DE GRUYTER Stud. Nonlinear Dyn. E. 2024; 28(2): 177–199

Shih-Tang Hwu and Chang-Jin Kim*

Markov-Switching Models with Unknown Error Distributions: Identification and Inference Within the Bayesian Framework

https://doi.org/10.1515/snde-2022-0055
Received July 2, 2022; accepted October 19, 2023; published online November 28, 2023

Abstract: The basic Markov-switching model has been extended in various ways ever since the seminal work of Hamilton (1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57: 357–84). However, the estimation of Markov-switching models in the literature has relied upon parametric assumptions on the distribution of the error term. In this paper, we present a Bayesian approach for estimating Markov-switching models with unknown and potentially non-normal error distributions. We approximate the unknown distribution of the error term by the Dirichlet process mixture of normals, in which the number of mixtures is treated as a parameter to estimate. In doing so, we pay special attention to the identification of the model. We then apply the proposed model and MCMC procedure to the growth of the postwar U.S. industrial production index. Our model can effectively control for irregular components that are not related to business conditions. This leads to sharp and accurate inferences on recession probabilities.

Keywords: label switching problem; identification condition; unknown error distribution; mixture of normals; semi-parametric Bayesian inference; Markov chain Monte Carlo

JEL Classification: C11; C13; C22

1 Introduction

Since the seminal work of Hamilton (1989), the basic Markov-switching model has been extended in various ways. For example, Diebold, Lee, and Weinbach (1994) and Filardo (1994) extend the model to allow the transition probabilities governing the Markov process to be functions of exogenous or predetermined variables. Kim (1994) extends it to the case of the state-space model, which encompasses general dynamic models such as autoregressive moving average processes, unobserved components models, dynamic factor models, etc. Chib (1998) introduces a structural break model with unknown multiple change points by constraining the transition probabilities of the Markov-switching model so that the latent state variable can either stay at the current value or jump to the next higher value.[1] More recently, Kaufmann (2015) proposes a general K-state model with time-varying transition probabilities by employing the multinomial Logit specification. Fox et al. (2011), Song (2014), and Bauwens, Carpantier, and Dufays (2017) introduce infinite hidden Markov models and generalize the finite state Markov switching model of Hamilton (1989) to the case of an infinite number of states.

Estimations of the aforementioned models and the other Markov-switching models in the literature have relied upon parametric assumptions on the distribution of the error term. Most applications in the literature

1 For surveys of earlier literature on Markov switching models, refer to Frühwirth-Schnatter (2006) and Hamilton (2016).

***Corresponding author: Chang-Jin Kim**, Department of Economics, University of Washington, Seattle, USA, E-mail: changjin@uw.edu
Shih-Tang Hwu, Department of Economics, California State Polytechnic University, Pomona, USA, E-mail: shwu@cpp.edu

assume normally distributed errors, with rare exceptions like Dueker (1997) and Bulla et al. (2011) who proposed Markov-switching models of stock returns in which the innovations are assumed to be drawn from a Student-*t* distribution; and Angelis and Cinzia (2017) who assumed a normal-inverse Gaussian distribution as the conditional form of financial returns and model innovations.

In this paper, we deal with a Bayesian semi-parametric approach to making inferences on the Markov-switching model when the unknown error distribution is approximated by the mixture of normals.[2] We address two identification issues that are necessary for the estimation of the model. They include: (i) the problem of label switching for the Markov-switching regime indicator variable; and (ii) the problem of disentangling the Markov-switching regime indicator variable from the serially independent mixture indicator variable. If we do not take care of these identification issues, the marginal posterior distributions of the model parameters obtained from the MCMC output may be misleading. Without loss of generality, these issues are discussed within a basic model with no serial correlation or heteroscedasticity in the error term. We then present an MCMC procedure for estimating a generalized version of the model that allows for serial dependence as well as heteroscedasticity in the error term. In our generalized model, we approximate the unknown distribution of the error term by the Dirichlet process mixtures of normals. Our simulation study shows that the identification schemes and the proposed MCMC procedure work well.

We apply the proposed model and the MCMC algorithm to the monthly index of industrial production (1947:M1–2019:M9). It turns out that the posterior mean for the number of mixtures is about 3. The estimates of the recession probabilities from the proposed model are much sharper and agree much more closely with the NBER reference cycles than those from a model with a normality assumption. Besides, while results from the model with normality assumption are very sensitive to the priors employed, those from the proposed model are robust to them.

The rest of this paper is organized as follows. In Section 2, we motivate our paper by performing a Monte Carlo experiment, which is designed to investigate the effect of maximizing a normal log-likelihood when the normality assumption is violated for the error term. In Section 3, we deliver the two identification issues necessary for the estimation of the model when the unknown error distribution is approximated by the mixture of normals. In Section 4, we present an MCMC algorithm for making inferences on the model. In Section 5, we apply the proposed identification schemes and the MCMC algorithm to the log-differenced monthly postwar U.S. industrial production index [1947:M1–2019:M9]. Section 6 concludes the paper.

