percentage points for the posterior odds procedures and 20-25 percentage points for the SIC procedures. Table 2 also demonstrates that the preferred procedures select the correct values of *m* and *R* close to as often as the correct value of *m* only, suggesting that the procedure is reasonably successful at identifying the correct subset of parameters that break at each break date.

In the final set of simulations, we evaluate the performance of the model selection procedures when the true model has two structural breaks. In particular, we generate data from the model in (13) with $m = 2$. We again consider six potential data generating processes. In the first two, we generate data from a model with two breaks in β_0 only. We again consider both small and large breaks, calibrated as discussed above for the case of a single structural break. These values are given by $\{\beta_0^1 = 0, \beta_0^2 = 0.67, \beta_0^3 = 0\}$ for the small break case and $\{\beta_0^1 = 0, \beta_0^2 = 0.91, \beta_0^3 = 0\}$ for the large break case. For each case, we set $\beta_1 = \sigma = 1$. For the second set of data generating processes, we consider a model with two breaks in β_1 only. Here, small and large breaks are parameterized as

$\{\beta_1^1 = 0, \beta_1^2 = 0.53, \beta_1^2 = 0\}$ and $\{\beta_1^1 = 0, \beta_1^2 = 0.74, \beta_1^2 = 0\}$ respectively. For each case, we set $\beta_0 = \sigma = 1$. Finally, we consider two cases for which there are two breaks in σ only, with small and large breaks parameterized as $\{\sigma^1 = 0.62, \sigma^2 = 1, \sigma^3 = 0.62\}$ and $\{\sigma^1 = 0.51, \sigma^2 = 1, \sigma^3 = 0.51\}$ respectively. For each case, we set $\beta_0 = \beta_1 = 1$. For all simulations, we space the break dates, τ_1 and τ_2, equally throughout the sample at dates 67 and 134 respectively. The autoregressive parameter for the regressor, ρ, is set equal to 0.6.

Table 3 records the proportion of simulations for which the indicated model selection procedure chose the $m = 2$ model. For the preferred posterior odds and SIC procedures, the table also records the proportion of simulations for which the model selection procedure chose both *m* and *R* correctly. The results in Table 3 give a similar message to those in Table 2 for the case of a single structural break. Specifically, in all cases considered, the preferred posterior odds procedure selects the correct value of *m* substantially more often than the baseline procedure. This improvement is even larger than was the case for a single structural break, which is not surprising given that the number of unnecessary parameter breaks allowed by the baseline procedure grows in the simulations with two breaks from those with a single break. Similar results are also obtained for the SIC procedures, although, again, the benefits to conducting model selection over *R* in addition to *m* are more pronounced. Finally, Table 3 again demonstrates that the preferred procedures are fairly successful at selecting the correct value of both *m* and *R*, and thus the correct subset of parameters that break at each break date.

In summary, the results from these simulation exercises are suggestive that model selection conducted over both the number of structural breaks and the subset of parameters that change at each break date can yield important benefits over model selection procedures in which all parameters are allowed to break at each break date. One such benefit is an improved frequency with which the correct number of structural breaks are chosen, which, as would be expected, is particularly the case as the total number of potential parameter breaks grows relative to the total number of actual parameter breaks. Another benefit is that the model selection procedures conducted over both m and R provide some reliable information regarding the subset of parameters that undergo breaks at each break date, information that is absent from procedures that simply allow all parameters to change at each break date.

4. Application to U.S. Inflation Dynamics

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 quite low prior to the "great inflation" and after the Volcker disinflation, while being quite high between these episodes. The Cogley and Sargent results were 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. Other authors, such as Levin and Piger (2002), have argued that there are important structural breaks in the intercept parameter of an autoregression for

inflation, and that allowing for such breaks is important to properly characterize inflation persistence. 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 Bayesian model selection procedures described above to evaluate the evidence for structural breaks 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 Deflator, sampled from the first quarter of 1953 through the second quarter of 2005. A plot of this data is shown in Figure 1. We are particularly interested in four questions: 1) Is there strong statistical evidence of structural change in the parameters of an autoregression fit to the U.S. inflation rate? 2) If so, is there evidence of structural change in conditional mean parameters or only in residual variance parameters? 3) If there is evidence of changes in conditional mean parameters, is there evidence for changes in the persistence of the inflation process? 4) Does allowing for structural change in the parameters of an autoregression for inflation alter our estimates of inflation persistence?

We fit the following k^{th}-order autoregressive model with potential structural breaks in intercept, autoregressive parameters, and residual variance to the inflation rate y_t:

$$
\begin{aligned}
&y_t = \alpha^1 + \rho_1^1 y_{t-1} + \rho_2^1 y_{t-2} + ... + \rho_k^1 y_{t-k} + \sigma^1 \varepsilon_t && \tau_0 + 1 \le t \le \tau_1, \\
&y_t = \alpha^2 + \rho_1^2 y_{t-1} + \rho_2^2 y_{t-2} + ... + \rho_k^2 y_{t-k} + \sigma^2 \varepsilon_t && \tau_1 + 1 \le t \le \tau_2, \\
&. && . \\
&. && . \\
&y_t = \alpha^{m+1} + \rho_1^{m+1} y_{t-1} + \rho_2^{m+1} y_{t-2} + ... + \rho_k^{m+1} y_{t-k} + \sigma^{m+1} \varepsilon_t && \tau_m + 1 \le t \le \tau_{m+1}, \\
&t = 1,...., T ,
\end{aligned}
\qquad (14)
$$

A key quantity of interest in the model in (14) is the persistence of y_t, or the extent to which an innovation, ε_t, has long-lived effects on the level of the inflation rate. We measure persistence via the sum of the autoregressive coefficients, denoted $\rho^i = \sum_{j=1}^{k} \rho_j^i$. As discussed in Pivetta and Reis (2007), for $\left|\rho^i\right| < 1$, $1/\left(1-\rho^i\right)$ gives the area under the impulse response function, and is thus an intuitively appealing measure of the persistence of a time-series process.

