Structural Breaks in U.S. Macroeconomic Time Series: A Bayesian Model Averaging Approach*

Adam Check[†] Jeremy Piger[‡]

This Draft: October 2018

Abstract: We investigate the evidence for structural breaks in the parameters of autoregressive models of U.S. post-war macroeconomic time series. There is substantial model uncertainty associated with such models, including uncertainty related to lag selection, the number of structural changes, and the specific parameters that change at each break date. We develop a feasible approach to Bayesian Model Averaging (BMA), where the model space encompasses each of these sources of uncertainty. This BMA procedure performs very well in Monte Carlo simulations calibrated to match relevant macroeconomic time series. We then apply the BMA approach to a cross-section of U.S. macroeconomic variables measuring inflation, production growth, and labor market conditions, finding substantial evidence for structural breaks in all of these series. For most series there are multiple structural breaks detected. We find pervasive evidence for at least one, and often multiple, breaks in conditional variance parameters, and for price inflation series we find strong evidence of changes in persistence. We find little evidence for changes in trend growth rates of production series, or in the natural rate of unemployment. For most series there is substantial uncertainty along one or more dimension of model specification, calling into question the common practice of basing inference on a single selected structural break model.

Keywords: posterior model probability, model selection, model uncertainty, changepoint, inflation persistence, Great Moderation

JEL Classifications: C13, C22, E32, G12

*We thank Sacha Gelfer and seminar participants at the Federal Reserve Bank of St. Louis Econometrics Workshop, the 2018 Western Economic Association International Conference, and the 2018 Midwest Economic Research Group for helpful comments.

[†]Department of Economics, University of St. Thomas, (ajc@stthomas.edu)

[‡]Department of Economics, University of Oregon, (jpiger@uoregon.edu)

1 Introduction

There are a host of economic and political developments that could be expected to alter the reduced-form dynamics of macroeconomic time series, and a large literature is devoted to evaluating the evidence for such changes. In this paper we are interested in a substantial subset of this literature that investigates the evidence for discrete structural breaks in the intercept, slope and residual variance parameters of autoregressive (AR) models fit to macroeconomic time series. Examples of the application of the AR model with structural changes include studies of U.S. trend productivity growth (Hansen (2001)), the dynamics of international unemployment rates (Papell et al. (2000); Summers (2004)), the dynamics of real interest rates and inflation (Perron (1990); Wang and Zivot (2000); Rapach and Wohar (2005); Clark (2006); Eo (2016)), and the volatility of macroeconomic and financial time series (Inclán (1993); Stock and Watson (2002); Sensier and van Dijk (2004)).

The potential for structural parameter changes in AR models 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 the number of AR lags to include 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 short lag lengths, the space of potential models to consider becomes enormous.

In the face of this model uncertainty, the existing literature has primarily made *a priori* choices that eliminate portions of the model space, and then used data-based model selection procedures along the remaining dimensions of model specification. For example, a substantial literature has devoted attention to the model selection question of the number of structural changes, while generally conditioning on other elements of model specification, such as the number of AR lags or the type of structural changes.[1] From the classical framework, Bai

[1]In some cases, model specification proceeds sequentially. For example, the number of AR lags might be

and Perron (1998) and Andrews and Ploberger (1999) develop sequential testing procedures designed to reveal the number of, perhaps multiple, structural changes. Using a Bayesian framework, Inclán (1993), Chib (1998) and Wang and Zivot (2000) use posterior odds ratios to compare AR models with different numbers of in-sample breaks, and Pesaran et al. (2006) additionally use hierarchical methods to allow for a random number of out-of-sample breaks when forecasting. Finally, a number of papers have developed flexible Bayesian models that do not fix the number of in-sample breaks for a particular model, but instead treat it as random in estimation. Examples include Koop and Potter (2007), Giordani and Kohn (2008), Geweke and Jiang (2011) and Maheu and Song (2014).

Far less attention has been given to dimensions of model uncertainty beyond the number of structural changes. 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), Levin and Piger (2008), Chen and Zivot (2010) and Eo (2016) use Bayesian posterior model probabilities to compare alternative AR models on dimensions beyond the number of structural changes. In each case, the authors restrict the model space such that it is feasible to consider all possible models. In particular, Hultblad and Karlsson (2008) do not consider uncertainty regarding the type of structural change, Chen and Zivot (2010) do not allow AR parameters to change, Eo (2016) chooses the AR lag length sequentially prior to considering structural breaks, and Levin and Piger (2008) and Eo (2016) do not allow some AR parameters to break with others remaining constant.

Meanwhile, the focus of nearly all of the existing literature has been on selection of a particular AR model with structural changes, 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);

chosen based on a model without structural breaks, and then the number of structural breaks established conditional on this chosen number of lags. This effectively eliminates a portion of the model space that would have been considered through a joint consideration of model uncertainty.

Moulton (1991); Draper (1995); Kass and Raftery (1995); Raftery (1996); Fernández et al. (2001a)). 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.[2] Unfortunately, the direct averaging of all possible models will not be feasible when the model space becomes too large, as is the case for all but the simplest AR models with structural parameter changes.

In this paper, we develop a feasible approach to conduct BMA for AR models with structural changes, where the model space encompasses lag selection, the number of structural changes, and the type of each structural change. Specifically, we design a MCMC sampler to obtain draws from the model space, which are in turn used to estimate posterior model probabilities. We investigate the performance of our proposed BMA procedure for characterizing structural breaks via a Monte Carlo simulation study. We simulate a variety of autoregressive data generating processes, with alternative types and numbers of structural breaks, that are meant to mimic key features of U.S. macroeconomic time series. Our BMA procedure performs very well in these simulations, identifying both the type and number of parameter changes accurately, as well as the lag order of the autoregressive process. The procedure is not overly prone to false positives, only infrequently identifying structural breaks when the true data generating process does not contain any.

Our objective is closely related to that in Groen et al. (2013), who propose a feasible BMA procedure to investigate variable inclusion and the number and type of structural breaks in Phillips Curve specifications for U.S. inflation. The primary difference from our approach is that Groen et al. (2013) impose hierarchical priors on the size of parameter breaks and

[2] For an introduction to BMA and a review of related literature, see Hoeting et al. (1999) and Steel (2017).

the break dates, whereas we leave these as unrestricted parameters to be estimated. While the additional structure imposed by the hierarchical priors can be useful for forecasting, which is the goal of Groen et al. (2013), we prefer to pursue an unrestricted estimation for retrospective analysis. Our design will necessitate an alternative scheme to sample the model space from that used in Groen et al. (2013), who are able to use the efficient sampling algorithm of Gerlach et al. (2000) to sample their specification of the structural break process. We will instead proceed by directly sampling the model space using an extension of the Markov Chain Monte Carlo Model Composition (MC^3) procedure of Madigan and York (1995).

We use the proposed feasible BMA procedure to investigate structural breaks in AR models for a cross-section of U.S. macroeconomic time series measuring wage and price inflation, production growth, and labor market conditions, measured from 1959 through the present. Our results produce several conclusions. First, there is overwhelming evidence of structural breaks for all U.S. macroeconomic series that we consider, and for most series there are multiple structural breaks detected. For all series, there appear to be at least one, and usually multiple structural changes in conditional variance. We find substantial evidence for breaks in the autoregressive parameters of price inflation series. Finally, the evidence suggests that short order AR models are preferred by the data, with only one to two lags included with meaningful probability.

For most series, there is substantial model uncertainty in at least one dimension, whether it be the exact number of structural breaks, the type of structural breaks, or the AR lag order. This provides empirical support for the importance of accounting for model uncertainty in autoregressive models with structural breaks, and calls into question the common practice of basing inference on a single selected structural break model. We use our BMA procedure to produce model-averaged inference regarding a number of macroeconomic phenomena, including changes in trend growth rates of production series, changes in the natural rate of unemployment, changes in the level and persistence of inflation, and the continuation of the macroeconomic "Great Moderation". We do not find any evidence for substantive changes in the trend growth rate of real GDP or in the natural rate of unemployment over the entire sample period. We do find strong evidence for changes over time in the persistence of core inflation series, with this persistence rising during the 1960s and falling around 1990. Finally, our results suggest that the Great Moderation is alive and well, notwithstanding a burst in volatility associated with the Great Recession.

The remainder of the paper is organized as follows. Section 2 defines the class of AR models with structural changes that will be considered, and establishes aspects of model uncertainty. Section 3 discusses BMA in the context of the structural change model, and lays out our proposed feasible BMA procedure. Section 4 evaluates the performance of this BMA procedure using a Monte Carlo simulation experiment. Finally, Section 5 investigates the evidence for structural changes in the parameters of AR models for U.S. macroeconomic times series. Section 6 concludes.

2 Model Specification

This section defines the autoregressive model with multiple structural breaks in intercept, autoregressive parameters and conditional variance, and discusses the various choices required to fully specify the model. To begin, consider the Gaussian autoregressive model with constant parameters:

$$\begin{gathered} y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t, \quad t = 1, \dots, T \\ \varepsilon_t \sim i.i.d.\ N\left(0, h^{-1}\right), \end{gathered} \tag{1}$$

where h is the precision parameter, defined as the inverse of the conditional variance. We assume throughout that we have an appropriate number of initial observations, $y_0, y_{-1}, \dots, y_{p-1}$, upon which to condition inference.

The autoregressive model with structural breaks augments (1) to allow for m episodes of parameter change that occur at the ordered dates $\tau = \{\tau_1, \tau_2, ..., \tau_m\}$, with τ_i a positive integer. We wish to write this model such that it allows for specification choice regarding which parameters change at each breakdate. Define an m element row vector of indicator variables based on these dates as $I_t = [I(t \geq \tau_1) \;\; I(t \geq \tau_2) \;\; \ldots \;\; I(t \geq \tau_m)]$. We then write the autoregressive model with breaks as follows:

$$
\begin{aligned}
y_t &= \alpha_t + \sum_{i=1}^{p} \phi_{i,t} y_{t-i} + \varepsilon_t, \quad t = 1, 2, \ldots, T \\
\varepsilon_t &\sim i.i.d.\ N\left(0, h_t^{-1}\right) \\
\alpha_t &= \alpha + I_t^{\alpha}\, \widetilde{\alpha} \\
h_t &= h + I_t^{h}\, \widetilde{h} \\
\phi_{i,t} &= \phi_i + I_t^{\phi_i}\, \widetilde{\phi}_i, \quad i = 1, \ldots, p.
\end{aligned} \tag{2}
$$

Here, I_t^{α} is a row vector holding the subset of the indicator variables in I_t that are relevant for measuring changes in the intercept parameter. If α_t changes at each break date, then $I_t^{\alpha} = I_t$, while if α_t experiences no structural breaks then I_t^{α} is the empty set. The vector $\widetilde{\alpha}$ holds the vector of changes in the intercept for each date where the intercept changes. $\{I_t^h, \widetilde{h}\}$ and $\{I_t^{\phi_i}, \widetilde{\phi}_i\}$, $i = 1, \ldots, p$, are defined similarly to capture structural breaks in the precision parameter and each autoregressive parameter respectively.[3]

As is common in the recent literature on structural change models, the break dates are assumed unknown and treated as additional objects to be estimated. We restrict the break dates so that the model parameters are constant for a minimum of b time periods between

[3]We have chosen to write the AR model with structural breaks in an intercept form rather than a deviation from mean form. This implies that a change in the autoregressive parameters, holding other parameters constant, will generate a change in the unconditional mean of the process.

breaks, as well as at the beginning and end of the sample:

$$\begin{aligned} &\tau_1 > b, \\ &(\tau_i - \tau_{i-1}) \geq b, \quad i = 2, ..., m, \\ &(T + 1 - \tau_m) \geq b \end{aligned} \tag{3}$$

When estimating this model, we treat b as a prior parameter set by the econometrician.

The autoregressive model with multiple structural changes of unknown type necessitates several dimensions of model specification. First, as with (1), the number of lagged variables in the model must be chosen, which we denote as a choice of the lag order, p. Second, the number of structural changes, m, must be set. Third, we must specify which of the intercept, autoregressive parameters, and precision parameter are allowed to change at each of the m break dates. This is determined by choices of I_t^{α}, I_t^{h}, and $I_t^{\phi_i}$, $i = 1, \ldots, p$. We compactly denote this choice as $R = \left\{ I_t^{\alpha},\ I_t^{h},\ I_t^{\phi_1},\ \ldots,\ I_t^{\phi_p} \right\}_{t=1}^{T}$.