2 Pitfalls of Ignoring Non-normality and Maximizing a Normal Log Likelihood

In order to investigate the small-sample performance of the maximum likelihood estimation when a normal log-likelihood is maximized but the normality assumption is violated, we consider the following simple model with Markov-switching mean:

$$y_t = \beta_{S_t} + \sigma\varepsilon_t^*, \quad \varepsilon_t^* \sim (0,1), \quad S_t = 1,2, \quad t = 1,2,\ldots,T, \tag{1}$$

where ε_t^* is independently distributed and S_t is a 2-state Markov-switching process with transition probabilities

$$\Pr[S_t = 1|S_{t-1} = 1] = p_{S,11}, \quad \Pr[S_t = 2|S_{t-1} = 2] = p_{S,22}. \tag{2}$$

2 Defining x_t and y_t as a vector of covariates and the response variable, respectively, Taddy and Kottas (2009) consider a model in which the unknown joint distribution of x_t and y_t depends upon a latent state variable (S_t) that follows a $K-$ state first-order Markov-switching process. On the contrary, we deal with a model in which the distribution of y_t conditional on x_t, S_t, and past information is unknown. An example of our model is the Hamilton model (1989) in which the normally distributed error term is replaced by the Dirichlet process mixture of normals.

We consider the following four alternative distributions for the error term ε_t, the first two of which are symmetric and the other two are asymmetric:

Case #1: $\varepsilon_t^* \sim i.i.d.\ \ N(0,1)$
Case #2: $\varepsilon_t^* = \frac{u_t}{\sqrt{\nu/(\nu-2)}},\quad u_t \sim i.i.d.\ \ \text{t}-\text{distribution}\ \ \text{with}\ \ d.f. = \nu$
Case #3: $\varepsilon_t^* = \frac{\ln(u_t^2)-E(\ln(u_t^2))}{\sqrt{(var(\ln(u_t^2))}}, u_t \sim i.i.d.N(0,1)$, where $E(\ln\ u_t^2) = -1.2704$, and $var(\ln\ u_t^2) = \pi^2/2$.
Case #4: $\varepsilon_t^*|D_t \sim i.i.d.N\left(\mu_{D_t}^*, h_{D_t}^{*2}\right),\ D_t = 1,2,3$, where $\Pr[D_t = i] = p_{D,i}, i = 1,2,3$

For each of the above four cases, we generate 10,000 sets of data. For each data set generated, we estimate the model in equations (1) and (2) by maximizing a normal log-likelihood. We consider three alternative sample sizes: $T = 500$, $T = 5,000$ and $T = 50,000$. The parameter values we assign are given below:

$$\beta_1 = -0.6,\quad \beta_2 = 0.7;\quad \sigma^2 = 1.1;\quad p_{S,11} = 0.9,\quad p_{S,22} = 0.95;\quad \nu = 5;$$

$$\mu_1^* = 1.05,\quad \mu_2^* = 0.1,\quad \mu_3^* = -1.35;\quad h_1^{*2} = 0.2,\quad h_2^{*2} = 0.05,\quad h_3^{*2} = 1.695;$$

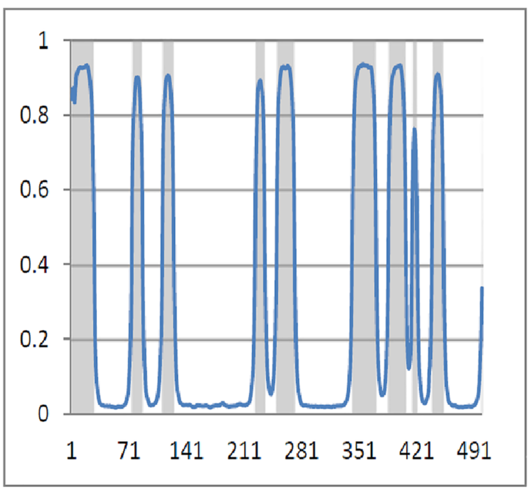
$$p_{D,1} = 0.2,\quad p_{D,2} = 0.6,\quad p_{D,3} = 0.2$$

Table 1 reports the mean of the estimates for each parameter in each case, as well as the root mean squared error (RMSE) for the estimates. For all the cases we consider, the maximum likelihood estimation seems to result in consistent parameter estimates, in the sense that both the biases and RMSE's decrease as the sample size increases. When the normality assumption is violated, however, the maximization of a normal log likelihood results in poor small sample properties of the estimators. In particular, in a situation like Case #4 in which the degree of asymmetry in the error distribution is the highest, the biases remain sizable even when the sample size is as large as 50,000.