To estimate ρ^i directly, we rewrite (14) using the Dickey-Fuller transformation:

$$
\begin{aligned}
&y_t = \alpha^1 + \rho^1 y_{t-1} + \gamma_1^1 \Delta y_{t-1} + ... + \gamma_{k-1}^1 \Delta y_{t-(k-1)} + \sigma^1 \varepsilon_t && \tau_0 + 1 \le t \le \tau_1, \\
&y_t = \alpha^2 + \rho^2 y_{t-1} + \gamma_1^2 \Delta y_{t-1} + ... + \gamma_{k-1}^2 \Delta y_{t-(k-1)} + \sigma^2 \varepsilon_t && \tau_1 + 1 \le t \le \tau_2, \\
&. && . \\
&. && . \\
&y_t = \alpha^{m+1} + \rho^{m+1} y_{t-1} + \gamma_1^{m+1} \Delta y_{t-1} + ... + \gamma_{k-1}^{m+1} \Delta y_{t-(k-1)} + \sigma^{m+1} \varepsilon_t && \tau_m + 1 \le t \le \tau_{m+1}, \\
&t = 1,....,T ,
\end{aligned}
\qquad (15)
$$

where $\gamma_s^i = -\left(\sum_{j=s+1}^{k} \left(\rho_j^i\right) \right)$ and $\varepsilon_t \sim i.i.d\ N(0,1)$. In the notation of the model in (1), we have $x_t = \left(1, y_{t-1}, \Delta y_{t-1}, ..., \Delta y_{t-(k-1)}\right)'$ and $\beta^i = \left(\alpha^i, \rho^i, \gamma_j^i, ..., \gamma_{k-1}^i\right)'$, $i = 1,..., m+1$.

We use the following parameter prior specifications for the model in (15). For given values of *m*, *R*, and *k*, we have the vector of conditional mean parameters β_R. The elements of this vector are assigned prior independence, and each is given a Gaussian prior distribution with variance equal to 1 and mean equal to 1 for α^i and ρ^i and 0 for γ^i, $i = 1,..., m+1$. Each σ^i is assigned an inverted gamma prior distribution with

parameters $\nu = 2.001$ and $\delta = 1$. Finally, the minimum regime length is given by $b = 12$ quarters. Results are based on 10,000 draws of the Gibbs Sampler after an initial 5000 "burn-in" draws to obtain convergence.

To conduct model selection for the model in (15), we use the preferred posterior odds procedure outlined in Section 2.4. In particular, we consider all possible combinations of shifts in the intercept parameter, the autoregressive parameters, and the residual variance parameter at each potential break date. To economize on the model space considered, we consider the autoregressive parameters as a single block and thus assume that these parameters undergo structural breaks together. In addition, we extend the Bayesian model selection procedure to allow for model selection over the lag length, *k*. In particular, a model is defined by values for *m*, *R*, and *k*, denoted $M(m, R, k)$. We augment our prior over models with a flat prior over *k*, where we consider a maximum lag length of k^*. That is, we have:

$$P(M(m, R, k)) = \frac{1}{m^* + 1} * \frac{1}{N_m} * \frac{1}{k^*}. \tag{16}$$

We then conduct model selection using the posterior model probabilities, $P(M(m, R, k) \mid \tilde{y})$. As a means of comparison, we also generate posterior model probabilities from our baseline posterior odds procedure, where we assume that all parameters break at each potential break date, and thus do not conduct model selection over *R* , but instead over *m* and *k* only. Finally, for all the results presented below, we set $k^* = 4$ and $m^* = 4$.

Table 4 presents the posterior probabilities for alternative values of the number of structural breaks, *m*, defined as $\sum_{k}\sum_{R} P\left(M(m,R,k) \mid \tilde{y}\right)$. Table 4 also presents posterior probabilities for alternative *m* obtained from the baseline posterior odds procedure**.** The results for the preferred posterior odds procedure demonstrate that there is overwhelming evidence for structural breaks, and that $m = 3$ is the preferred number of structural breaks. The results for the baseline posterior odds procedure also suggest overwhelming evidence for structural breaks, although the chosen number of breaks is four rather than three.

Next we move to evaluating the nature of this structural change. We are first interested in evaluating the claim that structural breaks in the inflation rate are confined to structural breaks in the residual variance parameter only, and do not extend to the conditional mean parameters. Table 5 compares the posterior probability of models with structural breaks in only residual variance with the posterior probability of models that contain structural breaks in conditional mean parameters. These probabilities are computed using two procedures. The first is analogous to our "baseline" posterior odds procedure, and compares the posterior probability of models that contain only residual variance breaks with the posterior probability of models that contain breaks in all conditional mean parameters. The second is our preferred posterior odds procedure, which will differ from the baseline procedure by considering models in which only a subset of the conditional mean parameters are allowed to change at each break date.

As Table 5 makes clear, when the baseline procedure is used, the posterior probability that the model contains only breaks in residual variance dominates the posterior probability that the model contains structural breaks in conditional mean parameters.

However, when the preferred procedure is used, these probabilities reverse, with models containing structural breaks in conditional mean parameters dominating those with structural breaks in only residual variance. [4] Put simply, the preferred procedure reveals high posterior probability models, not considered by the baseline procedure, in which only a subset of the conditional mean parameters break. These results demonstrate that model selection over *R* can be crucial for evaluating the evidence for alternative types of parameter breaks.

What is the nature of the structural breaks in conditional mean parameters? We are particularly interested in whether there is evidence in favor of shifts in autoregressive parameters, or if the breaks in conditional mean parameters are confined to intercept shifts. An advantage of conducting model selection over *R* is that it allows us to provide evidence on this question. In particular, we begin by restricting the model space to only those models that have a structural break in conditional mean parameters. We then divide this model space into those models that do and do not contain structural breaks in autoregressive parameters and construct the posterior probability for each class of models. The results suggest that the models with breaks in autoregressive parameters are given 33% posterior probability, while the models without breaks in autoregressive parameters are given 67% posterior probability. Thus, while there does not seem to be

[4] For a given value for the number of structural breaks, *m*, and lag length, *k*, the prior model probabilities in (16) give equal prior probability to each possible choice of which parameters are allowed to break, or *R*. As for each value of *m* and *k* there is only one model that has breaks in only the residual variance parameter, the prior probability assigned to models with only variance breaks shrinks relative to the prior probability assigned to models with breaks in conditional mean parameters as *m* or *k* increases. To assess the sensitivity of the results in Table 5 to this prior, we computed posterior probabilities using an alternative prior that assigns equal weight to the class of models with breaks in only residual variance vs. the class of models with breaks in conditional mean parameters. Under this prior, the results in Table 5 are qualitatively similar. In particular, the class of models with breaks in conditional mean parameters are assigned substantially more posterior weight than the class of models with breaks in only the residual variance parameter.

strong evidence in favor of shifts in autoregressive parameters, the data is not speaking strongly against such shifts either. This suggests that while there is clear evidence in favor of shifts in conditional mean parameters, the evidence is not clear as to whether these shifts include breaks in autoregressive parameters, and thus in the persistence of the inflation process.