A particular model specification, denoted M, is then determined by a choice of $\{p, m, R\}$, and the number of potential models is determined by the number of unique combinations of $\{p, m, R\}$ considered. Here we consider values of p that take on any integer value from 0 to a maximum of p^*, and values of m that take on any integer value from 0 to a maximum of m^*. We allow R to be any possible choice of $\left\{ I_t^{\alpha},\ I_t^{h},\ I_t^{\phi_1}, \ldots,\ I_t^{\phi_p} \right\}$ that is consistent with m structural breaks. In other words, we allow each of I_t^{α}, I_t^{h}, $I_t^{\phi_1}$, $\ldots$, $I_t^{\phi_p}$ to be any subset of I_t, provided that R is consistent with at least one parameter change at each of the m break dates. The number of potential specifications of the autoregressive model with multiple structural changes is then:

$$N = \sum_{m=0}^{m^*} \left(\sum_{p=0}^{p^*} \left(2^{p+2} - 1\right)^m \right)$$

Finally, it will be useful to define (2) in a matrix form. Let $Y = (y_1, y_2, \ldots, y_T)'$ be the $T \times 1$ vector of observations on y_t, and $e = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$ be the $T \times 1$ vector of disturbance terms. Let Z be a $T \times J$ matrix holding the relevant, non-constant, right-hand side variables for a particular structural break model. In other words, Z is a matrix such that (2) can be written in matrix form as:

$$Y = [\iota_T \;\; Z]\, \beta + e,$$

where β is a vector holding the conditional mean parameters:

$$\beta = \left[\alpha, \widetilde{\alpha}', \phi_1, \widetilde{\phi}_1{}', \ldots, \phi_p, \widetilde{\phi}_p{}'\right]'$$

3 A Feasible Approach to Bayesian Model Averaging

3.1 Bayesian Model Averaging for the Multiple Break Model

The large number of specification choices required for the autoregressive model with multiple structural breaks is likely to lead to substantial model uncertainty. The typical practice in the literature is to make some model selection choices *a priori*, and use data-based procedures to determine the remaining elements of model specification. For example, consider the question of which parameters are allowed to change at each structural break date. A common *a priori* choice is a model in which a subset of parameters are allowed to change at every break date, while other parameters are never allowed to change.[4] Given this choice, other, possibly data based, procedures may be used to select the number of structural breaks and the autoregressive lag order in order to arrive at a specific model. These data based procedures are often applied sequentially. For example, the autoregressive lag order might be chosen based on a model without structural breaks, and then the number of structural changes selected conditional on this chosen lag order. In the end, a particular model is selected upon which to base inference.

Focusing on a particular model, selected either through data based or *a priori* means, ignores information in models other than the selected model, and thus does not yield infer-

[4]This is referred to as a partial structural change model in the literature. See, e.g., Wang and Zivot (2000).

ences that fully incorporate model uncertainty. Instead of basing inference on a single model, we will instead average inference about objects of interest across alternative models, where averaging is with respect to the Bayesian posterior probability that each model is the true model. Label the j^{th} model specification as M_j, where $j = 1, \ldots, N$. The Bayesian posterior model probability is then:

$$\Pr(M_j|Y) = \frac{f(Y|M_j)\Pr(M_j)}{\sum_{n=1}^{N} f(Y|M_n)\Pr(M_n)}, \tag{4}$$

In (4), $\Pr(M_j)$ is the prior probability that M_j is the true model and $f(Y|M_j)$ is the marginal likelihood for model M_j:

$$f(Y|M_j) = \sum_{\tau \in \Pi} \left[\int_{\theta} f(Y|\theta, \tau, M_j)\, p(\theta, \tau|M_j)\, d\theta \right], \tag{5}$$

where $\theta = \{\beta, h, \widetilde{h}\}$ collects the intercept, autoregressive, and precision parameters for model M_j, $f(Y|\theta, \tau, M_j)$ is the likelihood function for M_j and $p(\theta, \tau|M_j)$ is the prior distribution over the parameters of M_j. Finally, Π indicates the set of possible locations for the m break dates.

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 structural changes:

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

where $\Pr(m|Y, M_n) = \Pr(m|M_n)$ is 1 if model M_n contains m structural breaks and is 0 otherwise. As another example, consider the intercept parameter at time t, α_t. The BMA posterior density for α_t is:

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

where $p(\alpha_t|Y, M_n)$ is the posterior density for the intercept conditional on model M_n being the true model.

While BMA for the autoregressive model with multiple structural changes is conceptually straightforward, the potentially enormous model space makes direct calculation of each $\Pr(M_j|Y)$ based on (4) practically infeasible for all but the simplest cases. For example, consider an empirically relevant case where the maximum lag order allowed is ($p^* = 6$) and the maximum number of structural breaks allowed is $m^* = 5$. In this case there are $N > 1$ trillion alternative models to consider. For our study of structural breaks in U.S. macroeconomic time series, the model space will be even larger.

When faced with an infeasibly large sample space, a popular 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^3) algorithm of Madigan and York (1995), which uses a posterior simulator based on the Metropolis-Hastings algorithm. MC^3 was implemented by Raftery et al. (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 et al. (2001b)). One could apply an MC^3 algorithm in the current context to sample directly from the N-element model space. However, in order to calculate the Metropolis-Hastings acceptance probability, this algorithm would require computation of the marginal likelihood in (5), which is very computationally expensive.[5]

To sidestep this computational difficulty, we instead design a posterior sampler to obtain

[5] The number of summations required to cover all possible values of τ in (5) can be enormous. For example, for a small number of breaks ($m = 2$), a sample size similar to our empirical applications ($T = 230$), and assuming $b = 10$, there are over 20,000 possible locations for the break dates.

draws from a joint posterior distribution that includes the model space and the timing of structural changes. In other words, we obtain samples from the distribution $\Pr(M, \tau|Y)$. By extending the target posterior distribution to include the timing of structural change, the relevant marginal likelihood is no longer (5), but instead is for an autoregression with structural breaks of known timing, which can be computed relatively quickly. To construct these posterior draws we design a two-block Metropolis-Hastings sampler, the details of which are presented in the Appendix.

The posterior sampler produces draws from $\Pr(M, \tau|Y)$, which we denote as $M^{\{g\}}$ and $\tau^{\{g\}}$ for the g^{th} draw. Given G such draws, one can then form an estimate of $\Pr(M_j|Y)$ by simply counting the relative frequency that M_j is visited by the sampler. Specifically, an estimate of $\Pr(M_j|Y)$ is given by:

$$\frac{1}{G}\sum_{g=1}^{G} I\left(M^{\{g\}} = M_j\right), \tag{6}$$

where $I\left(M^{\{g\}} = M_j\right)$ is an indicator function that equals one if $M^{\{g\}} = M_j$ and is zero otherwise. Note that this approach gives zero posterior probability to any model that is not visited by the sampler. This fact suggests an alternative estimate of $\Pr(M_j|Y)$ based on equation 4, where the summation over models is restricted to only those models that are drawn by the sampler:

$$\Pr(M_j|Y) = \frac{f(Y|M_j)\Pr(M_j)}{\sum_{n\in S} f(Y|M_n)\Pr(M_n)}, \quad j \in S \tag{7}$$

If the models never visited by the sampler are assumed to have zero probability, model probabilities based on 7 will be exact, while those based on 6 will contain estimation error. Gvien this, all results we present are based on model probabilities constructed from 7.

3.2 Prior Specification

The BMA approach requires the specification of a prior distribution for the parameters of each model as well as a prior distribution across alternative models. As there are a large number of models, we do not attempt to elicit priors for each individual model, but instead use a strategy that generates priors automatically. This is a common practical approach to set priors in the BMA literature (e.g. Fernández et al. (2001b) and Blonigen and Piger (2014)). We begin with the joint prior density for the parameters of each model, $p(\theta, \tau|M)$, which we factor as follows:

$$p(\theta, \tau|M) = \Pr(\tau|\theta, M)p(\theta|M)$$

For $\Pr(\tau|\theta, M)$, we place equal prior probability on all possible locations of the m break dates. This follows the literature investigating Bayesian multiple break models (e.g. Inclán, 1993; Stephens, 1994; Wang and Zivot, 2000). Recall that Π indicates the set of possible locations for the m break dates, and label the total number of such locations as C. The probability distribution function is then uniform:

$$\Pr(\tau|\theta, M) = \frac{1}{C}, \quad \tau \in \Pi, \tag{8}$$

where:

$$C = \frac{((T - 2b + 1) - (m - 1)(b - 1))!}{m!\,((T - 2b + 1) - (m - 1)(b - 1) - m)!}$$

Note that C depends on the minimum regime length parameter b, which we treat as a prior hyperparameter set by the econometrician. For the quarterly data used in our application we set $b = 8$, so that one structural break can happen every two years.

The remaining prior distribution can be further factored:

$$p(\theta|M) = p(\beta, h|\widetilde{h}, M)p(\widetilde{h}|M)$$

For $p(\widetilde{h}|M)$ we use independent and identical Gamma$(\widetilde{a}, \widetilde{v})$ distributions for each element of $\widetilde{h}$. For the remaining parameters we use a Normal-Gamma prior that is independent of $\widetilde{h}$:

$$[\beta, h]' \,|\widetilde{h}, M \sim NG\left(\mu, h^{-1}V, a, v\right),$$

This is the natural conjugate prior for the case where there are no structural changes in the precision parameter, and its use will simply marginal likelihood calculations considerably for this special case.

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 we follow a strategy 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.

The specifics are as follows. First, we set the prior mean for β equal to a vector of zeros ($\mu = 0_J$). The prior variance for conditional mean parameters is set as follows:

$$V = \begin{bmatrix} c & 0'_{J-1} \\ 0_{J-1} & V^* \end{bmatrix}.$$

The scalar c captures the prior variance on the intercept parameter, α, which is assumed to be in all models considered. As such, we set $c = 10^8$, which approximates the standard non-informative prior for α. The matrix V^* captures the prior variance-covariance matrix of the remaining conditional mean parameters, which are not in each possible model. To set V^* we use:

$$V^* = dI,$$

which simplifies the specification of V^* to the choice of a single value, d. For our primary results reported below we set $d = 16$. We have also evaluated the sensitivity of our results to values of d that are one-half and twice our baseline value, and found nearly identical results.

The first-regime precision parameter, h, is in all models considered, and as such we specify a prior distribution that approximates the standard non-informative prior for the model precision, setting $a = 1$ and $v = 10^{-10}$. This yields a Gamma density with mean of 1 and variance 2 x 10^{10}. For the remaining precision parameters, we set $\widetilde{a} = 1$ and $\widetilde{v} = 0.1$, which yields a Gamma density with mean of 1 and variance of 20.

Finally, we describe our strategy to provide automatic prior distributions for alternative models, $\Pr(M) = \Pr(p, m, R)$. Define the "model size" as the number of intercept, autoregressive, and precision parameters in the model. In the notation used above this is the number of elements in the set of parameters θ. We set $\Pr(M)$ indirectly by setting a flat prior over model size. That is, the set of all feasible choices of p, m, R that produce models with q conditional mean, autoregressive and precision parameters would have the same total probability as the set of all feasible choices of p, m, R that produce models with r such parameters.[6] Then, within a set of feasible models that have the same model size, we place equal prior probability on each model. Formally, let q equal the model size implied by a particular feasible choice of p, m, R, where $q = 2, 3, \ldots, Q$, and let F_q equal the number of unique feasible choices of p, m, R that yield model size equal to q.[7] Then:

$$\Pr(M) = \Pr(p, m, R) = \frac{1}{F_q (Q-1)} \tag{9}$$

[6] "Feasible" in this context refers to internal consistency of the elements that make up the model. For example, an AR(0) model with a structural breaks in an AR(1) parameter is an internally inconsistent model, and would receive zero weight in our prior over models.

[7] As all models contain α and h, the minimum value of $q = 2$.

Both F_q and Q depend on the maximum lag length allowed, $p^\star$, and the maximum number of structural breaks allowed, $m^\star$. We treat each of these values as a prior hyperparameter set by the econometrician. For the quarterly data used in our study we set $p^\star = 6$ and $m^\star = 10$. For all the data series we study, the data prefers values of p and m that are well below these maximum values.

4 Monte Carlo Simulation

To assess the performance of the BMA procedure described above, we perform a Monte Carlo exercise where we apply the procedure to simulated datasets that mimic relevant features of U.S. post-war macroeconomic time series. We consider five data generating processes (DGPs), detailed in Table 1.

The first DGP, labeled "No Break", is an AR(1) process with no structural breaks in any parameters. This DGP allows us to evaluate the frequency with which our our procedure falsely detects structural change. In the "No Break" DGP we set the AR(1) parameter to 0.9, thus generating a highly persistent process. This is motivated by an existing literature, e.g. Diebold and Chen (1996) and Prodan (2008), showing that frequentist tests for structural breaks can have severe size distortions when faced with persistent autoregressive processes.

The second and third DGPs, labeled "Intercept Large" and "Intercept Small" respectively, are each AR(1) processes with two breaks occurring in the intercept of the process only. In the "Intercept Large" case, these breaks are of a size 1.4 times the standard deviation of the model disturbance term, while in the "Intercept Small" case these breaks are of a size 0.7 times the standard deviation of the model disturbance terms. These types of processes are often used to investigate changes in the unconditional mean of series such as the growth rate of U.S. Real Gross Domestic Product or the U.S. unemployment rate.

The fourth DGP, labeled "Persistence" is an AR(2) process with two shifts in the persistence of the process, such that both the unconditional mean and variance of the process starts low, moves higher, and then returns to a lower level. Such a process has been suggested for post-war U.S. inflation series, and we calibrate the process to generate series that roughly match the variation in the level and volatility of these series over the post-war period. The final DGP, labeled "Intercept / Variance", is an AR(2) process with two shifts in both the intercept and conditional variance, calibrated such that the unconditional mean and variance of the process starts low, moves higher, and then returns to a lower level. Such a process generates a similar pattern as the "Persistence" DGP, and allows us to evaluate the ability of the sampler to detect the true DGP from models with competing features.

For each DGP the sample size is set equal to that used in our quarterly study of U.S. macroeconomic time series, $T = 226$, and the break dates are placed roughly equidistant throughout the sample. All Monte Carlo simulation results are based on applications of the BMA procedure to 100 simulated time series from each DGP. In implementing the Metropolis-Hastings sampler, we use 125,000 samples as our pre-convergence period, and base results on a subsequent 250,000 samples. All prior hyperparameters were set as discussed in Section 3.2.