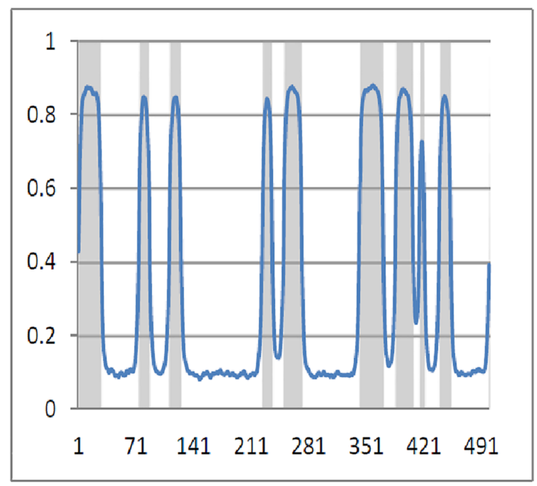
Table 1: Maximizing the normal log likelihood function when the error distribution is potentially non-normal: Monte Carlo experiment.

	True	Case #1	Case #2	Case #3	Case #4
$T = 500$					
β_1	−0.6	−0.609 (0.130)	−0.656 (0.555)	−0.983 (0.780)	−1.121 (0.892)
β_2	0.7	0.708 (0.088)	0.707 (0.083)	0.713 (0.117)	0.721 (0.127)
σ^2	1.1	1.087 (0.090)	1.074 (0.140)	0.967 (0.210)	0.907 (0.264)
P_{11}	0.9	0.892 (0.049)	0.892 (0.056)	0.788 (0.203)	0.754 (0.228)
P_{22}	0.95	0.942 (0.029)	0.946 (0.023)	0.928 (0.038)	0.920 (0.043)
$T = 5000$					
β_1	−0.6	−0.602 (0.039)	−0.610 (0.041)	−0.742 (0.246)	−0.959 (0.629)
β_2	0.7	0.701 (0.025)	0.696 (0.024)	0.730 (0.046)	0.735 (0.088)
σ^2	1.1	1.100 (0.027)	1.100 (0.049)	1.013 (0.107)	0.944 (0.176)
P_{11}	0.9	0.899 (0.012)	0.902 (0.012)	0.848 (0.098)	0.788 (0.169)
P_{22}	0.95	0.949 (0.007)	0.951 (0.006)	0.934 (0.020)	0.921 (0.033)
$T = 50000$					
β_1	−0.6	−0.602 (0.013)	−0.610 (0.025)	−0.694 (0.096)	−0.758 (0.169)
β_2	0.7	0.700 (0.009)	0.695 (0.024)	0.734 (0.035)	0.764 (0.065)
σ^2	1.1	1.100 (0.008)	1.098 (0.038)	1.028 (0.074)	0.966 (0.137)
P_{11}	0.9	0.900 (0.004)	0.901 (0.029)	0.868 (0.033)	0.835 (0.069)
P_{22}	0.95	0.950 (0.002)	0.951 (0.030)	0.936 (0.015)	0.920 (0.031)

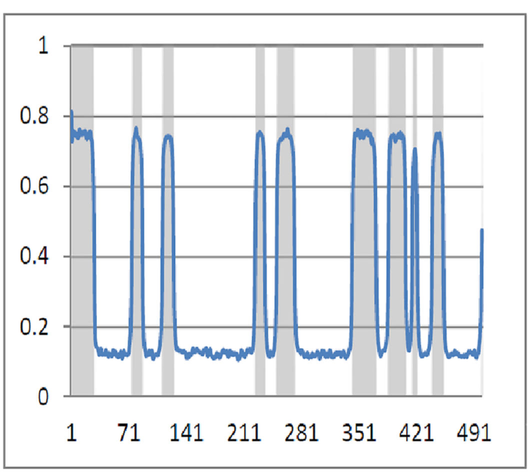
In order to investigate how the inferences on the regime probabilities are affected by the violation of the normality assumption, we conduct another simulation study. When generating data, we consider the same data generating processes as given above, except that we generate S_t, $t = 1, 2, \ldots, T$, only once and fix them in repeated sampling. The sample size we consider is $T = 500$. For each data set generated in this way, we estimate the model in equations (1) and (2) by maximizing a normal log-likelihood and then calculate the smoothed probabilities conditional on estimated parameters. Figure 1 plots the average smoothed probabilities of low-mean regime ($S_t = 1$) for each case. The shaded areas represent the true low-mean regime. Case #1 with the normally distributed error term provides us with the most accurate and sharpest regime inferences. However, as the distribution of the error term deviates from normality, the inferences about the regime probabilities deteriorate a lot, especially for Case #4 in which the degree of asymmetry in the error distribution is the highest.