We now study those models for which the number of structural breaks is equal to the chosen value of $m = 3$ in more detail. Table 6 contains specification and estimation details for the highest posterior probability models with $m = 3$. In particular, each model presented in Table 6 has a posterior probability that is no less than $1/10^{th}$ that of the most preferred model with $m = 3$. Several conclusions can be drawn from these results. First, there are a large number and variety of models in Table 6, suggesting that the data does not speak definitively about the exact form of the preferred model with three structural breaks. Second, the timing of the structural breaks generally fall into one of two categories. In the first, the structural breaks occur in the late 1960s, early 1970s and early 1980s, while in the second the structural breaks occur in the late 1960s, early 1980s and early 1990s. Third, structural breaks in the intercept and residual variance parameters appear to be a dominant feature of the data. In particular, of the 25 models presented in Table 6, all allow for at least two breaks in residual variance, while 13 allow for three breaks in residual variance. Correspondingly, 24 of the models allow for at least two shifts in intercept, while 12 allow for three shifts in intercept. Fourth, there is less evidence of any shifts in autoregressive parameters. Specifically, only 7 of the 25 models in Table 6 allow for a shift in the autoregressive parameters.

When the data does not definitely select an exact model specification, as is the case in Table 6, one would likely be hesitant to base conclusions about key parameters on the results from any one most preferred specification. An advantage of the Bayesian procedures employed in this paper is the ability to characterize key parameters of interest without conditioning on a particular value for *m*, *R*, or *k*. For example, we might be interested in estimating the sum of the autoregressive coefficients at each quarter in the sample, defined as ρ_t. A posterior distribution for this quantity that is averaged over potential values for *m*, *R* and *k* is given by:

$$f(\rho_t \mid \tilde{y}) = \sum_k \sum_m \sum_R f\left(\rho_t \mid M(m,R,k), \tilde{y}\right) * P\left(M(m,R,k) \mid \tilde{y}\right). \qquad (17)$$

For comparison purposes, we will also be interested in an estimate of ρ_t conditional on a particular value for *m*, given by:

$$f(\rho_t \mid m, \tilde{y}) = \sum_k \sum_R f\left(\rho_t \mid M(m,R,k), \tilde{y}\right) * P\left(M(m,R,k) \mid \tilde{y}\right). \qquad (18)$$

Figure 2 presents the 5th, 50th and 95th percentile of $f(\rho_t \mid m = 0, \tilde{y})$, which is the estimate of the persistence of the inflation process assuming there have been no structural breaks. The estimated persistence of the inflation process is quite high, with a median posterior value above 0.9 and 95th posterior percentile approaching one. This is consistent with a large existing literature documenting high inflation persistence when measured using a constant parameter autoregression over the post-war period (e.g. Nelson and Plosser, 1982; Fuhrer and Moore, 1995). Figure 3 instead presents the 5th, 50th and 95th percentile of $f(\rho_t \mid \tilde{y})$, that is the estimate of persistence allowing for

structural breaks. There are at least three items of interest in Figure 3. First of all, the median estimate of the persistence process is largely constant. Second, the median estimate of the persistence process is substantially lower than that obtained for the constant parameter autoregression presented in Figure 2. For example, over the entire sample, the median estimate is below the 5th percentile of the posterior for the constant parameter autoregression. This lowered persistence comes from allowing for intercept shifts in the model with structural breaks. Finally, the uncertainty surrounding the persistence parameter has become quite wide over the last two decades, with the 95% highest posterior density interval spanning from 0.2 to 0.9 toward the end of the sample.

It is also interesting to characterize the posterior distribution of the residual standard deviation at each quarter in the sample, denoted σ_t. Figure 4 presents the 5th, 50th and 95th percentile of $f(\sigma_t \mid \tilde{y})$, that is the estimate of residual standard deviation averaging over different values of *m*, *R*, and *k*. This residual standard deviation has been far from constant. In particular it has varied from a low volatility regime in the 1950s and much of the 1960s to a high volatility regime from the late-1960s to the early 1980s, before returning to a low volatility regime beginning in the early 1980s.

In summary, these results suggest that there have been important structural breaks in both the conditional mean parameters and conditional variance parameters of an autoregression for post-war U.S. inflation. We find that the evidence for shifts in conditional mean parameters are only revealed once we conduct model selection over not just the number of structural breaks, but also the subset of parameters that break at each break date, demonstrating the empirical relevance of the model selection procedures developed in this paper. Finally, we find no strong evidence for time variation in the

persistence of U.S. inflation. However, the estimates of persistence that we obtain are substantially lower than those obtained in the existing literature using constant parameter autoregressions, as these models ignore important shifts in the intercept parameter of the inflation autoregression.

5. Conclusion

We have developed a Bayesian approach to model selection in regression models with structural breaks in conditional mean and residual variance parameters. A novel feature of our approach is that it does not assume knowledge of the parameter subset that undergoes structural breaks, but instead conducts model selection jointly over the number of structural breaks and the subset of the parameter vector that changes at each break date. Simulation experiments suggest there are potentially sizeable gains for break detection to conducting model selection over the subset of parameters that undergo breaks rather than simply allowing all parameters to change.

We apply the proposed model selection procedure to characterize possible structural breaks in conditional mean and residual variance parameters in an autoregressive model for the U.S. inflation rate. We find substantial evidence for changes in both residual variance and conditional mean parameters, the latter of which is revealed only when one conducts model selection over the subset of conditional mean parameters that undergo breaks, a result that demonstrates the empirical relevance of our proposed model selection procedure. We obtain estimates of inflation persistence that are substantially lower than those obtained assuming a constant parameter autoregression.