In summarizing the Monte Carlo evidence, we analyze three categories of performance for the BMA procedure: (1) accuracy of inclusion for each autoregressive lag, (2) identification of the number and type of breaks, and (3) estimation of parameter values. We find that our procedure performs well on each criterion under all five DGPs.

Regardless of the true DGP, the Bayesian procedure detected the presence of the appropriate number of autoregressive lags very accurately. Table 2 shows the Bayesian posterior inclusion probability (PIP), defined as the posterior probability that a particular autoregressive lag belongs in the true model, for all autoregressive lags up to $p^{\star} = 6$. In all cases, the BMA procedure places greater than 80% posterior probability on the true number of autoregressive lags, and in the majority of cases this posterior probability is above 90%. For autoregressive lags that were not included in the true DGP, the estimated inclusion probabilities were typically very close to 0%.

We next turn to accuracy of break detection. Table 3 shows the posterior probability of alternative numbers of structural breaks. In all cases, the BMA procedure places substantial posterior probability on the correct number of structural breaks. For four of the five DGPs, the posterior probability for the correct number of structural breaks is near to or above 80%, and in three cases is close to 90%. For the "Intercept Small" case the posterior probability for the correct number of breaks (2) is 44%, while the posterior probability for a single break is 52%. Given the relatively small size of the structural break for this DGP, the BMA procedure's ability to detect at least one break in the process at a high rate is encouraging. This case also demonstrates the benefits of a BMA approach, which would give substantial weight to both one and two break models, vs. selecting only one highest probability model, which would focus only on the incorrect one break model. Finally, the BMA procedure does not frequently identify structural breaks that are not present. For example, for the "No Break" DGP, the procedure identifies false breaks only 9% of the time.

We are interested not only in accurate identification of the number of breaks, but also in accurate identification of the types of structural breaks. Table 4 shows the mean number of structural breaks included for each potential model parameter in each DGP. The BMA procedure is extremely accurate at detecting which parameters change at each break date. In all cases, the parameters that actually break in each DGP are the ones that are identified as changing, while very few breaks are identified for parameters that do not actually undergo structural changes.

Finally, we evaluate how well the Bayesian procedure performed at estimating the true parameter values at each point in time. In particular, for each Monte Carlo simulation, we compute the RMSE for each parameter, defined as follows:

$$RMSE = \sqrt{\sum_{t=1}^{T} \frac{1}{T} \left(E\left(\gamma_t | Y\right) - \gamma_t \right)^2}$$

where γ_t is the true value of a specific parameter at time t and $E\left(\gamma_t | Y\right)$ is the Bayesian posterior mean of this parameter at time t. We then average the RMSE values across the 100 Monte Carlo simulations, and present the mean RMSE value in Table 5. We can see that the parameter values were estimated very accurately by our sampler, with the largest mean RMSE being 0.25, and most far smaller. The mean RMSE for variables that were not included in the full model were particularly small, since our sampler was able to identify that these variables did not belong in the model with a high degree of accuracy, and were therefore able to perfectly set the coefficient value equal to zero in most simulations.

5 Empirical Results

Having demonstrated the efficacy of our feasible BMA procedure in Monte Carlo simulations, we now present evidence from the application of this procedure to study structural breaks in autoregressive models of U.S. macroeconomic time series. Our sample consists of a cross section of eight time series, measuring production, employment, wages, and prices.[8] We consider growth rates of the underlying variables for all series except the unemployment rate, which is each considered in levels. The data frequency is quarterly, and the sample period extends from a maximum of the first quarter of 1959 to the second quarter of 2018. For some series the sample period is smaller due to data availability. Table 6 provides additional details regarding the series considered.

The feasible BMA procedure proceeds by sampling from the model space using a Metropolis-Hastings sampler. Our results are based on one million draws from the sampler, following 250,000 pre-convergence draws. We took a number of checks on the performance of the sampler and the sufficiency of the number of pre-convergence draws. First, the acceptance rates from each block of the Metropolis-Hastings sampler were within reasonable ranges, averaging between 10% and 30% for the various series under consideration. Second, the one million post-convergence draws were generated by four separate chains, each of which was allowed

[8] We focus on core price index series to avoid issues with detecting breaks in the presence of large outliers in food and energy prices.

to produce 250,000 pre-convergence draws and 250,000 post-convergence draws. The results produced by each of these chains were very similar to each other, suggesting the chains had converged to their stationary distribution. Finally, we produced a number of standard visual diagnostic checks, such as trace plots and running mean plots. Each of these suggested that our convergence period was sufficient.

5.1 BMA Evidence on Autoregressive Lag Order

We begin by considering the evidence for autoregressive lag inclusion. For each series, Table 7 gives the PIP for each autoregressive lag up to $p^* = 6$. These PIPs are not conditioned on a particular number or type of structural breaks, but instead incorporate uncertainty about these elements. The PIPs make clear that autoregressive dynamics are a clear feature for nearly all of these series, with the only exception being the growth rate of real investment. For all other series, the PIP for the AR(1) parameter is 100%. Inclusion of higher order AR dynamics is less uniform. Five of the series have PIPs for AR(2) parameter dynamics that are 100%, while one series, real GDP growth, has a PIP for the AR(2) parameter of 65%. Finally, one series, core CPI inflation, has a PIP for the AR(3) parameter that is above 50%. PIPs are uniformly low for AR lags above 3. Thus, in general, the results suggest relatively small AR models for most series considered, with short order (primarily AR(1) or AR(2)) dynamics prevalent for most series. It is also worth highlighting that while many PIPs are near zero or one, this is not uniformly the case. For example, there appears to be non-trivial model uncertainty regarding whether real GDP growth displays AR(2) dynamics. BMA is particularly well suited to incorporate this model uncertainty in estimation of other model elements.

5.2 BMA Evidence on the Number, Type and Timing of Breaks

We now turn to the evidence for structural breaks in these series. For each series, Table 8 presents BMA posterior model probabilities for models with different numbers of structural breaks, including models with no breaks. Again, it is important to remember that these posterior probabilities incorporate uncertainty about lag selection and the type of structural break. Looking down the column for zero breaks, it is clear that there is overwhelming evidence for structural breaks in these time series. For all but one series, the probability of a model with no structural breaks is identically zero, meaning that models with no structural breaks were never drawn following convergence of our sampler. For the remaining series, which is the unemployment rate, the posterior probability of the model without structural breaks is only 9%.

Given that models with structural breaks are strongly preferred, we next consider the evidence for the number of structural breaks. From Table 8, we can draw a number of conclusions. First, for all series, there is non-trivial weight placed on models with different numbers of structural breaks. In other words, there is substantial posterior uncertainty about the true number of structural breaks. This provides empirical motivation for the use of BMA methods, and casts doubt on procedures that draw inference by focusing on models with a specific number of structural breaks. More generally, it calls into question the enterprise of attempting to select a specific number of structural breaks for U.S. macroeconomic time series. Second, for all series, there is significant posterior probability placed on models with more than one structural break. Indeed, for six of the eight series considered, the highest probability models contain at least 2 breaks, and for five series the highest probability model contains at least 3 breaks. Third, the two series that place high weight on models with only a single structural break are each NIPA production components, namely real GDP and Consumption.

What is the nature of the structural breaks found for U.S. macroeconomic time series? Tables 9 - 11 give posterior probabilities for the number of breaks in the intercept parameter, AR parameters, and conditional variance parameter respectively. Beginning with Table 9, for most series, the posterior probability of zero breaks in intercept is close to or greater than 90%. The two exceptions are the growth rate of hourly earnings and core CPI inflation, which place 20% and 66% probability on the model with no structural breaks in intercept. It is worth highlighting the case of CPI inflation, for which the non-trivial possibility of a structural break in intercept would have been ignored in a study that chose a single "best" model. In our study this model will be averaged into inference.[9] Finally, for each series, Figure 1 shows plots of the posterior probability that a structural break in intercept occurs at each quarter in the sample period. This shows that the timing of intercept breaks for hourly earnings growth are dated to the early 1970s, mid 1970s, and early 1980s, while the timing of the intercept break for core CPI inflation is dated to the early 1990s.

Turning to Table 10, the evidence for breaks in AR parameters splits out along two groups of variables. For the production and employment variables, there is little evidence for breaks in AR parameters, as the posterior probability of the model with zero structural breaks is greater than 95% in all cases. For the inflation series, substantial posterior weight is placed on models with structural breaks in AR parameters, although there are some differences in the number of breaks. For core CPI inflation, substantial weight is placed on models with both one and two structural breaks in AR parameters, while for core PCE inflation nearly all posterior weight is placed on models with two and three breaks. For wage inflation, the highest probability model is one with no breaks in AR parameters, although substantial (43%) probability is placed on a model with 1 break. Overall, there seems to be strong evidence for structural breaks in the AR parameters of U.S. price inflation series, and less, though still substantial, evidence for such breaks in wage inflation. In terms of timing of breaks, Figure 2 shows that for the price inflation series, breaks in AR parameters are dated to two primary time periods, the mid-1960s and the early to mid 1990s. For wage inflation the breaks are dated in the 1970s and early 1980s.

Finally, we turn to Table 11, which provides posterior probabilities for alternative numbers of breaks in the conditional variance parameter. Here we see very strong and uniform

[9]It is worth noting that intercept breaks (or the lack of such breaks) does not necessarily imply a change in the unconditional mean of the series, as this will also depend on breaks in AR parameters, the evidence for which we turn to next. Later, in Section 5.3 we investigate the implications for changes in both the intercept and AR parameters for changes in the unconditional mean of these series.

evidence for breaks, with the probability of a model with no breaks in conditional variance receiving zero posterior weight for all series except the unemployment rate, for which the no break model receives 9% posterior probability. The probability of alternative numbers of structural breaks in conditional variance tends to be spread out among different, non-zero, numbers of breaks, with the pattern being series specific. It is easiest to summarize this information by considering the timing of structural breaks, which is shown in Figure 3, and demonstrates that breaks appear to cluster around three main time period. First, there are a large number of series for which a break is dated to the late 1960s or early 1970s. These elevated probabilities are seen in many of the nominal series, as well as the labor market variables. Second, most series have elevated break probabilities in the mid 1980s, consistent with the timing of the so-called "Great Moderation" in U.S. macroeconomic variables. For a small number of variables there is no elevation in probabilities in the early 1980s, but instead in the late 1980s or early 1990s. Finally, for a few series, there is an elevated probability of a structural break corresponding to the beginning and the end of the Great Recession.

5.3 BMA Evidence on Breaks in Trend Growth Rates, Inflation Dynamics, and Macroeconomic Volatility

The above results establish evidence regarding the number, timing and nature of structural breaks in the parameters of autoregressive models of U.S. macroeconomic time series. We now use these results to revisit a variety of specific questions from the existing literature. First, we investigate the evidence for structural breaks in trend growth rates of production. Second, we investigate the evidence for changes in the natural rate of unemployment. Third, we evaluate the possibility for level and persistence shifts in inflation series. Finally, we present new evidence regarding the nature of the Great Moderation in U.S. economic activity.

A long and prolific literature has investigated structural breaks in trend U.S. real GDP growth rates over the post-war period, with a primary focus being evidence for a reduction in trend real GDP growth during the early 1970s.[10] Recently, interest in this topic has experienced a resurgence, with a number of papers investigating the possibility of additional structural changes in trend real GDP growth in the mid 2000s (prior to the start of the Great Recession). Recent contributions in this literature include Luo and Startz (2014), Antolin-Diaz et al. (2017), Grant and Chan (2017), Kamber et al. (2018) and Eo and Morley (2018).

To investigate changes in trend growth rates, we look for breaks in the unconditional mean of both real GDP and real Consumption growth. Recall, the evidence in Tables 9 and 10 provided only weak evidence for breaks in the intercept or AR parameters for real GDP and Consumption growth. Given this, there is similarly weak evidence for breaks in the unconditional mean of real GDP growth, as this mean is a function of the intercept and AR parameters only. To provide some visual evidence, Figures 4-5 shows the posterior median and 68% coverage intervals for the unconditional mean of annualized real GDP growth and annualized real Consumption growth respectively. It is important to emphasize that this is a BMA estimate, meaning that the estimates incorporate uncertainty about lag selection, the presence of structural breaks, the location of breaks, and which parameters break. For both real GDP and real Consumption growth, the posterior median does show some downward drift over time, with small reductions visible in the mid 1980s and early 2000s, and a larger reduction visible in the mid 2000s. The sum of these reductions are relatively small, on the order of 0.07 of an annualized percentage point for real GDP. Also, these changes are small relative to the uncertainty with which the unconditional mean is estimated. Thus, our BMA procedure does not provide strong evidence in favor of changes in trend growth rates in the U.S. economy. It is true that our BMA posterior median estimates are qualitatively consistent with a reduction in trend growth rates over time in the U.S. economy. However, as of now, additional data is needed to strengthen this evidence.

Similarly, our BMA results provide very little evidence for changes in the natural rate of

[10] As just two examples in this large literature, see Perron (1989) and Bai et al. (1998).

the unemployment rate, as measured by its unconditional mean. Figure 6 shows the BMA posterior median and 68% coverage intervals, and displays very little visible change over time. Consistent with Staiger et al. (1997), the 68% coverage intervals are relatively large, suggesting the natural rate is a difficult concept to measure empirically.

We now turn to evidence on changes in the dynamics of post-war U.S. inflation rates, which is the focus of a substantial existing literature. In particular, Cogley and Sargent (2001) argue that the persistence of shocks to the U.S. inflation rate have varied considerably over the post-war 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 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. 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.[11] 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.