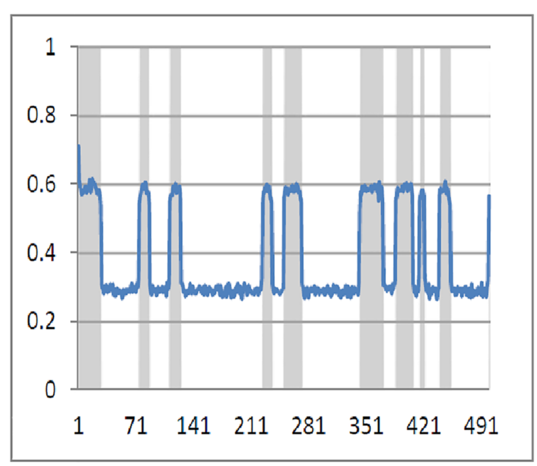
(a) Case #1 : Normal Distribution



(b) Case #2: t-Distribution



(c) Case #3: Log Chi-Square Distribution



(d) Case #4: Mixture of Normals

Figure 1: Smoothed probabilities of regime 1 based on Quasi-maximum likelihood estimation under different error distributions [T = 500; shaded area: True regime 1].

3 Basic Model and Two Identification Issues

Without loss of generality in dealing with the identification issues, let us consider the following basic model in which the unknown error distribution is approximated by the mixture of normals:

$$\begin{aligned} y_t &= \beta_{S_t} + \sigma\varepsilon_t^*, \quad S_t = 1,2,\ldots,K; \quad t = 1,2,\ldots.,T, \\ \varepsilon_t^*|D_t &\sim i.i.d.\ N\left(\mu_{D_t}^*, h_{D_t}^{*2}\right), \quad D_t = 1,2,\ldots,M, \end{aligned} \tag{3}$$

where S_t is a first order Markov-switching process with the following transition probabilities:

$$\Pr[S_t = j|S_{t-1} = i] = p_{S,ij}, \quad \sum_{j=1}^{K} p_{S,ij} = 1, \quad i,j = 1,2,\ldots,K, \tag{4}$$

and the mixture indicator variable D_t is serially independent.

3.1 Identification Issue #1: The Label Switching Problem

For our basic model, a typical way of labeling the states for S_t and D_t is given below:

$$\begin{aligned} \beta_{S_t} &= \sum_{k=1}^{K} \beta_k S_{k,t}, \\ \mu_{D_t}^* &= \sum_{m=1}^{M} \mu_m^* D_{m,t} \quad and \quad h_{D_t}^{*2} = \sum_{m=1}^{M} h_m^{*2} D_{m,t}, \end{aligned} \tag{5}$$

where

$$\begin{aligned} S_{k,t} &= \begin{cases} 1, & \text{if} \quad S_t = k; \quad k = 1,2,\ldots,K \\ 0, & \text{otherwise,} \end{cases} \\ D_{m,t} &= \begin{cases} 1, & \text{if} \quad D_t = m; \quad m = 1,2,\ldots,M \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{6}$$

The above labeling is not unique and the unconstrained parameter spaces for $\beta's$ and $\mu^*'s$ (or $h^{*2}{}'s$) contain $K!$ and $M!$ subspaces, respectively, each corresponding to different way to label states. As discussed in Stephens (2000) and Frühwirth-Schnatter (2001), when sampling from the unconstrained posterior via MCMC methods, it is impossible to know which component of the sampled parameter corresponds to which state due to potential label switching. Thus, as noted by Stephens (2000), summarizing joint posterior distributions by marginal distribution may lead to nonsensical answers due to the lack of identification.

The label switching problem is not an issue at all for the serially independent mixture indicator variable D_t, as we are not interested in the marginal distribution of $\mu's$ or $h^2{}'s$, or in the inferences on D_t. Furthermore, the complete data likelihood $f(y_1, \ldots, y_T|D_1, \ldots, D_T; .)$ and the prior for D_t is invariant to the relabeling of the states in D_t. However, it is critical that we take care of the label switching problem for S_t during the MCMC procedure, given that we want to obtain inferences on S_t and the regime-specific parameters based on their marginal posterior distributions.

The label switching problem for S_t can be solved by imposing the following ordering constraint on the regime-specific parameters:

$$\beta_1 < \beta_2 < \ldots < \beta_K. \tag{7}$$

A conventional way to incorporate the above constraint in the MCMC sampler is to employ a rejection method after drawing $\{\beta_1\beta_2 \ldots \beta_K\}$ jointly from the unconstrained joint posterior. However, note that the ordering constraint in equation (7) results in correlations among β_k, $k = 1,2,\ldots,K$. For example, the smaller the distances among the β parameters, the higher will be the correlations among them. Every time the β parameters are discarded and redrawn when the ordering constraint fails, we lose sample information about these correlations. This is why the rejection method may fail, especially when the distances among the β parameters are not large enough relative to the standard deviation of the error term.