Appendix: Gibbs Sampling Algorithm for Simulating the Posterior of the Regression Model with Structural Breaks

This appendix describes the Gibbs Sampling algorithm used to simulate the posterior distribution of model parameters for the regression model with multiple structural breaks given in equation 3, conditional on the prior distributions given in Section 2.2. The Gibbs sampling algorithm proceeds by drawing iteratively from the full set of conditional posterior densities in the following steps:

A) Generate a draw of β_R, denoted $\overline{\overline{\beta}}_R$, from $f(\beta_R \mid \overline{\sigma}_R, \overline{\tau}, \tilde{y})$

B) Generate a draw of σ_R, denoted $\overline{\overline{\sigma}}_R$, from $f(\sigma_R \mid \overline{\overline{\beta}}_R, \overline{\tau}, \tilde{y})$

C) Generate a draw of τ, denoted $\overline{\overline{\tau}}$, using the following *m* steps:

C1) Generate a draw of τ_1, denoted $\overline{\overline{\tau}}_1$, from $f(\tau_1 \mid \overline{\overline{\beta}}_R, \overline{\overline{\sigma}}_R, \overline{\tau}_2, ..., \overline{\tau}_m, \tilde{y})$

C2) Generate a draw of τ_2, denoted $\overline{\overline{\tau}}_2$, from $f(\tau_2 \mid \overline{\overline{\beta}}_R, \overline{\overline{\sigma}}_R, \overline{\overline{\tau}}_1, \overline{\tau}_3, ..., \overline{\tau}_m, \tilde{y})$

.

.

Cm) Generate a draw of τ_m, denoted $\overline{\overline{\tau}}_m$, from $f(\tau_m \mid \overline{\overline{\beta}}_R, \overline{\overline{\sigma}}_R, \overline{\overline{\tau}}_1, ..., \overline{\overline{\tau}}_{m-1}, \tilde{y})$

D) Set $\overline{\sigma}_R = \overline{\overline{\sigma}}_R$ and $\overline{\tau} = \overline{\overline{\tau}}$ and repeat steps A-D.

The algorithm is initiated with arbitrary values of $\overline{\sigma}_R$ and $\overline{\tau}$. Assuming certain regularity conditions are met (Tierney, 1994) draws from this algorithm will converge to draws from the posterior density of interest $f(\beta_R, \sigma_R, \tau \mid \tilde{y})$, which can be used to form point estimates and highest posterior density intervals.

Given the Gaussian likelihood function and conjugate priors, draws from $f(\beta_R \mid \sigma_R, \tau, \tilde{y})$ and $f(\sigma_R \mid \beta_R, \tau, \tilde{y})$ are straightforward. Specifically, conditional on τ,

the model in (3) is a linear regression model with dummy variables. The conditional posterior density function for β_R is then:

$$f(\beta_R \mid \sigma_R, \tau, \tilde{y}) \sim \tag{A1}$$

$$N\left(\left(\Sigma^{-1} + X_R^{'}(I_T \bullet \tilde{s}_R)^{-2} X_R\right)^{-1}\left(\Sigma^{-1}C + X_R^{'}(I_T \bullet \tilde{s}_R)^{-2} \tilde{y}\right), \left(\Sigma^{-1} + X_R^{'}(I_T \bullet \tilde{s}_R)^{-2} X_R\right)^{-1}\right),$$

from which samples can be easily obtained via draws from a multivariate random normal density. It can also be shown that $f(\sigma_R \mid \beta_R, \tau, \tilde{y}) = \prod_{i=1}^{q_\sigma} f(\sigma^i \mid \beta_R, \tau, \tilde{y})$, and:

$$f(\sigma^i \mid \beta_R, \tau, \tilde{y}) \sim IG\left(\nu + \tau_i - \tau_{i-1}, \delta + (\tilde{y}^i - X_R^i \beta_R)'(\tilde{y}^i - X_R^i \beta_R)\right), \tag{A2}$$

where $\tilde{y}^i$ and X_R^i hold the $\tau_{i-1}+1$ through τ_i rows of $\tilde{y}$ and X_R respectively. A draw from $f(\sigma_R \mid \beta_R, \tau, \tilde{y})$ is then generated as q_σ draws from $f(\sigma^i \mid \beta_R, \tau, \tilde{y})$, $i = 1, \ldots, q_\sigma$.

To complete the Gibbs-sampling algorithm, we require draws from $f(\tau_i \mid \beta_R, \sigma_R, \tau_{\neq i}, \tilde{y})$, which is given by:

$$f(\tau_i \mid \beta_R, \sigma_R, \tau_{\neq i}, \tilde{y}) = \frac{L(\tau_i \mid \beta_R, \sigma_R, \tau_{\neq i}, \tilde{y}) f(\tau_i \mid \beta_R, \sigma_R, \tau_{\neq i})}{f(\tilde{y} \mid \beta_R, \sigma_R, \tau_{\neq i})}. \tag{A3}$$

In A3, $L(\tau_i \mid \beta_R, \sigma_R, \tau_{\neq i}, \tilde{y})$ is the model likelihood function for τ_i holding all other parameters fixed. Given the uniform prior, $f(\tau)$, and the fact that the denominator of (A3) does not depend on τ_i we have:

$$f(\tau_i \mid \beta_R, \sigma_R, \tau_{\neq i}, \tilde{y}) \propto L(\tau_i \mid \beta_R, \sigma_R, \tau_{\neq i}, \tilde{y}). \tag{A4}$$

Expression (A4) can be simplified further by noting that:

$$L(\tau_i \mid \beta_R, \sigma_R, \tau_{\neq i}, \tilde{y}) \propto L(\tau_i \mid \beta_R, \sigma_R, \tau_{i-1}, \tau_{i+1}, y_{\tau_{i-1}+1}, ..., y_{\tau_{i+1}}). \quad \text{(A5)}$$

Thus, the conditional posterior distribution function for τ_i is a multinomial distribution defined over the admissible range for τ_i, given by $\tau_{i-1} + b, ..., \tau_{i+1} - b$, with probabilities proportional to the likelihood function for τ_i evaluated using only data from $\tau_{i-1} + 1, ..., \tau_{i+1}$.