Figure 7 presents the BMA posterior median of the persistence of wage and price inflation series. We measure persistence using the sum of the autoregressive coefficients (SAR), a commonly used statistic that is recommended by Andrews and Chen (1994) as the best scalar measure of persistence. Our BMA results yield a number of conclusions. First, both core CPI and core PCE inflation display large changes in the persistence of inflation, with persistence increasing in the mid 1960s and declining around 1990. The size of these changes are largest for core PCE inflation, for which the SAR is very low prior to the first break (around 0.2), rises to a high level (around 0.8), and then declines dramatically after the

[11] See, for example, Kang et al. (2009), Eo (2016), Stock and Watson (2007), Benati (2008), and Giordani and Kohn (2008).

second break (around 0.4). Core CPI inflation persistence displays a very similar pattern with similar timing, but the changes are less pronounced. For both series, the 1990 timing for the reduction in inflation persistence is substantially later than the commonly held perception that inflation persistence fell immediately after the Volcker disinflation of the early 1980s.

Figure 8 shows how the various breaks in the inflation process have translated into changes in the unconditional mean of inflation. For both core CPI and PCE inflation, there is a substantial rise and then fall in unconditional mean, with the magnitude of the increase and subsequent decrease of a similar size. For core PCE inflation, nearly all of the changes in unconditional mean are driven by changes in the SAR, as there is only slight evidence of breaks in intercept for this series. For core CPI inflation, there was some (33%) posterior probability placed on a single break in the intercept parameter, corresponding to a decline in intercept around 1990, and some (22%) posterior probability placed on a second break in the AR parameters, corresponding to a decline in the SAR around 1990. These breaks are responsible for the size of the decline in the BMA estimate of the unconditional mean of core CPI inflation around 1990, but would have been ignored by a typical analysis based on only a single highest probability model. Finally, Figures 7 and 8 also show evidence for breaks in persistence and unconditional mean of wage inflation. For this series, changes in persistence are small, and persistence is relatively high over the whole sample. However, we see increases in the unconditional mean of wage inflation during the 1970s, followed by a large drop in the unconditional mean of the series in the early 1980s. These changes are driven by breaks in the intercept of the process.

Finally, we use our procedure to re-evaluate the evidence for the Great Moderation in U.S. macroeconomic time series. As was first documented in McConnell and Perez-Quiros (2000) and Kim and Nelson (1999), there is substantial evidence for a reduction in the volatility of measures of production growth occurring in the early 1980s. This reduction in volatility is also observed in measures of inflation, as well as in labor market variables. More recently, some authors have questioned whether the Great Recession signaled the end of the Great

Moderation.[12] Figures 9 - 11 shows the BMA posterior median for the unconditional variance of each series in our sample. Beginning with Figure 9, we see the very standard result that a large and sharp reduction in volatility occurred in the mid 1980s in real GDP growth. This reduction is also seen in real Investment growth and real Consumption growth, although the decline in Consumption growth volatility occurs in the early 1990s. For real GDP and Consumption growth, there is no increase in volatility associated with the Great Recession. For real Investment growth, there is a short-lived increase in volatility associated with the start of the Great Recession, but this is immediately reversed at the end of the recession, with volatility returning to an even lower level than the pre-Great Recession period. Thus, the original nature and timing of the Great Moderation appears to be confirmed by our BMA results, and our results suggest that the Great Moderation did not end with the Great Recession.

As shown in Figure 10, and consistent with the existing literature, we observe an increase in volatility of price inflation variables in the early 1970s, followed by declines in volatility during the 1980s and early 1990s. Interestingly, wage inflation volatility experiences an increase and decline that is dated earlier, and largely completed by 1970. Finally, as shown in Figure 11, the volatility of labor market variables follows a similar pattern to the price inflation variables, with an increase in volatility during the 1970s, a decline in volatility in the mid 1980s, and, for employment growth, a brief burst in volatility during the Great Recession.

6 Conclusion

We have investigated the evidence for structural changes in the parameters of autoregressive models for U.S. macroeconomic time series, including series measuring production growth, inflation, and labor market conditions. To account for the substantial *a priori* uncertainty associated with the correct specification for such models, we have developed a feasible

[12]See, for example Charles et al. (2018).

approach to conduct Bayesian Model Averaging for autoregressive models with structural changes, where the model space encompasses lag selection, the number of structural changes, and the type of each structural change. Using an extensive Monte Carlo experiment, we find that the proposed procedure performs very well for identifying the correct autoregressive lag structure, and the correct number and types of structural changes.

Our results suggest overwhelming evidence of structural breaks for all U.S. macroeconomic series that we consider over the post-war sample period. For most series, there are multiple structural breaks detected. We find ubiquitous evidence for at least one, and often multiple, breaks in conditional variance parameters, and for price inflation series we find evidence of substantial changes in autoregressive slope parameters. For most series, the exceptions being wage inflation and core CPI inflation, we find only slight evidence for breaks in model intercept parameters.

We use our procedure to provide Bayesian Model Averaged evidence on a variety of macroeconomic phenomena, including changes in the trend growth rate of real GDP, changes in the natural rate of unemployment, changes in the level and persistence of inflation, and the continuation of the so-called macroeconomic "Great Moderation". We do not find any evidence for substantive changes in the trend growth rate of real GDP or in the natural rate of unemployment over the entire sample period. We do find strong evidence for changes over time in the persistence of core inflation series, with this persistence rising during the 1960s and falling around 1990. Finally, our results suggest that the Great Moderation is alive and well, notwithstanding a burst in volatility associated with the Great Recession.

Appendix: Posterior Sampler for Model Space

In this appendix we describe the details of our two-block Metropolis-Hastings Sampler to obtain draws from $\Pr(M, \tau|Y)$, or equivalently $\Pr(p, m, R, \tau|Y)$. Define a $(p^\star + 2)$ column vector ψ_t holding each of the potential time-varying parameters at time t:

$$\psi_t = (\alpha_t, \phi_{1,t}, \phi_{2,t}, \ldots, \phi_{p^\star,t}, h_t)'$$

where $\phi_{j,t} = 0$, $\forall t$, if y_{t-j} is not included in the model. Next, define a $(p^\star + 2)$ column vector A_t whose i^{th} element is 1 if the i^{th} parameter in ψ_t undergoes a break at time t and is zero otherwise. Finally, define a $(p^\star + 2) \times T$ matrix A, which holds the A_t vectors for $t = 1, \ldots, T$:

$$A = [A_1, \quad A_2, \quad \ldots, \quad A_T]$$

A choice of the matrix A, subject to the minimum distance constraints in (3), is equivalent to a choice of $\{m, R, \tau\}$. Each A_t indicates whether a structural break occurs at time t, as well as which variables experience a break on this date. The entirety of the A matrix then indicates the number of breaks (m), the location of breaks (τ), and which parameters change at each break date (R). Our target distribution from which to obtain draws can then be written as $\Pr(p, m, R, \tau|Y) = \Pr(p, A|Y)$. Our two-block Metropolis-Hastings Sampler will proceed by sampling iteratively from the following two conditional posterior distributions:

$$\Pr(p|A, Y)$$

$$\Pr(A|p, Y)$$

Sampling from $\Pr(p|A, Y)$

To sample from $\Pr(p|A, Y)$ we use a Metropolis-Hastings step that is based on the MC^3 algorithm of Madigan and York (1995). Given the previous draw of p, denoted $p^{[g]}$, we generate a proposal for p, denoted p' by randomly selecting with equal probability from all possible values of p ranging from 0 to $p^{\star}$.[13] Given this proposal, we compute the Metropolis-Hastings acceptance probability:

$$\alpha\left(g,\prime\right)=\min\left(\frac{f(Y|p',A)\Pr\left(p',A\right)}{f(Y|p^{[g]},A)\Pr\left(p^{[g]},A\right)},1\right) \tag{A.1}$$

$p^{[g+1]}$ is then set equal to p' with probability equal to (A.1). Otherwise, $p^{[g+1]}=p^{[g]}$.

The Metropolis-Hastings acceptance probability is calculated as follows. First, the prior probabilities, $\Pr\left(p,A\right)$, are factored as:

$$\Pr\left(p,A\right)=\Pr\left(p,m,R,\tau\right)=\Pr\left(\tau|p,m,R\right)\Pr\left(p,m,R\right),$$

where $\Pr\left(\tau|p,m,R\right)$ and $\Pr\left(p,m,R\right)$ are calculated as in (8) and (9). Second, the marginal likelihood, $f\left(Y|p,A\right)$ is calculated using one of two strategies. Note that conditional on p and A, the structural break model is simply a Normal linear regression model with known regressor matrix 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 is available analytically:

$$f(Y|p,A)=t\left(Y|\Xi\mu,m^{-1}\left(\Xi V\Xi'+I_T\right),b\right), \tag{A.2}$$

where $\Xi=[\iota_T\ \ Z]$ and $t\left(Y|l,k,v\right)$ denotes the probability density function of a T-variate multivariate student-t distribution with location vector l, scale matrix k and degrees of freedom v. In the case where there are structural changes in disturbance variance, the marginal likelihood is no longer available analytically. However, there are a number of

[13]Two aspects of this proposal deserve further comment. First, the proposal distribution is symmetric, such that the probability of proposing a move from p' to $p^{[g]}$ is equal to the probability of proposing a move from $p^{[g]}$ to p'. Second, the proposal allows internally inconsistent models to be proposed if an autoregressive lag is removed in the proposal, but the A matrix indicates that the parameter on this autoregressive lag undergoes breaks. As discussed in Section 3.2, such a model will receive zero prior weight, and will thus receive zero probability of acceptance by the Metropolis-Hastings sampler.

approaches that provide very accurate estimates of the marginal likelihood for such a model with little computational expense. Here, we 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, we found that the procedure produced estimates of the log marginal likelihood across 1000 separate Gibbs runs that were never more than 0.12% apart, even when using a very small number of total Gibbs simulations (100 simulations following 10 burn-in simulations).[14] Such a small number of Gibbs simulations can be produced in approximately one-100^{th} of a second at current computing speeds, making use of the Chib approach as a step in the Metropolis-Hastings sampler feasible.

Sampling from $\Pr(A|p,Y)$

To sample from $\Pr(A|p,Y)$ we again use a Metropolis-Hastings step. Given the previous draw of A, denoted $A^{[g]}$, we generate a proposal, denoted A', by randomly selecting with equal probability from one of two proposal densities, labeled $q_1\left(A|A^{[g]}\right)$ and $q_2\left(A|A^{[g]}\right)$. The first proposal can remove a single existing structural break, can move the timing and / or change which parameters change at a single existing structural break, or can add a single new structural break:

$\boldsymbol{q_1\left(A|A^{[g]}\right)}$

1. Randomly select with equal probability a contiguous block of size b from $A^{[g]}$.

2. Replace this block with a block of size b that indicates at most one break date, chosen

[14]These results are consistent with the Monte Carlo simulations reported in Bos (2002), which demonstrate the Chib (1995) approach provides accurate and robust estimates of the marginal likelihood for the linear regression model.

with equal probability from the set of all possible such blocks. The resulting A matrix is A'.

In experiments, we found that relying only on $q_1\left(A|A^{[g]}\right)$ produced a sampler that would quickly explore areas of high probability in terms of the number and type of breaks, but was slow to explore high probability regions for the timing of structural breaks. The second proposal is designed to improve the sampler in this dimension, by proposing moves in the location of perhaps multiple existing breaks, without changing which parameters change at each break. Define $m^{[g]}$ as the number of breaks indicated by $A^{[g]}$, and let $\widetilde{A}_i$ be a window of A around the i^{th} break date:

$$\widetilde{A}_i = \{A_{\tau_i-4}, A_{\tau_i-3}, A_{\tau_i-2}, A_{\tau_i-1}, A_{\tau_i+1}, A_{\tau_i+2}, A_{\tau_i+3}, A_{\tau_i+4}\}$$

The second proposal then proceeds as follows:

$\boldsymbol{q_2\left(A|A^{[g]}\right)}$

1. Randomly select a number of breaks to move, denoted $\widetilde{m}$, from $0, 1, \ldots, m^{[g]}$.

2. For each $i = 1, \ldots, \widetilde{m}$, replace with equal probability one of the columns in $\widetilde{A}_i$ with the current value of A_{τ_i}. Then, replace A_{τ_i} with a vector of zeros. After these $\widetilde{m}$ replacements are made, the resulting A matrix is A'.

Once a proposal is generated, it is accepted or rejected based on the Metropolis-Hastings acceptance probability:

$$\alpha\left(g, \prime\right) = \min\left(\frac{f(Y|p, A')\Pr\left(p, A'\right)}{f(Y|p, A^{[g]})\Pr\left(p, A^{[g]}\right)}, 1\right) \tag{A.3}$$

which is computed as described in the previous section of this appendix.[15]

[15] Each of the proposal densities are symmetric, such that the probability of proposing a move from A' to $A^{[g]}$ is equal to the probability of proposing a move from $A^{[g]}$ to A'. Again, if an internally inconsistent model is proposed, this model will receive zero prior weight and will be rejected by the Metropolis-Hastings sampler.