For the permutation sampler proposed by Frühwirth-Schnatter (2001), the β parameters are first drawn from the unconstrained joint posterior. Then a suitable permutation of the labeling of the states is applied if the ordering constraint is violated. As the permutation is applied without discarding the β parameters drawn in this case, there is no loss of sample information unlike in the case of the rejection method. This is why the permutation sampler improves upon the rejection method, as illustrated by Frühwirth-Schnatter (2001). However, one potential drawback of the permutation sampler is that we need to set the marginal priors for β_k, $k = 1, 2, \ldots, K$, to be identical, but independent. This is because the prior densities for the individual β coefficients should be permutation-invariant. For this reason, it would be impossible to specify a joint prior that appropriately reflects the potential correlations among the β parameters.

As an alternative method for dealing with the label switching problem, we consider the following transformation of the β parameters in equation (7):

$$\beta_{S_t} = \beta_1 + \sum_{k=2}^{S_t} a_k, \;\; S_t = 2, 3, \ldots, K, \quad a_k > 0 \;\; for\ all\ k. \tag{8}$$

An advantage of the above specification is that we can indirectly specify the potential prior dependence among β_k, $k = 1, 2, \ldots, K$, by employing independent marginal priors for β_1 and a_k, $k = 2, 3, \ldots, K$. We first draw β_1 conditional on a_k, $k = 2, 3, \ldots, K$, and then, draw a_k for $k = 2, 3, \ldots, K$, from appropriate truncated marginal posteriors conditional on β_1 and $\tilde{a}_{\neq k} = \{a_2, \ldots, a_{k-1}; a_{k+1}, \ldots a_K\}$. We can then recover the $\beta_2, \ldots, \beta_K$ parameters based on equation (8). Here, an important issue to consider is that the likelihood function for a_k depends only on the observations for which $S_t = j$, $j = k, k+1, \ldots, K$, while the likelihood function for β_1 depends on all the observations in the sample. As in the case of the permutation sampler, there is no loss of sample information in the course of the proposed MCMC procedure. Unlike the case of the permutation sampler, however, the proposed prior and sampling procedure allow us to accommodate the non-sample information on the potential dependence among the β parameters that results from the inequality constraint in equation (7).[3] Besides, the implementation of the proposed sampler is much easier than the permutation sampler, especially for a model like the one in our empirical application section, in which the regime-specific means are subject to structural breaks within the Markov-switching framework.

3.2 Identification Issue #2: Disentangling the Markov-Switching Variable (S_t) from the Mixture Indicator Variable (D_t)

In this section, we consider the identification of the latent Markov-switching variable S_t from the latent and serially independent mixture indicator variable D_t in equation (3). For this purpose, we substitute equation (8) into equation (3) to obtain

$$y_t = \bar{\beta}'\bar{S}_t + \varepsilon_t, \;\; \varepsilon_t \sim (\beta_1, \sigma^2), \tag{9}$$

where $\bar{\beta} = [\bar{\beta}_2 \;\; \bar{\beta}_3 \;\; \ldots \;\; \bar{\beta}_K]'$, with $\bar{\beta}_k = \sum_{j=2}^{k} a_j$, $k = 2, 3, \ldots, K$; $\bar{S}_t = [S_{2,t} \;\; \ldots \;\; S_{K,t}]'$, with $S_{k,t}$, $k = 2, 3, \ldots, K$, being defined in equation (6); and $\varepsilon_t = \beta_1 + \sigma\varepsilon_t^*$. We can then approximate ε_t by the following mixture of normals:

$$\varepsilon_t | D_t \sim i.i.d. \;\; N\left(\mu_{D_t}, h_{D_t}^2\right), \quad D_t = 1, 2, \ldots, M. \tag{10}$$

---

3 When $K = 2$, for example, we have $\beta_2 = \beta_1 + a_2$, where we impose independent priors for β_1 and a_2 (e.g., $\beta_1 \sim N\left(b_1, \sigma_{\beta_1}^2\right)$ and $a_2 \sim N\left(\alpha_2, \sigma_{a_2}^2\right)_{1[a_2>0]}$, with 1[.] referring to the indicator function). In this case, it is easy to show that $\text{corr}(\beta_1, \beta_2) = \frac{\sigma_{\beta_1}^2}{\sqrt{\sigma_{\beta_1}^2\left(\sigma_{\beta_1}^2 + \sigma_{a_2}^2\right)}}$, which is a decreasing function of the prior mean of a_2. Note that, as a_2 has a truncated normal distribution, $\sigma_{a_2}^2$ is positively related to the prior mean α_2.