Note that steps C1-Cm in the Gibbs Sampler above could be replaced with a single draw from $f(\tau \mid \beta_R, \sigma_R, \tilde{y})$, for which it is straightforward to show:

$$f(\tau \mid \beta_R, \sigma_R, \tilde{y}) \propto L(\tau \mid \beta_R, \sigma_R, \tilde{y}). \quad \text{(A6)}$$

However, in practice, obtaining a draw from (A6) can be very computationally intensive, as it requires evaluating the likelihood function at all admissible τ. For example, for T=200, $m = 3$ and $b = 1$, each draw from (A6) requires over 1 million evaluations of the likelihood function. By contrast obtaining a draw of τ from (A4) requires less than $m * T$ evaluations of the likelihood function.[5]

[5] Chib (1998) develops an alternative, computationally efficient, approach to draw the entire vector of break dates simultaneously, which is based on modeling the structural break process as an m+1 regime Markov-switching process with absorbing states.

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., and 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. and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica,* 66, 47-78.

Boivin, J. (1999), "Revisiting the Evidence on the Stability of Monetary VAR's," manuscript, Columbia University.

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.

Clarida, R., Gali, J., and Gertler, M. (2000), "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics,* 115, 147-180.

Cogley, T., Sargent, T. (2001), "Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual 2001,* 331-372.

Cogley, T. and 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.

Diebold, F. X. and Chen, C. (1996), "Testing Structural Stability with Endogenous Breakpoint: A Size Comparison of Analytic and Bootstrap Procedures," *Journal of Econometrics,* 70, 221-241.

Fuhrer, J., and Moore, G. (1995), "Inflation Persistence," *Quarterly Journal of Economics,* 110, 127-159.

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

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

Kim, C.-J., and Nelson, C. R. (1999), "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-switching Model of the Business Cycle," *The Review of Economics and Statistics*, 81, 608-616.

Levin, A. and Piger, J. (2002), "Is Inflation Persistence Intrinsic In Industrial Economies?," Federal Reserve Bank of St. Louis Working Paper #2002-023.

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

McConnell, M. M. and Quiros, G. P. (2000), "Output fluctuations in the United States: What has Changed Since the Early 1980s?," *American Economic Review*, 90, 1464-1476.

Nelson, C.R. and Plosser, C.I. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139-162.

Pivetta, F., and Reis, R. (2007), "The Persistence of Inflation in the United States," *Journal of Economic Dynamics and Control*, 31, 1326-1358.

Rapach, D. E. and 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.

Sims, C. A. (1999), "Drift and Breaks in Monetary Policy," manuscript, Princeton University.

Sims, C. A. (2001), "Comment on Sargent and Cogley's 'Evolving U.S. Postwar Inflation Dynamics'," *NBER Macroeconomics Annual 2001*, 379-387.

Sims, C. A. and Zha, T. (2006), "Were there Regime Switches in US Monetary Policy?," *American Economic Review*, 96, 54-81.

Stephens, D. A. (1994), "Bayesian Retrospective Multiple Changepoint Identification," *Applied Statistics*, 43, 159-178.

Stock, J. (2001), "Comment on Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual 2001*, 379-387.

Stock, J. H. and Watson, M. W. (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11-30.

Summers, P. M. (2004), "Bayesian Evidence on the Structure of Unemployment," *Economics Letters*, 83, 299-306.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701-1762.

Wang, J. and Zivot, E. (2000), "A Bayesian Time Series Model of Multiple Structural Changes in Level, Trend, and Variance," *Journal of Business and Economic Statistics*, 18, 374-86.

Table 1
Simulation Results: No Structural Breaks

	Low Persistence	Moderate Persistence	High Persistence
Preferred Procedures			
Posterior Odds	97.8	98.8	98.0
SIC	92.4	93.6	90.8
Baseline Procedures			
Posterior Odds	98.8	99.4	99.6
SIC	99.8	99.6	99.6

Notes: This table holds the proportion of 500 simulations for which the indicated model selection procedure selected the correct value of $m = 0$ structural breaks when the data generating process is given by equation (13). "Preferred Procedures" indicate model selection is conducted over both the number of structural breaks and the subset of the parameter vector that changes at each break date. "Baseline Procedures" indicate model selection is conducted over the number of structural breaks only, with all parameters allowed to change at each break date. "Low Persistence", "Moderate Persistence" and "High Persistence" refer to the persistence of the regressor in equation (13).

Table 2
Simulation Results: One Structural Break

	Small Break		Large Break	
Break in β_0	***Correct Number of Breaks***	***Correct Model***	***Correct Number of Breaks***	***Correct Model***
Preferred Procedures				
Posterior Odds	46.2	39.2	85.6	77.6
SIC	50.6	43.6	86.2	80.6
Baseline Procedures				
Posterior Odds	26.0	NA	75.8	NA
SIC	16.4	NA	63.6	NA
Break in β_1				
Preferred Procedures				
Posterior Odds	42.0	36.2	86.2	77.6
SIC	49.2	43.4	88.0	82.8
Baseline Procedures				
Posterior Odds	24.0	NA	74.8	NA
SIC	13.8	NA	62.0	NA
Break in σ				
Preferred Procedures				
Posterior Odds	56.0	53.0	91.4	88.4
SIC	55.4	47.8	87.6	84.0
Baseline Procedures				
Posterior Odds	24.0	NA	74.0	NA
SIC	20.2	NA	68.2	NA

Notes: This table holds the proportion of 500 simulations for which the indicated model selection procedure selected the correct value of $m = 1$ structural break (given in the column labeled "Correct Number of Breaks"), as well as the correct subset of the parameter vector that changes at the break date (given in the column labeled "Correct Model"). The data generating process for the simulated data is given by equation (13). "Preferred Procedures" indicate model selection is conducted over both the number of structural breaks and the subset of the parameter vector that changes at each break date. "Baseline Procedures" indicate model selection is conducted over the number of structural breaks only, with all parameters allowed to change at each break date.