Table 1
Monte Carlo Data Generating Processes

	Regime 1 (75 periods)	**Regime 2** . (75 periods)	**Regime 3** (76 periods)
No Break			
Intercept	0.0	0.0	0.0
AR(1)	0.9	0.9	0.9
AR(2)	0.0	0.0	0.0
Variance	1.0	1.0	1.0
Intercept Small			
Intercept	0.7	0.0	-0.7
AR(1)	0.3	0.3	0.3
AR(2)	0.0	0.0	0.0
Variance	1.0	1.0	1.0
Intercept Large			
Intercept	1.4	0.0	-1.4
AR(1)	0.3	0.3	0.3
AR(2)	0.0	0.0	0.0
Variance	1.0	1.0	1.0
Persistence			
Intercept	0.3	0.3	0.3
AR(1)	0.6	1.2	0.6
AR(2)	-0.3	-0.3	-0.3
Variance	0.5	0.5	0.5
Intercept / Variance			
Intercept	0.3	2.1	0.3
AR(1)	0.6	0.6	0.6
AR(2)	-0.3	-0.3	-0.3
Variance	0.5	2.65	0.5

Table 2
Posterior Inclusion Probabilities for Autoregressive Lags

Autoregressive Lag	No Break	Intercept Small	Intercept Large	Persistence	Intercept / Variance
1	**100**	**86**	**83**	**100**	**100**
2	2	1	1	**95**	**93**
3	0	0	0	3	1
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0

Notes: Results shown are averages across Monte Carlo simulations. Bold type in a column indicates an autoregressive lag that is present in the DGP for that column.

Table 3
Posterior Probability for Alternative Numbers of Structural Breaks

Number of Breaks	No Break	Intercept Small	Intercept Large	Persistence	Intercept / Variance
0	**91**	0	0	1	0
1	9	52	0	1	0
2	0	**44**	**88**	**87**	**78**
3	0	3	11	11	20
4	0	0	1	1	2
> 4	0	0	0	0	0

Notes: Results shown are averages across Monte Carlo simulations. Bold type in a column indicates the true number of structural breaks present in the DGP for that column.

Table 4
Mean Number of Structural Breaks in Individual Model Parameters

Parameter	**No Break**	**Intercept Small**	**Intercept Large**	**Persistence**	**Intercept / Variance**
Intercept	0.07 (0)	**1.48 (2)**	**2.09 (2)**	0.10 (0)	**2.04 (2)**
AR(1)	0.01 (0)	0.05 (0)	0.05 (0)	**1.98 (2)**	0.08 (0)
AR(2)	0.00 (0)	0.00 (0)	0.00 (0)	0.06 (0)	0.04 (0)
AR(3)	0.00 (0)	0.00 (0)	0.00 (0)	0.01 (0)	0.00 (0)
AR(4)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
AR(5)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
AR(6)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
Variance	0.02 (0)	0.03 (0)	0.05 (0)	0.04 (0)	**2.04 (2)**

Notes: Results shown are averages across Monte Carlo simulations. Entries in parentheses give the true number of structural breaks for the respective parameter and DGP. Bold type in a column indicates a parameter that experienced structural breaks in the DGP for that column.

Table 5
Average RMSE of Parameter Estimates

Parameter	**No Break**	**Intercept Small**	**Intercept Large**	**Persistence**	**Intercept / Variance**
Intercept	0.07	0.25	0.24	0.06	0.23
AR(1)	0.03	0.09	0.09	0.12	0.06
AR(2)	0.00	0.00	0.00	0.07	0.07
AR(3)	0.00	0.00	0.00	0.00	0.00
AR(4)	0.00	0.00	0.00	0.00	0.00
AR(5)	0.00	0.00	0.00	0.00	0.00
AR(6)	0.00	0.00	0.00	0.00	0.00

Notes: RMSE is the root mean squared error of the Bayesian posterior mean from the true parameter value, calculated from $t = 1, \ldots, T$ for a single Monte Carlo simulation. Average RMSE is the RMSE averaged across Monte Carlo simulations.

Table 6
Data Series

Series Name	Mnemonic	Transformation	Sample
Wages and Prices			
Hourly Earnings of Production and Nonsupervisory Employees	AHETPI	Growth Rates	1964:Q2 - 2018:Q2
Consumer Price Index: Excluding Food and Energy	CPILFESL	Growth Rates	1959:Q1 - 2018:Q2
Personal Consumption Expenditures Price Index: Excluding Food and Energy	PCEPILFE	Growth Rates	1959:Q2 - 2018:Q2
Production			
Real Gross Domestic Product	GDPC1	Growth Rates	1959:Q1 - 2018:Q2
Real Gross Private Domestic Investment	GPDIC1	Growth Rates	1959:Q1 - 2018:Q2
Real Personal Consumption Expenditures	PCECC96	Growth Rates	1959:Q1 - 2018:Q2
Employment			
Total Nonfarm Payroll Employment	PAYEMS	Growth Rates	1959:Q1 - 2018:Q2
Civilian Unemployment Rate	UNRATE	Levels	1959:Q1 - 2018:Q2

Notes: All data series were obtained from the Federal Reserve Economic Database (FRED): `https://fred.stlouisfed.org`. Mnemonics refer to FRED series identifiers.

Table 7
Posterior Inclusion Probabilities for Autoregressive Lags

Series Name	**Autoregressive Lag**					
	1	2	3	4	5	6
Wages and Prices						
Hourly Earnings of Production and Nonsupervisory Employees	100	100	11	0	0	0
Consumer Price Index: Excluding Food and Energy	100	100	1	0	0	0
Personal Consumption Expenditures Price Index: Excluding Food and Energy	100	4	0	0	0	0
Production						
Real Gross Domestic Product	100	65	0	0	0	0
Real Gross Private Domestic Investment	2	0	0	0	0	0
Real Personal Consumption Expenditures	100	100	82	1	0	0
Employment						
Total Nonfarm Payroll Employment	100	4	0	0	0	0
Civilian Unemployment Rate	100	100	7	0	0	0

Table 8
Posterior Probability of Alternative Numbers of Structural Breaks

Series Name	**Number of Breaks**							
	0	1	2	3	4	5	6	> 6
Wages and Prices								
Hourly Earnings of Production and Nonsupervisory Employees	0	1	11	53	34	1	0	0
Consumer Price Index: Excluding Food and Energy	0	0	15	32	50	2	0	0
Personal Consumption Expenditures Price Index: Excluding Food and Energy	0	0	0	0	81	18	1	0
Production								
Real Gross Domestic Product	0	77	22	0	0	0	0	0
Real Gross Private Domestic Investment	0	1	1	90	7	0	0	0
Real Personal Consumption Expenditures	0	90	10	0	0	0	0	0
Employment								
Total Nonfarm Payroll Employment	0	0	0	63	37	0	0	0
Civilian Unemployment Rate	9	17	68	6	0	0	0	0

Notes: Probabilities may not sum to one due to rounding.

Table 9
Posterior Probability of Alternative Numbers of Structural Breaks in Intercept

Series Name	**Number of Breaks in Intercept**							
	0	1	2	3	4	5	6	> 6
Wages and Prices								
Hourly Earnings of Production and Nonsupervisory Employees	20	57	23	1	0	0	0	0
Consumer Price Index: Excluding Food and Energy	66	33	1	0	0	0	0	0
Personal Consumption Expenditures Price Index: Excluding Food and Energy	96	3	1	0	0	0	0	0
Production								
Real Gross Domestic Product	92	8	0	0	0	0	0	0
Real Gross Private Domestic Investment	99	1	0	0	0	0	0	0
Real Personal Consumption Expenditures	93	7	0	0	0	0	0	0
Employment								
Total Nonfarm Payroll Employment	98	2	0	0	0	0	0	0
Civilian Unemployment Rate	98	2	0	0	0	0	0	0

Notes: Probabilities may not sum to one due to rounding.

Table 10
Posterior Probability of Alternative Numbers of Structural Breaks in Autoregressive Parameters

Series Name	Number of Breaks in AR Parameters							
	0	1	2	3	4	5	6	> 6
Wages and Prices								
Hourly Earnings of Production and Nonsupervisory Employees	56	43	1	0	0	0	0	0
Consumer Price Index: Excluding Food and Energy	16	60	22	2	0	0	0	0
Personal Consumption Expenditures Price Index: Excluding Food and Energy	0	0	65	33	2	0	0	0
Production								
Real Gross Domestic Product	97	3	0	0	0	0	0	0
Real Gross Private Domestic Investment	100	0	0	0	0	0	0	0
Real Personal Consumption Expenditures	96	4	0	0	0	0	0	0
Employment								
Total Nonfarm Payroll Employment	98	2	0	0	0	0	0	0
Civilian Unemployment Rate	99	1	0	0	0	0	0	0

Notes: Results shown are the posterior probability for a break in any autoregressive parameter. Probabilities may not sum to one due to rounding.

Table 11
Posterior Probability of Alternative Numbers of Structural Breaks in Conditional Variance

Series Name	**Number of Breaks in Conditional Variance**							
	0	1	2	3	4	5	6	> 6
Wages and Prices								
Hourly Earnings of Production and Nonsupervisory Employees	0	0	53	46	0	0	0	0
Consumer Price Index: Excluding Food and Energy	0	0	100	0	0	0	0	0
Personal Consumption Expenditures Price Index: Excluding Food and Energy	0	0	100	0	0	0	0	0
Production								
Real Gross Domestic Product	0	86	14	0	0	0	0	0
Real Gross Private Domestic Investment	0	1	1	91	7	0	0	0
Real Personal Consumption Expenditures	0	98	2	0	0	0	0	0
Employment								
Total Nonfarm Payroll Employment	0	0	0	64	36	0	0	0
Civilian Unemployment Rate	9	17	69	5	0	0	0	0

Notes: Probabilities may not sum to one due to rounding.