Note that the dynamics of S_t, given the transition probabilities in equation (4), can be represented by the following VAR process for $\bar{S}_t$:[4]

$$\bar{S}_t = Q_0 + Q_1\bar{S}_{t-1} + \bar{v}_t, \tag{11}$$

where the elements of the $(K-1)\times 1$ vector Q_0 and the $(K-1)\times(K-1)$ matrix Q_1 are functions of the transition probabilities. We define C as the collection of the eigenvectors of the Q_1 matrix. By pre-multiplying both sides of equation (11) by C^{-1}, and then, by rearranging the terms in the resulting equation and equation (9), we obtain:

$$y_t = \bar{\beta}^{*\prime}\bar{S}_t^* + \varepsilon_t,\ \varepsilon_t \sim \left(\beta_1, \sigma^2\right), \tag{12}$$

$$\bar{S}_t^* = \Lambda_0 + \Lambda_1\bar{S}_{t-1}^* + \bar{v}_t^*, \tag{13}$$

where $\bar{\beta}^* = C'\bar{\beta}$; $\bar{S}_t^* = C^{-1}S_t^*$; $\Lambda_0 = C^{-1}Q_0$; $\Lambda_1 = C^{-1}Q_1$; $C = \text{diag}\{\lambda_2, \lambda_3, \ldots, \lambda_K\}$, with λ_k, $k = 2, 3, \ldots, K$, referring to the eigenvalues of Q_1; and $\bar{v}_t^* = C^{-1}\bar{v}_t$.

Then, by noting that ε_t in equation (12) can be approximated by the mixture of normals as in equation (10) (i.e. $\varepsilon_t = \sum_{m=1}^{M}\mu_m D_{m,t} + \sqrt{\sum_{m=1}^{M}h_m^2 D_{m,t}}\varepsilon_t^*,\ \varepsilon_t^* \sim i.i.d.N(0,1)$), where $D_{m,t}$ as defined in equation (6) is serially independent and by noting that Λ_1 in equation (12) is diagonal, we can rewrite equations (12) and (13) as follows:

$$\begin{aligned}
y_t &= \sum_{k=2}^{K}\bar{\beta}_k^*\bar{S}_{k,t}^* + \sum_{m=1}^{M}\mu_m D_{m,t} + \sqrt{\sum_{m=1}^{M}h_m^2 D_{m,t}}\ \varepsilon_t^*,\ \varepsilon_t^* \sim i.i.d.N(0,1),\\
\bar{S}_{k,t}^* &= \Lambda_{k,0} + \lambda_k\bar{S}_{k,t-1}^* + \bar{v}_{k,t}^*,\quad k = 2, 3, \ldots, K,\\
D_{m,t} &= g_m + \eta_{m,t},\ m = 1, 2, \ldots, M,
\end{aligned} \tag{14}$$

where $\bar{\beta}_k^*$ is the $(k-1)$th row of $\bar{\beta}^*$; $\bar{S}_{k,t}^*$ refers to the $(k-1)$th row of $\bar{S}_t^*$; $g_m = \Pr[D_t = m]$; and $\bar{v}_{k,t}^*$, $k = 2, 3, \ldots, K$, and $\eta_{m,t}$, $m = 1, 2, \ldots, M$, are discrete and serially independent.

Equation (14) tells us that we have a K-state first-order Markov-switching process for S_t and a mixture of M normals for ε_t only when the following conditions hold:

$$\lambda_k \neq 0, \quad k = 2, 3, \ldots, K, \tag{15}$$

suggesting that all the eigenvalues of the Q_1 matrix in equation (11) should be non-zero.

---

4 Equation (6) and the transition probabilities in equation (4) allow us to represent the dynamics of the vector $\begin{bmatrix} S_{1,t} & S_{2,t} & \ldots & S_{K,t}\end{bmatrix}'$ in the following VAR form:

$$\begin{bmatrix} S_{1,t}\\ S_{2,t}\\ \vdots\\ S_{K,t}\end{bmatrix} = \begin{bmatrix} p_{S,11} & p_{S,21} & \cdots & p_{S,K1}\\ p_{S,12} & p_{S,22} & \cdots & p_{S,K2}\\ \vdots & \vdots & \ddots & \vdots\\ p_{S,1K} & p_{S,2K} & \cdots & p_{S,KK}\end{bmatrix}\begin{bmatrix} S_{1,t-1}\\ S_{2,t-1}\\ \vdots\\ S_{K,t-1}\end{bmatrix} + \begin{bmatrix} v_{1,t}\\ v_{2,t}\\ \vdots\\ v_{K,t}\end{bmatrix}$$

where $\begin{bmatrix} v_1 & v_2 & \ldots & v_K\end{bmatrix}'$ is a vector of martingale difference sequences. As $\sum_{j=1}^{K}p_{ij} = 1$ and $\sum_{j=1}^{K}S_{jt} = 1$, the first row in the above equation does not carry additional information beyond that contained in the rest of the rows. Thus, by imposing the constraint $S_{1,t-1} = 1 - \sum_{j=2}^{K}S_{j,t-1}$ on the second through $K-th$ rows of the above equation, we obtain the following dynamics for $\begin{bmatrix} S_{2,t} & S_{3,t} & \ldots & S_{K,t}\end{bmatrix}'$:

$$\begin{bmatrix} S_{2,t}\\ \vdots\\ S_{K,t}\end{bmatrix} = \begin{bmatrix} p_{S,12}\\ \vdots\\ p_{S,1K}\end{bmatrix} + \begin{bmatrix} (p_{S,22} - p_{S,12}) & \cdots & (p_{S,K2} - p_{S,12})\\ \vdots & \ddots & \vdots\\ (p_{S,2K} - p_{S,1K}) & \cdots & (p_{S,KK} - p_{S,1K})\end{bmatrix}\begin{bmatrix} S_{2,t-1}\\ \vdots\\ S_{K,t-1}\end{bmatrix} + \begin{bmatrix} v_{2,t}\\ \vdots\\ v_{K,t}\end{bmatrix},$$

$$\left(\ \bar{S}_t = Q_0 + Q_1\bar{S}_{t-1} + \bar{v}_t,\ \right)$$

It is easy to show that the model is not identified when equation (15) does not hold. For example, suppose that $\lambda_K = 0$ and $\lambda_k \neq 0$, for $k = 2, 3, \ldots, K - 1$. Then, equation (14) can be written as

$$y_t = \sum_{k=2}^{K-1} \bar{\beta}_k^* \bar{S}_{k,t}^* + \sum_{m=1}^{M+1} \mu_m D_{m,t} + \sqrt{\sum_{m=1}^{M+1} h_m^{*2} D_{m,t}} \quad \varepsilon_t^*, \quad \varepsilon_t^* \sim i.i.d.N(0,1),$$
$$\bar{S}_{k,t}^* = \Lambda_{k,0} + \lambda_k \bar{S}_{k,t-1}^* + \bar{v}_{k,t}^*, \quad k = 2, 3, \ldots, K-1,$$
$$D_{m,t} = g_m + \eta_{m,t}, \quad m = 1, 2, \ldots, M+1, \tag{14'}$$

where $\mu_{M+1}^* = \bar{\beta}_K^*$; $g_{M+1} = \Lambda_{K,0}$; $D_{M+1,t} = \bar{S}_{K,t}^*$; $h_{M+1}^* = 0$; and the transition probabilities for S_t and the mixture probabilities for D_t are redefined accordingly. Note that equation (14') describes a model with a $(K - 1)$-state Markov-switching process for S_t and a mixture of $(M + 1)$ normals for ε_t. The likelihood value for the model in Equation (14') is exactly the same as that for the model in equations (14), and this is a typical example of non-identification.

For economic data, a negative serial correlation in S_t does not seem to make a lot of sense. We thus impose the constraints that $\lambda_k > 0, k = 2, \ldots, K$, in order to achieve the identification. For this purpose, We set the prior mean of $p_{S,jj}$ to be larger than 0.5 for $j = 1, 2, \ldots, K$, as this is the sufficient condition for $\lambda_k > 0, k = 2, 3, \ldots, K$. Then, once the transition probabilities are drawn conditional on S_t, $t = 1, 2, \ldots, T$, we can construct the Q_1 matrix in equation (11) and calculate its eigenvalues λ_k, $k = 2, 3, \ldots, K$. Then, if the identifying constraints are not satisfied, we redraw S_t, $t = 1, 2, \ldots, T$, and the corresponding transition probabilities until the constraints are satisfied.

4 General Model Specification and the MCMC Procedure

4.1 Specification for a General Model

Consider the following generalized model:[5]

$$y_t = \beta_{S_t} + u_t, \quad S_t = 1, 2, \ldots, K,$$
$$\phi(L) u_t = \sigma_{W_t} \varepsilon_t^*, \quad \varepsilon_t^* \sim (0,1), \quad W_t = 1, 2, \ldots, N,$$
$$\beta_1 < \beta_2 < \ldots < \beta_K; \quad \sigma_1^2 < \sigma_2^2 < \ldots < \sigma_N^2, \tag{16}$$

where ε_t^* is independently distributed; $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ is a polynomial equation in the lag operator; all roots of $\phi(L) = 0$ lie outside the complex unit circle; the transitional dynamics of S_t is specified in equation (4). We assume that W_t is independent of S_t and follows an N-state, first-order Markov-switching process with the following transition probabilities:[6]

$$\Pr[W_t = j | W_{t-1} = i] = p_{W,ij}, \quad \sum_{j=1}^{N} p_{W,ij} = 1, \quad i, j = 1, 2, \ldots, N. \tag{17}$$

In order to avoid the non-identification resulting from the problem of label switching, we follow employ the following specifications for the β_{S_t} and $\sigma_{W_t}^2$ parameters:

5 The first equation in (16) can be further generalized to the following regression equation with a vector of covariates x_t:

$$y_t = \beta_{S_t} + \Gamma_{S_t}' x_t + u_t, \tag{16'}$$

where x_t is a vector of exogenous variables.