Table 3
Simulation Results: Two Structural Breaks

	Small Breaks		Large Breaks	
Break in β_0	***Correct Number of Breaks***	***Correct Model***	***Correct Number of Breaks***	***Correct Model***
Preferred Procedures				
Posterior Odds	55.4	42.2	92.8	73.4
SIC	67.2	58.0	96.0	86.0
Baseline Procedures				
Posterior Odds	25.0	NA	75.2	NA
SIC	11.2	NA	49.2	NA
Break in β_1				
Preferred Procedures				
Posterior Odds	48.8	36.6	89.8	71.4
SIC	57.2	47.4	93.4	82.4
Baseline Procedures				
Posterior Odds	25.2	NA	74.8	NA
SIC	8.4	NA	50.0	NA
Break in σ				
Preferred Procedures				
Posterior Odds	65.8	60.6	94.8	81.4
SIC	69.8	59.0	97.8	88.2
Baseline Procedures				
Posterior Odds	25.2	NA	75.4	NA
SIC	17.8	NA	75.6	NA

Notes: This table holds the proportion of 500 simulations for which the indicated model selection procedure selected the correct value of $m = 2$ structural breaks (given in the column labeled "Correct Number of Breaks"), as well as the correct subset of the parameter vector that changes at the break dates (given in the column labeled "Correct Model"). The data generating process for the simulated data is given by equation (13). "Preferred Procedures" indicate model selection is conducted over both the number of structural breaks and the subset of the parameter vector that changes at each break date. "Baseline Procedures" indicate model selection is conducted over the number of structural breaks only, with all parameters allowed to change at each break date.

Table 4
U.S. GDP Deflator Inflation Rate (1953:Q1 – 2005:Q2)
Posterior Probability of Alternative Numbers of Structural Breaks in Autoregressive Model

Number of Breaks	Preferred Procedure	Baseline Procedure
No Breaks	0.0%	0.0%
One Break	0.0%	0.0%
Two Breaks	27.3%	4.8%
Three Breaks	49.7%	42.2%
Four Breaks	23.0%	53.0%

Notes: This table holds posterior model probabilities for autoregressive models of U.S. GDP Deflator inflation under alternative assumptions for the number of structural breaks. The "Preferred Procedure" computes the posterior probability for a given number of breaks without conditioning on a choice for the subset of the parameter vector that is allowed to change at each break date. That is, this choice is integrated out of the reported model probability. The "Baseline Procedure" computes the posterior probability for a given number of breaks conditional on allowing all parameters to break at each break date. The sample period considered is 1953:Q1-2005:Q2.

Table 5
U.S. GDP Deflator Inflation Rate (1953:Q1 – 2005:Q2)
Posterior Probability of Autoregressive Models with Only Breaks in Residual Variance vs. Autoregressive Models with Breaks in Conditional Mean Parameters

	Models with Only Breaks in Variance	**Models with Breaks in Conditional Mean Parameters**
Baseline Procedure	94.5%	5.5%
Preferred Procedure	2.2%	97.8%

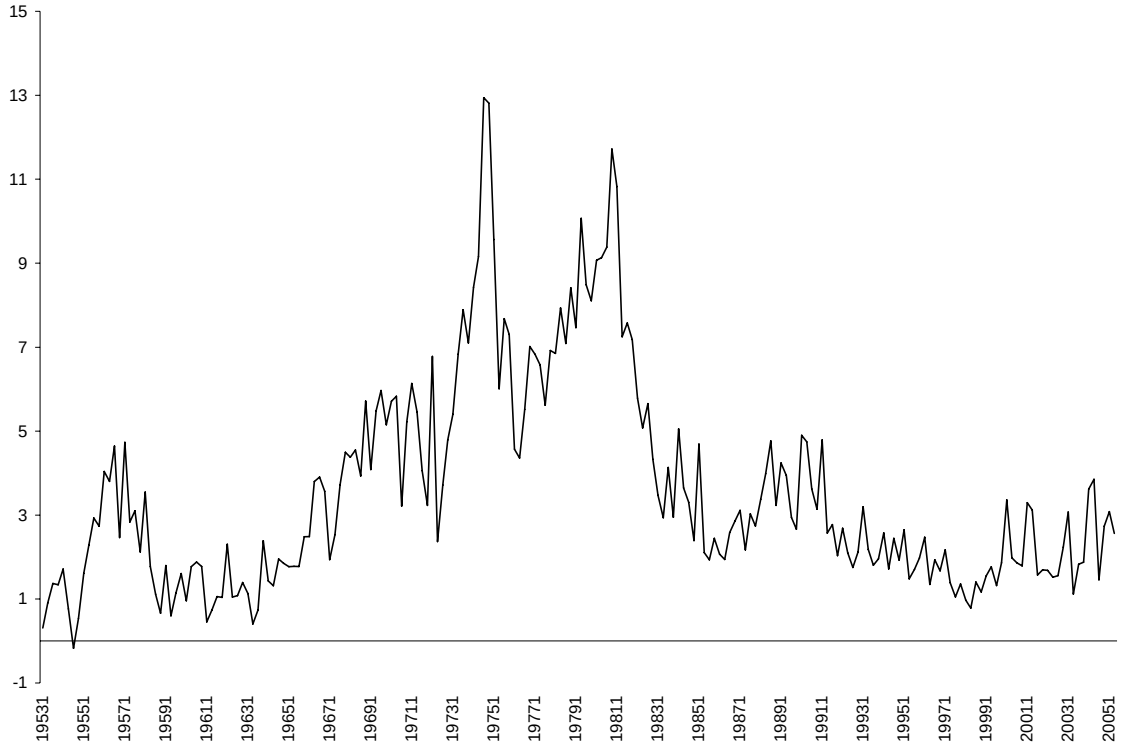
Notes: This table holds posterior model probabilities for autoregressive models of U.S. GDP Deflator inflation. The column labeled "Models with Only Breaks in Variance" contains posterior probabilities of the class of models that contain structural breaks in the residual variance parameter only. The column labeled "Models with Breaks in Conditional Mean Parameters" contains posterior probabilities of the class of models that contain structural breaks in conditional mean parameters, such as the intercept or autoregressive parameters. The probabilities in both columns are computed assuming a maximum of four structural breaks. The "Preferred Procedure" computes the posterior probability in the second column without conditioning on a choice for the subset of the conditional mean parameters that are allowed to change at each break date. The "Baseline Procedure" computes the posterior probability in the second column conditional on allowing all conditional mean parameters to break at each break date. The sample period considered is 1953:Q1-2005:Q2.