Figure 1
Probability of Structural Break in Intercept Parameter

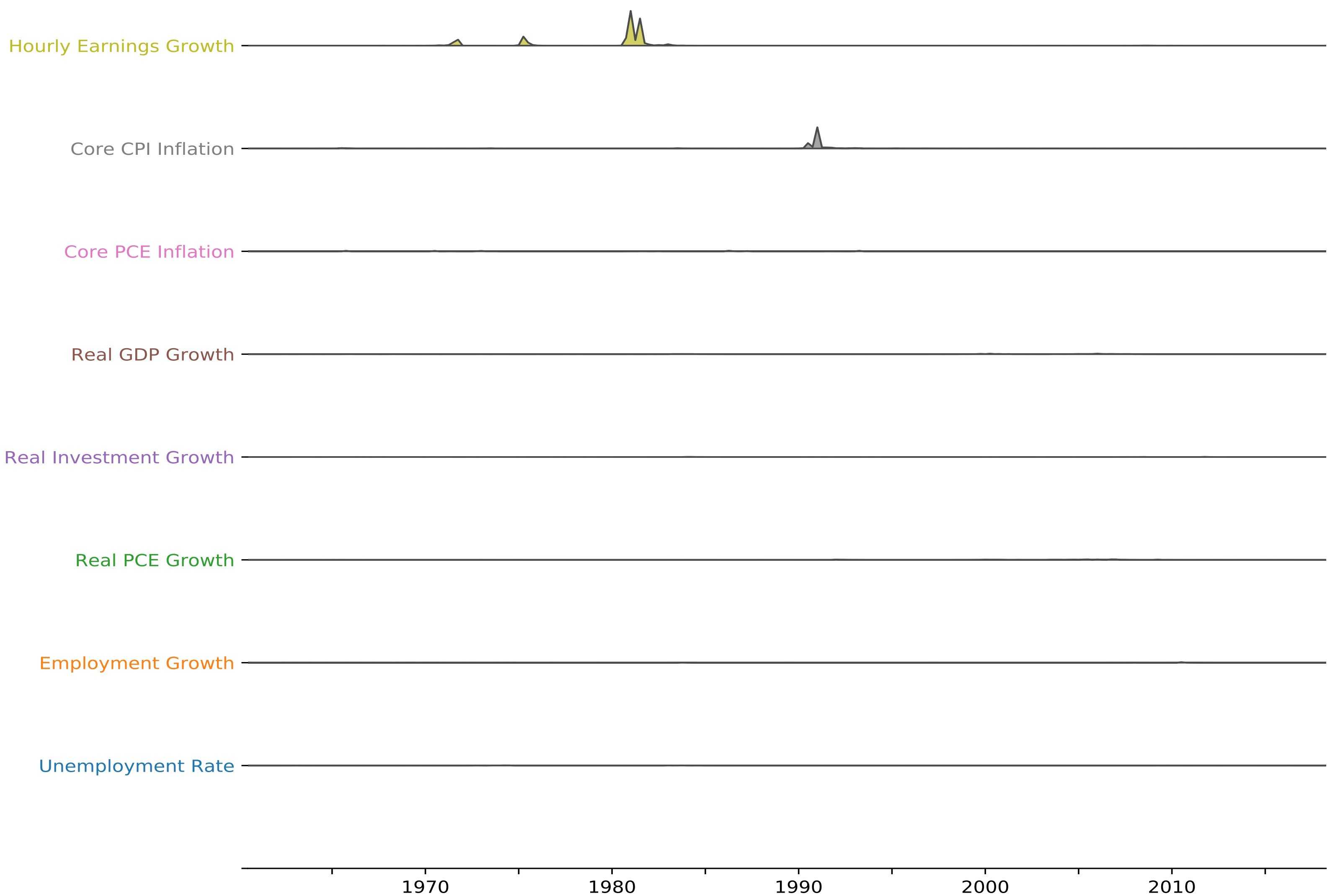


Figure 2
Probability of Structural Break in AR Parameters

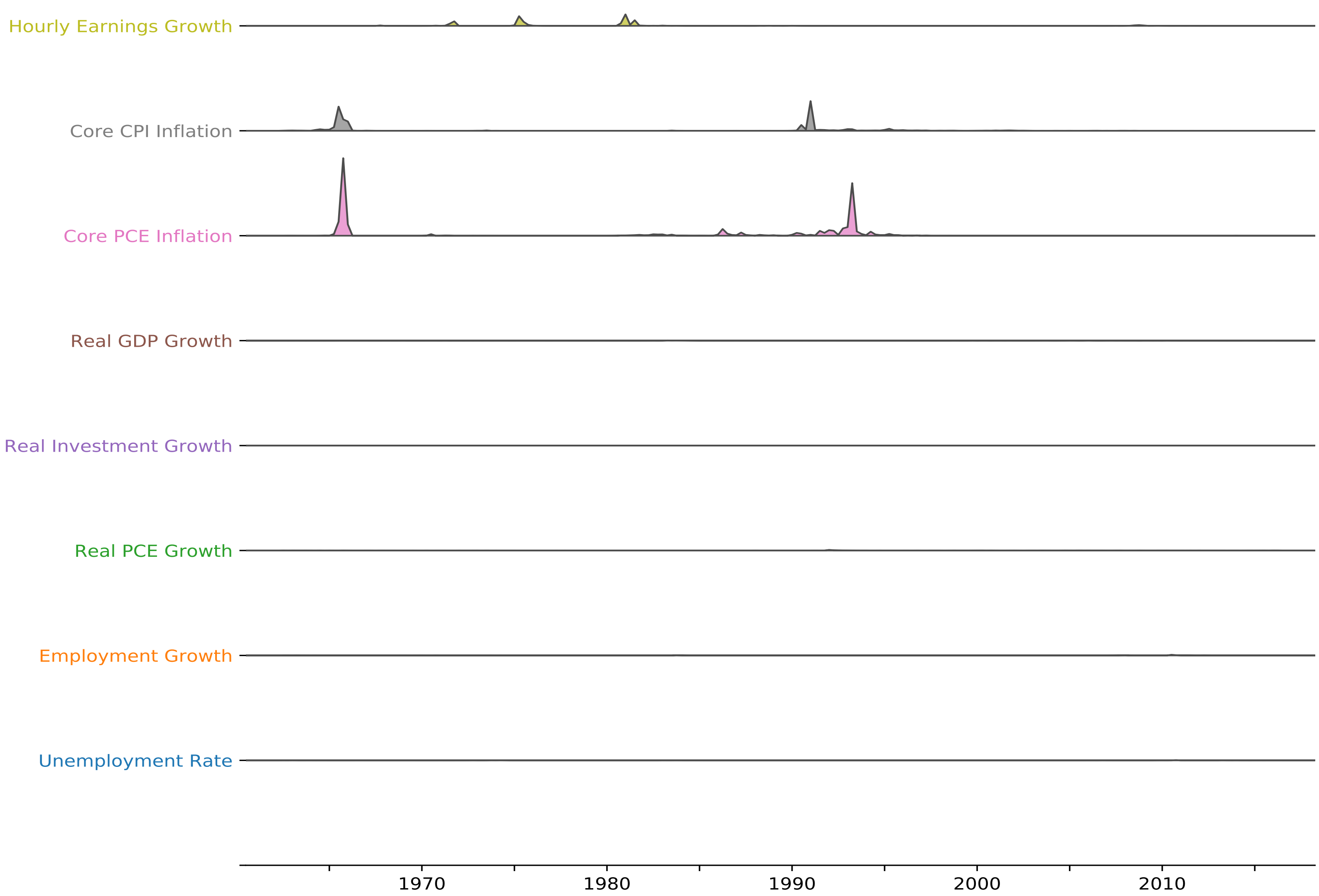


Figure 3
Probability of Structural Break in Conditional Variance Parameter

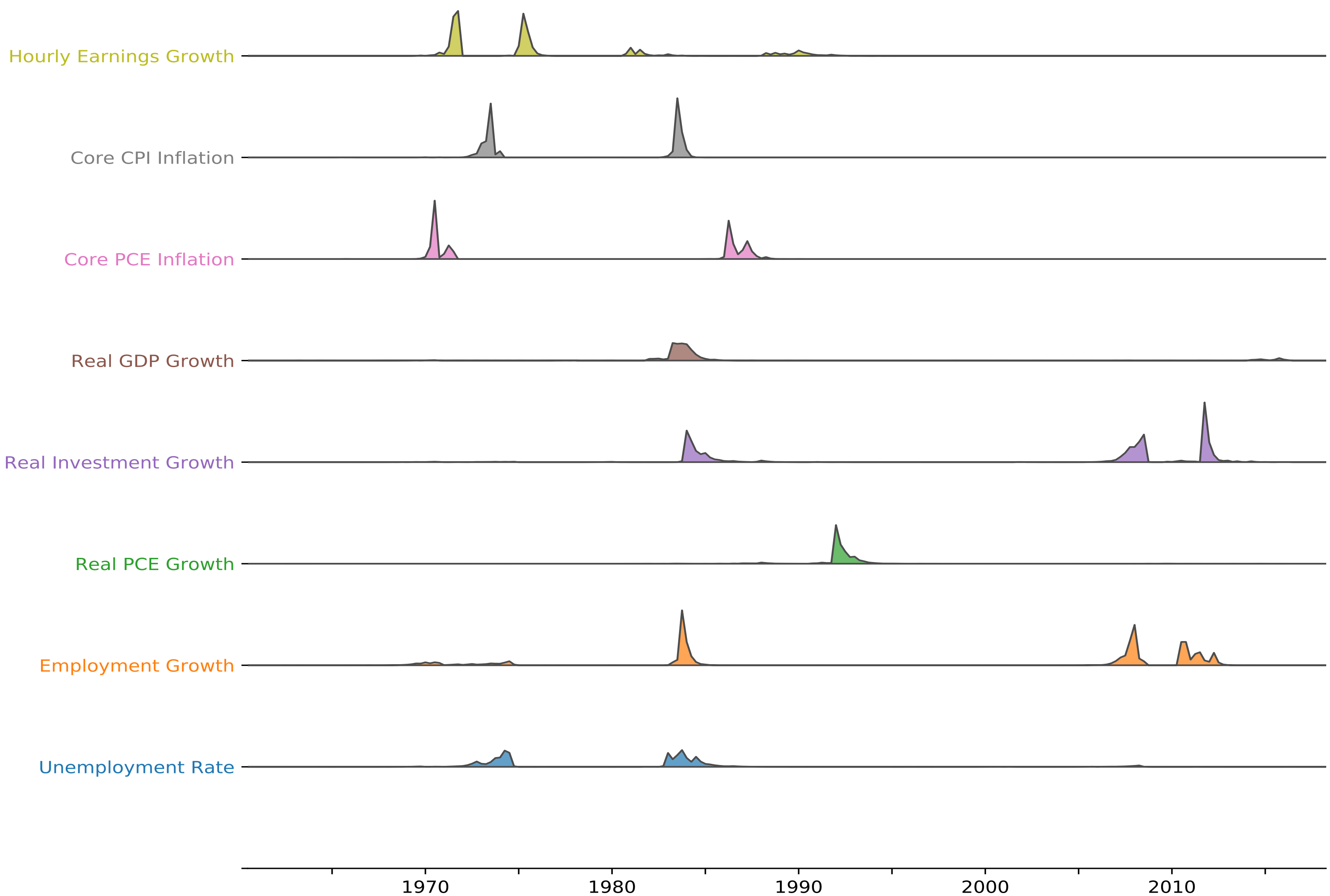


Figure 4
BMA Unconditional Mean of Annualized Real GDP Growth

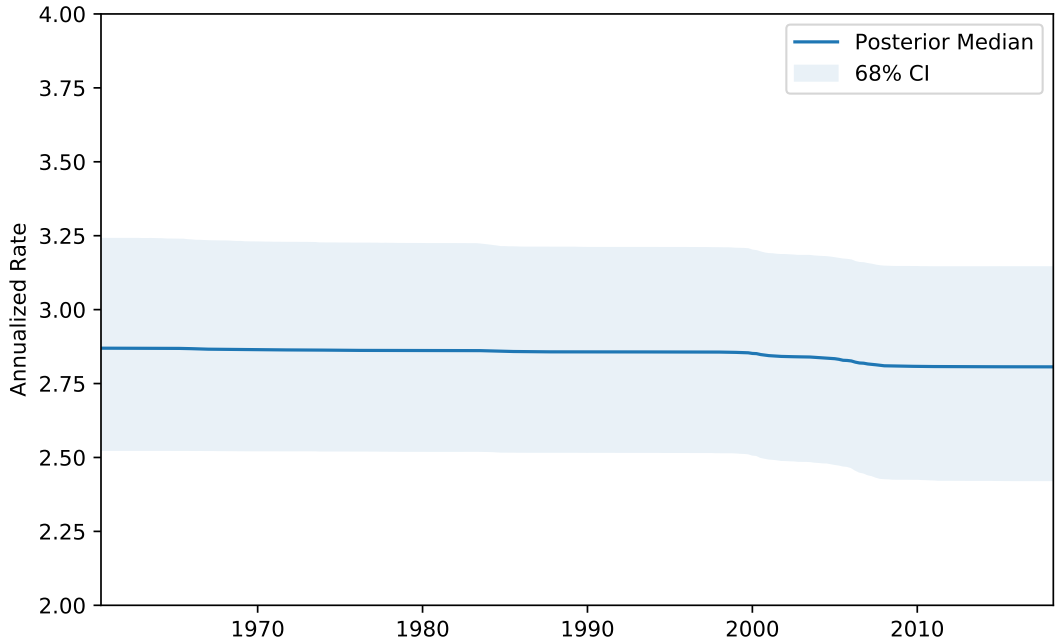


Figure 5
BMA Unconditional Mean of Annualized Real PCE Growth

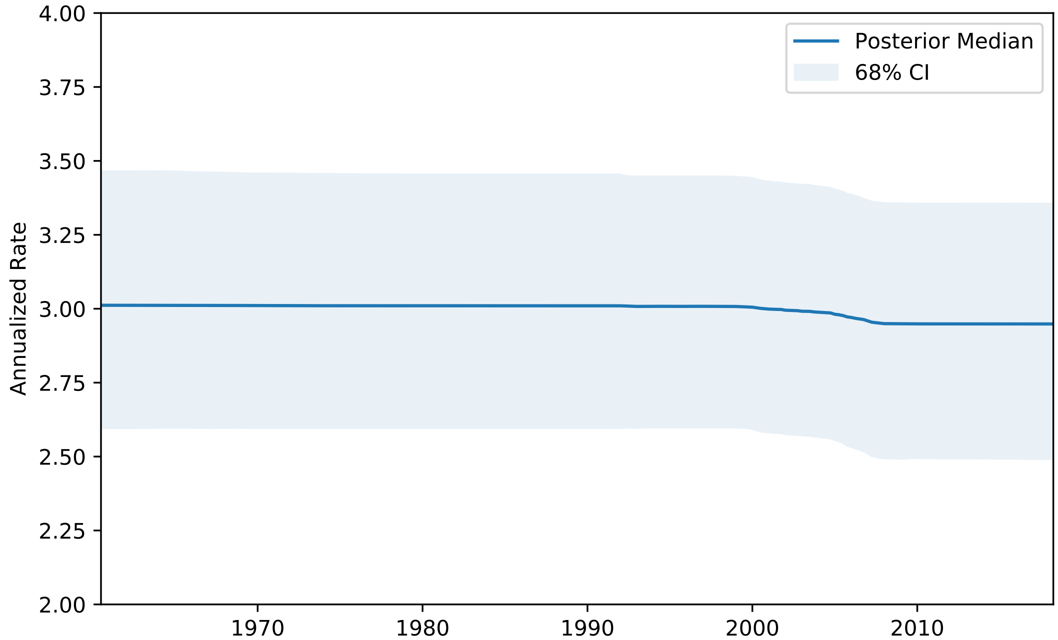


Figure 6
BMA Unconditional Mean of Unemployment Rate

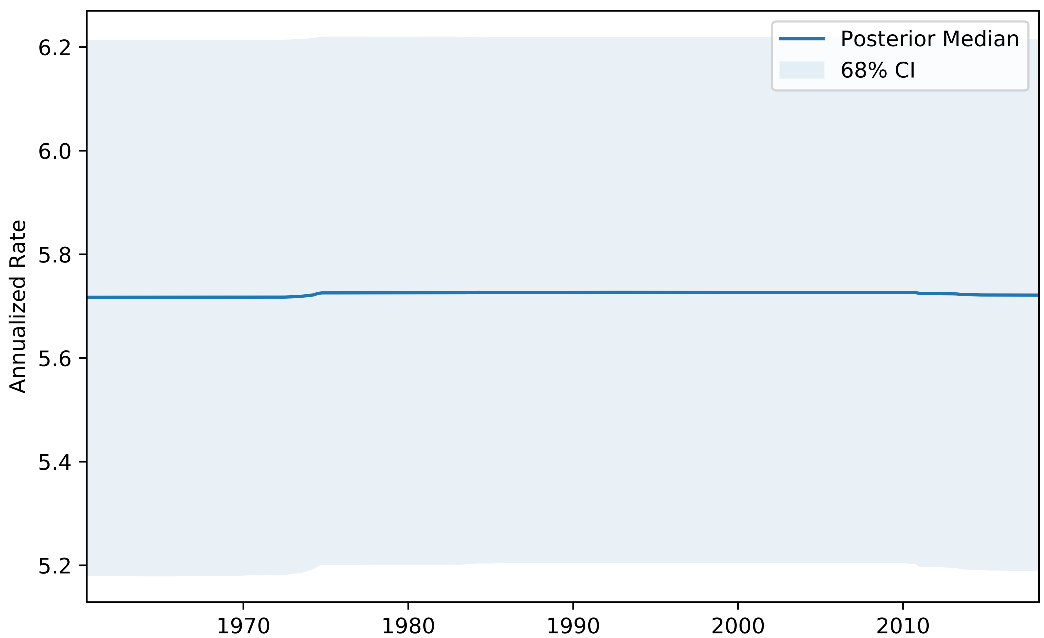


Figure 7
BMA Sum of Autoregressive Coefficients for Inflation Series

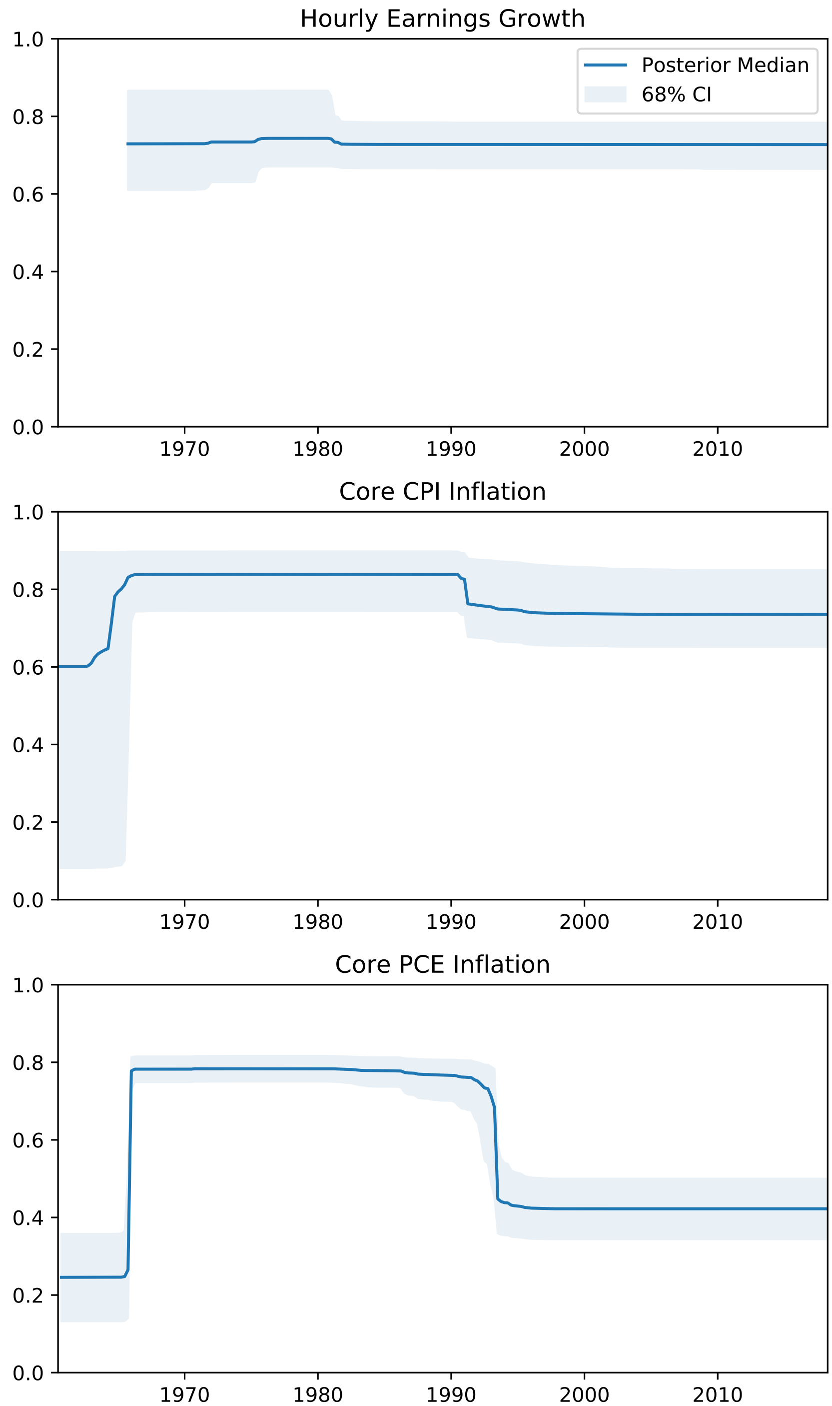


Figure 8
BMA Unconditional Mean of Annualized Inflation

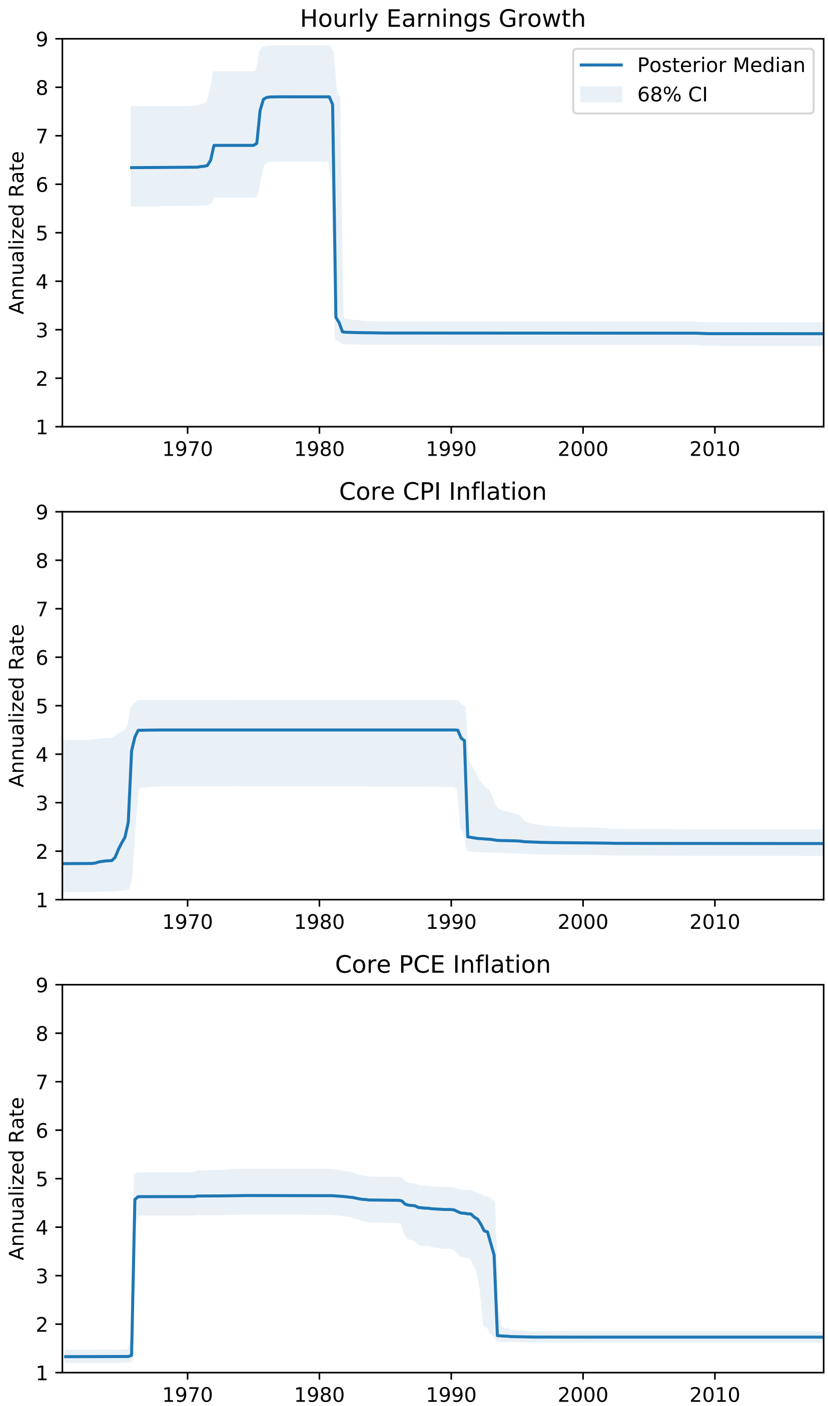


Figure 9
BMA Unconditional Standard Deviation of Production Series

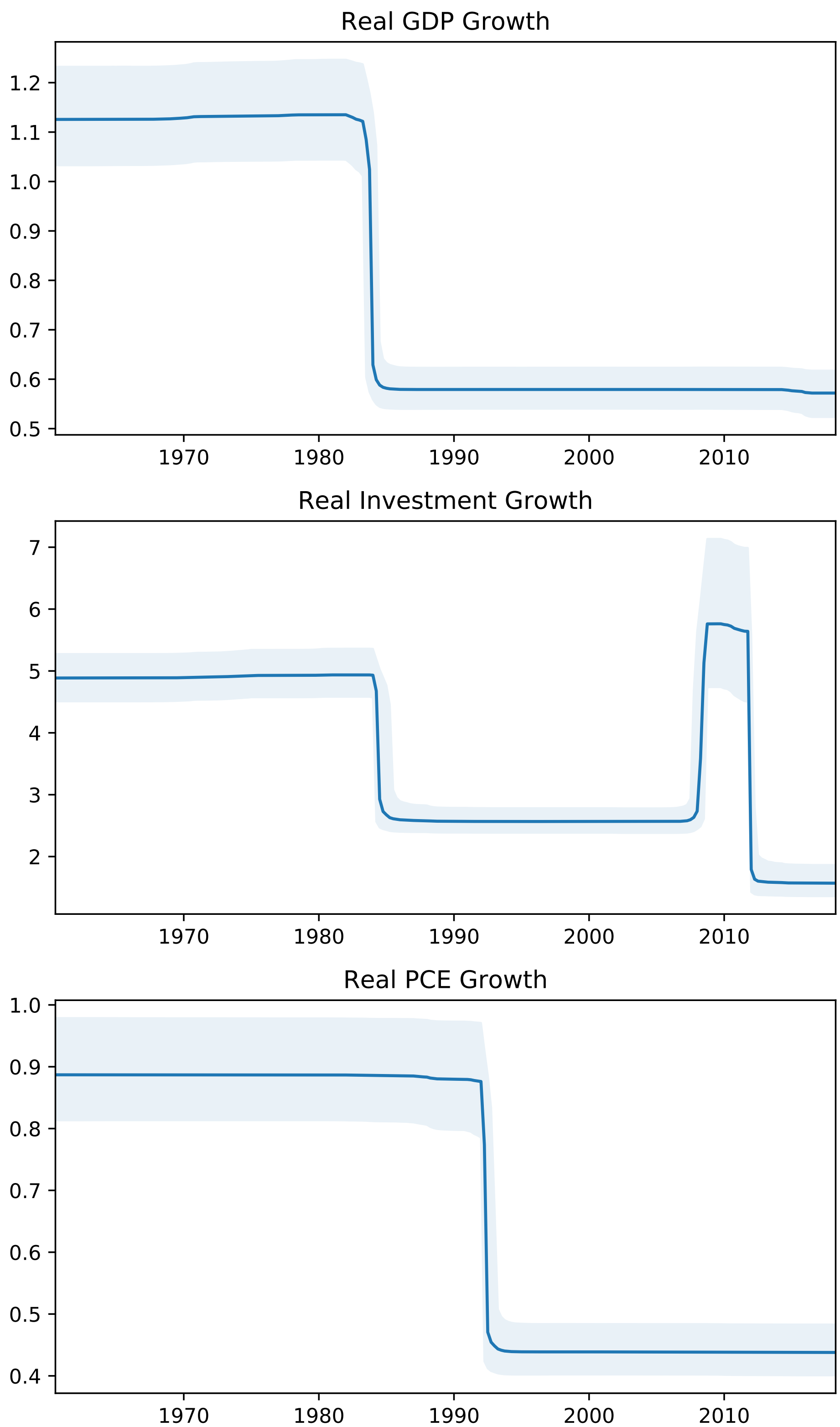


Figure 10
BMA Unconditional Standard Deviation of Inflation Series

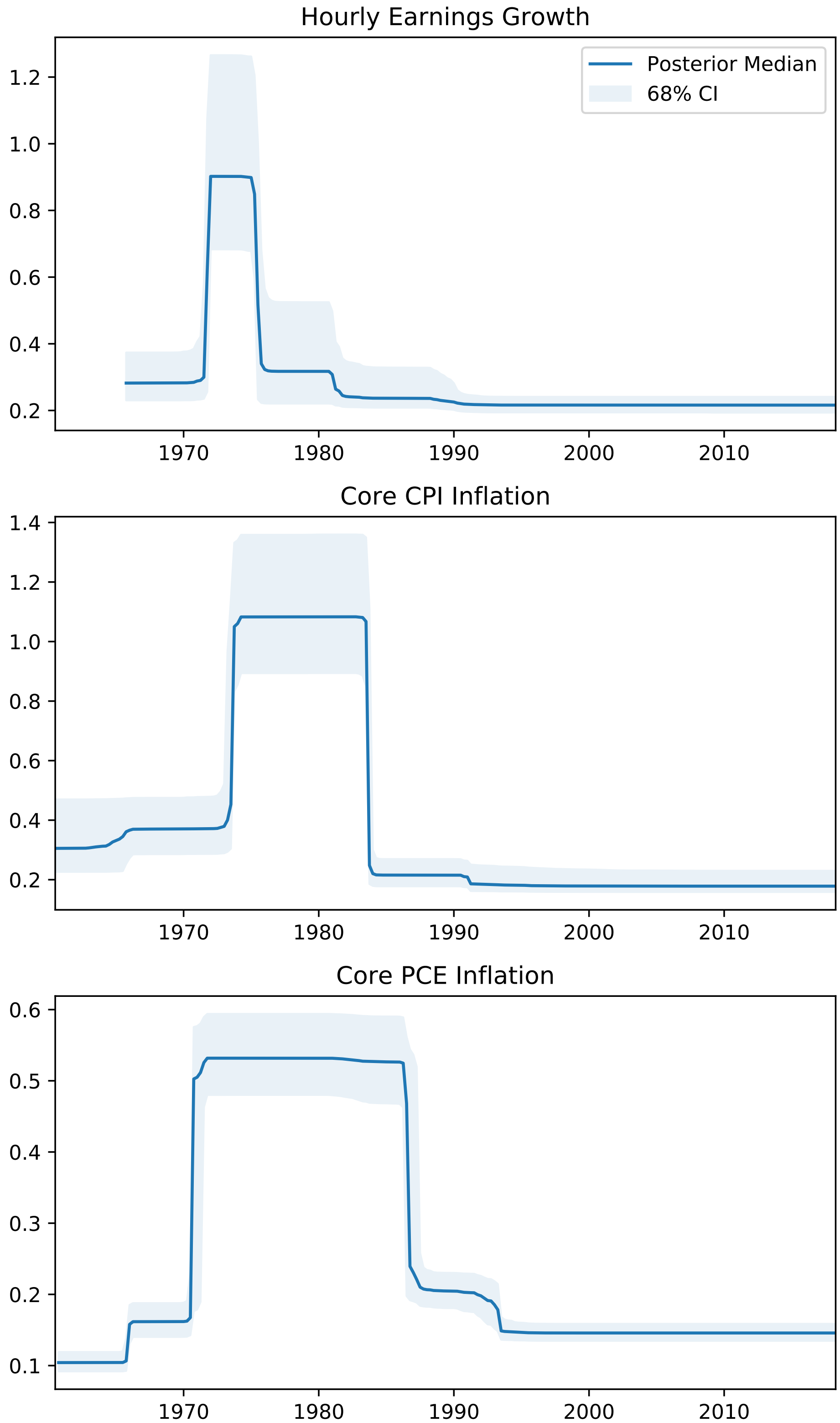
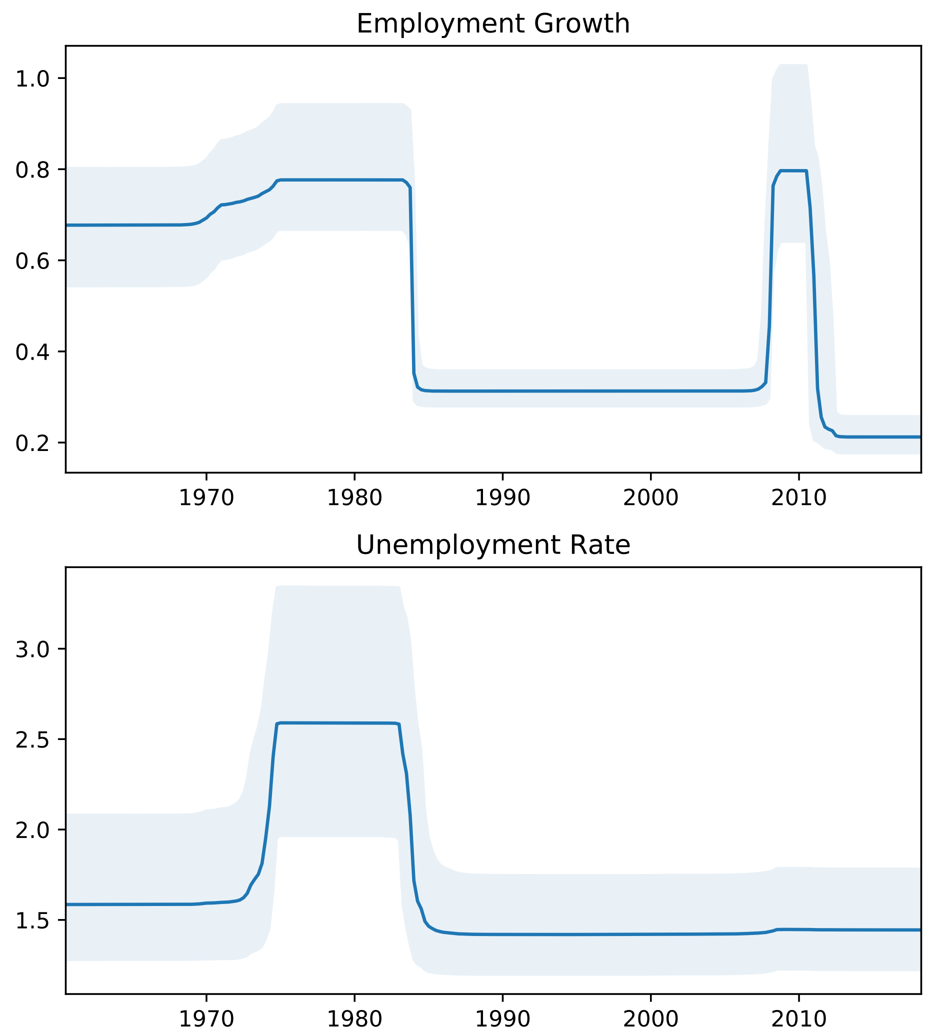


Figure 11
BMA Unconditional Standard Deviation of Labor Market Series



References

Andrews, D. W. and W. Ploberger (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics 91*(2), 299–323.

Andrews, D. W. K. and H. Y. Chen (1994). Approximately median-unbiased estimation of autoregressive models. *Journal of Business and Economics Statistics 12*(2), 187–204.

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2017). Tracking the slowdown in long-run gdp growth. *Review of Economics and Statistics 99*(2), 343–356.

Bai, J., R. L. Lumsdaine, and J. Stock (1998). Testing for and dating common breaks in multivariate time series. *Review of Economic Studies 65*(3), 395–432.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Barsky, R. (1987). The fisher hypothesis and the forecastibility and persistence of inflation. *Journal of Monetary Economics 19*, 3–24.

Benati, L. (2008). Investigating inflation persistence across monetary regimes. *Quarterly Journal of Economics 123*(3), 1005–1060.

Blonigen, B. and J. Piger (2014). Determinants of foreign direct investment. *Canadian Journal of Economics 47*(3), 775–812.

Bos, C. (2002). A comparison of marginal likelihood computation methods. In W. Hardle and B. Ronz (Eds.), *Compstat*, Heidelberg, pp. 111–116. Physica.

Charles, A., O. Darné, and L. Ferrara (2018). Does the great recession imply the end of the great moderation? international evidence. *Economic Inquiry 56*(2), 745–760.

Chen, Y.-C. and E. Zivot (2010). Postwar slowdowns and long-run growth: A bayesian analysis of structural break models. *Empirical Economics 39*(3), 897–921.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association 90*(432), 1313–1321.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics 86*(2), 221–241.

Clark, T. E. (2006). Disaggregate evidence on the persistence of consumer price inflation. *Journal of Applied Econometrics 21*(5), 563–587.

Cogley, T. and T. J. Sargent (2001). Evolving post-world war ii u.s. inflation dynamics. *NBER Macroeconomics Annual 16*, 331–372.