Table 6
U.S. GDP Deflator Inflation Rate (1953:Q1 – 2005:Q2)
Specification and Estimation Details for Highest Posterior Probability Autoregressive Models with Three Structural Breaks

Posterior Probability Relative to Most Preferred Model	First Break		Second Break		Third Break		Lag Order
1.00	c	65:Q4	c, σ	70:Q3	c, σ	81:Q3	4
0.90	c, σ	66:Q1	c, σ	71:Q3	c, σ	81:Q2	4
0.67	c	66:Q2	σ	70:Q3	c, σ	81:Q3	4
0.61	c, σ	68:Q2	ϕ, σ	82:Q4	c, σ	91:Q3	1
0.50	c, σ	67:Q2	σ	70:Q3	c, σ	81:Q3	4
0.41	c, σ	68:Q1	c, σ	81:Q3	σ	91:Q3	4
0.30	σ	60:Q3	c, σ	67:Q3	c, σ	81:Q3	4
0.27	c, σ	67:Q3	c	73:Q2	c, σ	81:Q2	4
0.27	c, σ	68:Q2	c, ϕ, σ	82:Q4	c, σ	91:Q3	1
0.25	c, σ	68:Q2	ϕ, σ	82:Q4	c	91:Q3	1
0.25	c	65:Q4	c, σ	70:Q3	c, σ	81:Q3	2
0.25	c, σ	66:Q3	c, σ	71:Q1	c, σ	81:Q4	2
0.20	c, σ	68:Q3	ϕ, σ	82:Q2	ϕ, σ	91:Q2	1
0.18	c	65:Q4	c, σ	70:Q4	c, σ	81:Q3	3
0.17	c	67:Q2	σ	70:Q3	c, σ	81:Q4	2
0.15	c, σ	66:Q2	c, σ	71:Q3	c, σ	81:Q3	3
0.15	c, σ	67:Q4	c, σ	81:Q3	c	91:Q2	4
0.14	c, σ	67:Q4	c, σ	81:Q4	ϕ, σ	91:Q2	1
0.14	c, σ	67:Q4	c, σ	81:Q3	σ	91:Q4	2
0.12	c, σ	67:Q3	σ	70:Q3	c, σ	81:Q4	2
0.12	c, σ	68:Q1	ϕ, σ	82:Q2	c, ϕ, σ	91:Q2	1
0.12	c, σ	68:Q2	c, ϕ, σ	82:Q4	c	91:Q3	1
0.11	c, σ	67:Q3	c	73:Q2	c, σ	81:Q3	2
0.11	c	67:Q2	σ	70:Q3	c, σ	81:Q2	3
0.10	c, σ	67:Q4	c, σ	81:Q3	c	91:Q2	2

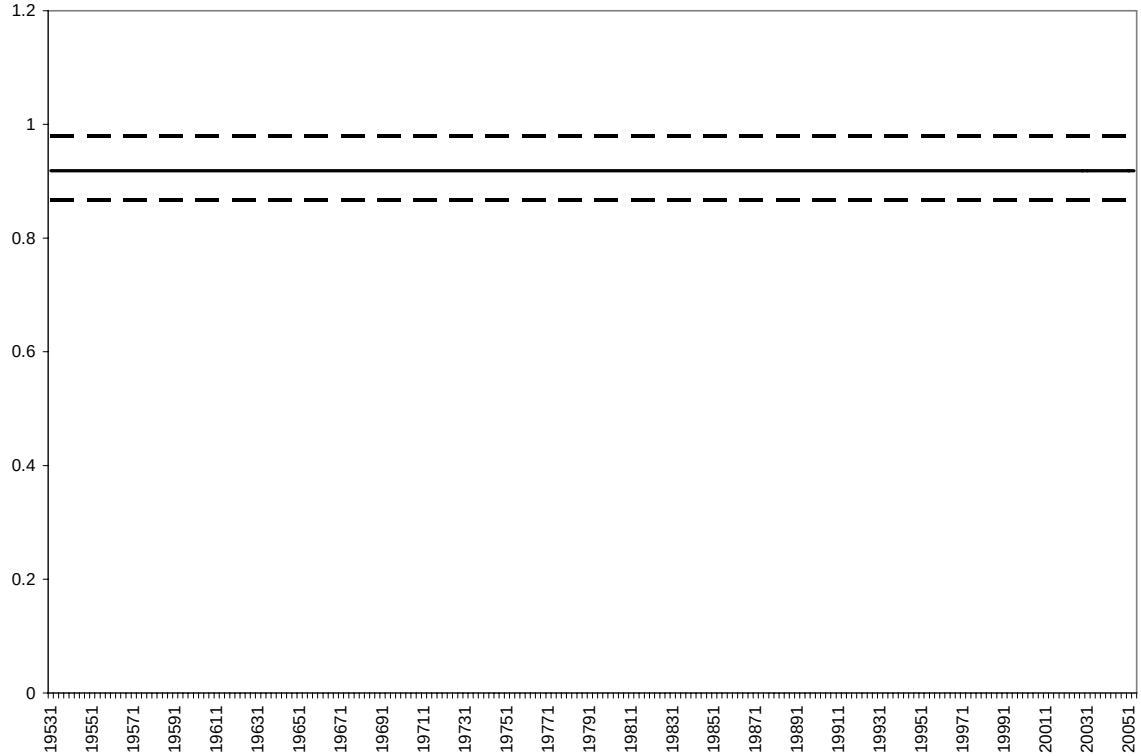
Notes: This table holds specification and estimation details for the highest posterior probability autoregressive models of U.S. GDP Deflator inflation with three structural breaks. For each of the three breaks, the table gives the parameters that are allowed to change at that break as well as the median of the posterior distribution of the break date. The sample period considered is 1953:Q1-2005:Q2.

Figure 1
U.S. GDP Deflator Inflation Rate (1953:Q1-2005:Q2)



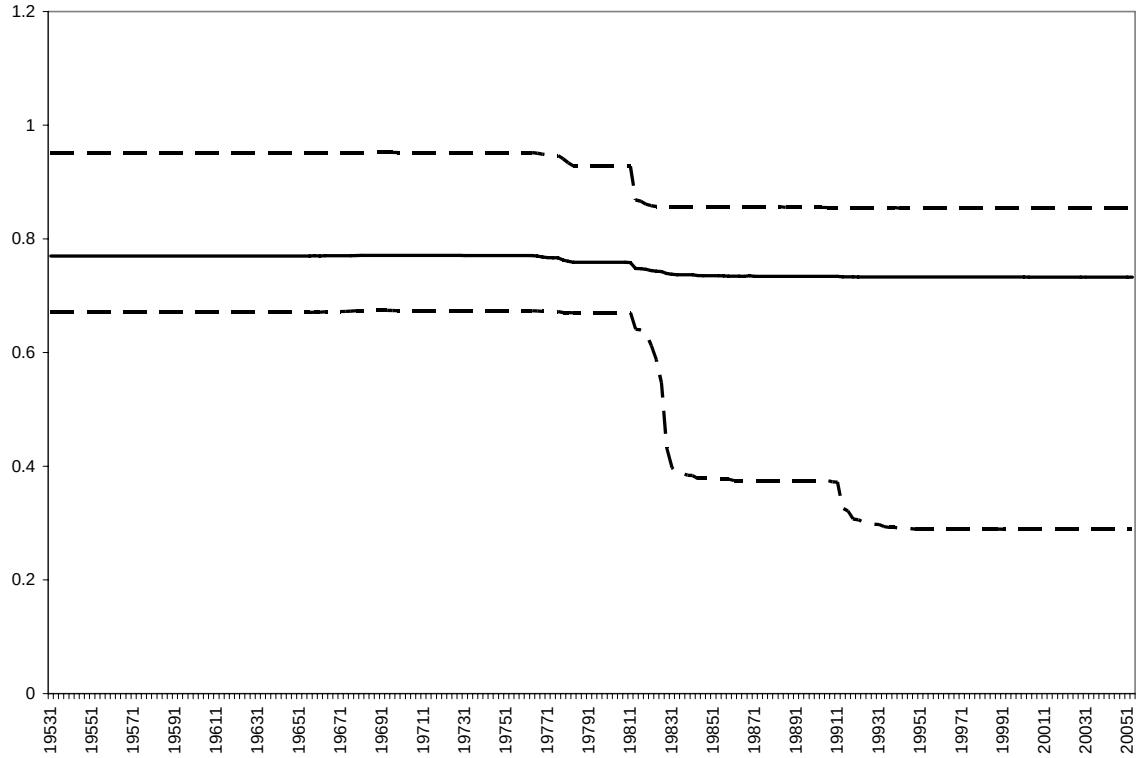
Notes: This figure plots quarterly, annualized growth rates of the U.S. GDP Deflator over the sample period 1953:Q1 – 2005:Q2.

Figure 2
U.S. GDP Deflator Inflation Rate (1953:Q1 – 2005:Q2)
5th, 50th, 95th Percentiles of Posterior Distribution for Sum of Autoregressive Coefficients from Model with No Break



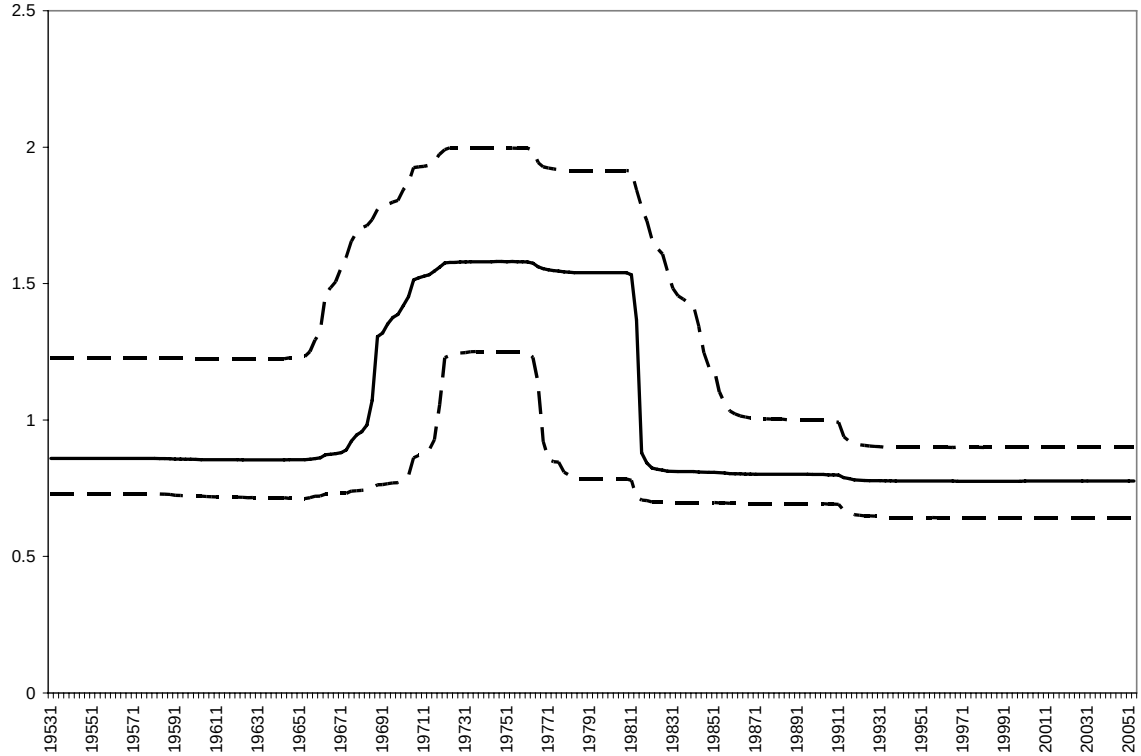
Notes: This figure plots percentiles of the posterior distribution of the sum of the autoregressive coefficients from a fixed parameter autoregression for U.S. GDP Deflator inflation over the sample period 1953:Q1 – 2005:Q2.

Figure 3
U.S. GDP Deflator Inflation Rate (1953:Q1 – 2005:Q2)
5th, 50th, 95th Percentiles of Posterior Distribution for Sum of Autoregressive Coefficients



Notes: This figure plots percentiles of the posterior distribution for the sum of the autoregressive coefficients from an autoregression for U.S. GDP Deflator inflation over the sample period 1953:Q1 – 2005:Q2. The posterior probability is not conditional on the choice of whether to include structural breaks, the number of structural breaks, or the subset of the parameter vector allowed to change at each break date.

Figure 4
U.S. GDP Deflator Inflation Rate (1953:Q1 – 2005:Q2)
5th, 50th, 95th Percentiles of Posterior Distribution for Residual Standard Deviation



Notes: This figure plots percentiles of the posterior distribution for the residual variance parameter from an autoregression for U.S. GDP Deflator inflation over the sample period 1953:Q1 – 2005:Q2. The posterior probability is not conditional on the choice of whether to include structural breaks, the number of structural breaks, or the subset of the parameter vector allowed to change at each break date.