Diebold, F. X. and C. Chen (1996). Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures. *Journal of Econometrics 70*(1), 221–241.

Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B 57*(1), 45–97.

Eo, Y. (2016). Structural changes in inflation dynamics: Multiple breaks at different dates for different parameters. *Studies in Nonlinear Dynamics and Econometrics 20*(3), 211–231.

Eo, Y. and J. Morley (2018). Why has the u.s. economy stagnated since the great recession? SSRN working paper: https://ssrn.com/abstract=3077864.

Evans, M. and P. Wachtel (1993). Inflation regimes and the sources of inflation uncertainty. *Journal of Money, Credit and Banking 25*, 475–511.

Fernández, C., E. Ley, and M. F. Steel (2001a). Benchmark priors for bayesian model averaging. *Journal of Econometrics 100*(2), 381–427.

Fernández, C., E. Ley, and M. F. Steel (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics 16*(5), 563–576.

Gerlach, R., C. Carter, and R. Kohn (2000). Efficient bayesian inference for dynamic mixture models. *Journal of the American Statistical Association 95*(451), 819–828.

Geweke, J. and Y. Jiang (2011). Inference and prediction in a multiple-structural-break model. *Journal of Econometrics 163*(2), 172–185.

Giordani, P. and R. Kohn (2008). Efficient bayesian inference for multiple change-point and mixture innovation models. *Journal of Business and Economic Statistics 26*(1), 66–77.

Grant, A. and J. C. Chan (2017). A bayesian model comparison for trend-cycle decompositions of output. *Journal of Money, Credit and Banking 49*(2-3), 525–552.

Groen, J. J. J., R. Paap, and F. Ravazzolo (2013). Real-time inflation forecasting in a changing world. *Journal of Business and Economic Statistics 31*(1), 29–44.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in u.s. labor productivity. *Journal of Economic Perspectives 15*(4), 117–128.

Hodges, J. S. (1987). Uncertainty, policy analysis, and statistics. *Statistical Science 2*(3), 259–275.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: A tutorial (with discussion). *Statistical Science 14*(4), 382–401.

Hultblad, B. and S. Karlsson (2008). Bayesian simultaneous determination of structural breaks and lag lengths. *Studies in Nonlinear Dynamics and Econometrics 12*(3), 1–29.

Inclán, C. (1993). Detection of multiple changes of variance using posterior odds. *Journal of Business and Economic Statistics 11*(3), 289–300.

Inoue, A. and B. Rossi (2008). Identifying the sources of instabilities in macroeconomic fluctuations. *The Review of Economics and Statistics 93*(4), 1186–1204.

Kamber, G., J. Morley, and B. Wong (2018). Intuitive and reliable estimates of the output gap from a beveridge-nelson filter. *Review of Economics and Statistics 100*, 550–566.

Kang, K. H., C.-J. Kim, and J. Morley (2009). Changes in u.s. inflation persistence. *Studies in Nonlinear Dynamics and Econometrics 13*(4).

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

Kim, C.-J. and C. R. Nelson (1999). Has the u.s. economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. *Review of Economics and Statistics 81*(4), 608–616.

Koop, G. and S. M. Potter (2007). Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies 74*(3), 763–789.

Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* New York, NY: John Wiley and Sons, Inc.

Levin, A. T. and J. Piger (2008). Bayesian model selection for structural break models. SSRN working paper: https://ssrn.com/abstract=1132463.

Luo, S. and R. Startz (2014). Is it one break or ongoing permanent shocks that explains us real gdp? *Journal of Monetary Economics 66*, 155–163.

Madigan, D. and J. York (1995). Bayesian graphical models for discrete data. *International Statistical Review 63*(2), 215–232.

Maheu, J. and Y. Song (2014). A new structural break model, with an application to canadian inflation forecasting. *International Journal of Forecasting 30*(1), 144–160.

McConnell, M. M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980's? *American Economic Review 90*(5), 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*(1-2), 169–193.

Papell, D. H., C. J. Murray, and H. Ghiblawi (2000). The structure of unemployment. *The Review of Economics and Statistics 82*(2), 309–315.

Perron, P. (1989). The great crash, the oil shock and the unit root hypothesis. *Econometrica 57*(6), 1361–1401.

Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics 8*(2), 153–162.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies 73*(4), 1057–1084.

Pivetta, F. and R. Reis (2007). The persistence of inflation in the united states. *Journal of Economic Dynamics and Control 31*(4), 1326–1358.

Prodan, R. (2008). Potential pitfalls in determining multiple structural changes with an application to purchasing power parity. *Journal of Business and Economic Statistics 26*(1), 50–65.

Raftery, A. E. (1996). Approximate bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika 83*(2), 251–266.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association 92*(437), 179–191.

Rapach, D. E. and M. E. Wohar (2005). Regime changes in international real interest rates: Are they a monetary phenomenon? *Journal of Money, Credit and Banking 37*(5), 887–906.

Sensier, M. and D. van Dijk (2004). Testing for volatility changes in u.s. macroeconomic time series. *The Review of Economics and Statistics 86*(3), 833–839.

Staiger, D., J. Stock, and M. Watson (1997). How precise are estimates of the natural rate of unemployment. In C. Romer and D. Romer (Eds.), *Reducing Inflation: Motivation and Strategy*, Chicago, pp. 195–242. University of Chicago Press for the National Bureau of Economic Research.

Steel, M. F. (2017). Model averaging and its use in economics. CRiSM Working Paper 17-06.

Stock, J. H. (2001). Comment on evolving post-world war ii u.s. inflation dynamics. *NBER Macroeconomics Annual 16*, 379–387.

Stock, J. H. and M. W. Watson (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*, 159–218.

Stock, J. H. and M. W. Watson (2007). Why has u.s. inflation become harder to forecast? *Journal of Money, Credit and Banking 39*(1), 3–33.

Summers, P. M. (2004). Bayesian evidence on the structure of unemployment. *Economics Letters 83*(3), 299–306.

Wang, J. and E. Zivot (2000). A bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business and Economic Statistics 18*(3), 374–386.