6 The independence assumption between W_t and S_t is for analytical/computational convenience. This assumption can be relaxed.

$$\beta_{S_t} = \beta_1 + \sum_{k=2}^{S_t} a_k, \quad a_k > 0 \text{ for all } k, \; S_t = 2,3,\ldots,K,$$
$$\sigma_{W_t}^2 = \sigma_1^2 \prod_{n=2}^{W_t} (1+b_n), \quad (1+b_n) > 1 \text{ for all } n, \quad W_t = 2,3,\ldots,N, \tag{18}$$

which allow us to employ independent priors for $\{\beta_1, a_k, k = 2,3,\ldots,K\}$ and for $\{\sigma_1^2, (1+b_n), n = 2,3,\ldots,N\}$. This way, we can also indirectly account for the dependence among β_k, $k = 1,2,\ldots,K$ and the dependence among σ_n^2, $n = 1,2,\ldots,N$ which result from the ordering constraints ($\beta_1 < \beta_2 < \ldots < \beta_K$ and $\sigma_1^2 < \sigma_2^2 < \ldots < \sigma_N^2$).

By substituting equation (18) into equation (16) and rearranging terms, we obtain:

$$y_t = \beta_1 + a_2 \sum_{k=2}^{K} S_{k,t} + a_3 \sum_{k=3}^{K} S_{k,t} + \cdots + a_{K-1} \sum_{k=K-1}^{K} S_{k,t} + a_K S_{K,t} + u_t,$$
$$\phi(L) u_t = g_{W_t} u_t^*, \quad u_t^* \sim (0, \sigma_1^2), \tag{19}$$

where

$$g_{W_t}^2 = \frac{\sigma_{W_t}^2}{\sigma_1^2} = \prod_{n=2}^{W_t} (1+b_n), \quad g_1^2 = 1, \quad W_t = 2,3,\ldots,N. \quad \left(g_{W_t}^2 = g_{W_{t-1}}^2 (1+b_W) \right) \tag{20}$$

Then, by defining $e_t = \beta_1 + u_t$, equation (19) can be rewritten as:[7]

Model with Transformed Parameters

$$\begin{aligned} y_t &= a_2 \sum_{k=2}^{K} S_{k,t} + a_3 \sum_{k=3}^{K} S_{k,t} + \cdots + a_{K-1} \sum_{k=K-1}^{K} S_{k,t} + a_K S_{K,t} + e_t, \\ \phi(L) e_t &= g_{W_t} \varepsilon_t, \\ \varepsilon_t | W_t &\sim \left(\frac{1}{g_{W_t}} \phi(1) \beta_1, \sigma_1^2\right), \end{aligned}$$

where $\varepsilon_t = \phi(1)\beta_1 / g_{W_t} + u_t^*$. The unknown distribution of $\varepsilon_t | W_t$ can be approximated by the following Dirichlet process mixture of normals:[8]

Dirichlet Process Mixture of Normals

$$\begin{aligned} &\varepsilon_t \,|W_t, D_t \sim i.i.d.N\left(\frac{1}{g_{W_t}} \mu_{D_t}, h_{D_t}^2\right), D_t = 1,2,3,\ldots \\ &\left(\mu_m, h_m^2\right) \sim G, \quad m = 1,2,3,\ldots \\ &G | G_0, \alpha \sim DP(\alpha, G_0), \\ &G_0 \equiv N\left(\lambda_0, \psi_0 h_m^2\right) IG\left(\frac{\nu_h}{2}, \frac{\delta_h}{2}\right), \end{aligned} \tag{22}$$

---

7 For a Markov-switching model with covariates, the first equation in (21) can be extended to:

$$y_t = a_2 \sum_{k=2}^{K} S_{k,t} + a_3 \sum_{k=3}^{K} S_{k,t} + \cdots + a_{K-1} \sum_{k=K-1}^{K} S_{k,t} + a_K S_{K,t} + \sum_{k=1}^{K} S_{k,t} \Gamma_k' x_t + e_t. \tag{21'}$$

8 Note that σ_1^2 and β_1 can be easily recovered from

$$\beta_1 = \frac{1}{\phi(1)} \sum_{m=1}^{M} \mu_m p_{D,m}; \quad \sigma_1^2 = \left(\widetilde{\frac{1}{g^2}}\right) \sum_{m=1}^{M} (\mu_m - \bar{\mu})^2 p_{D,m} + \bar{h}^2,$$

where $\left(\widetilde{\frac{1}{g^2}}\right) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{g_{W_t}^2}$, $\bar{\mu} = \sum_{m=1}^{M} \mu_m p_{D,m}$, and $\bar{h}^2 = \sum_{m=1}^{M} \sigma_m^2 p_{D,m}$.

where $DP(.,.)$ refers to the Dirichlet process;[9] G_0 and α are referred to as the base distribution and the concentration parameter, respectively.

The base distribution G_0 is like the mean of the Dirichlet Process. In other words, the Dirichlet Process draws distributions around the base distribution the way a normal distribution draws real numbers around its mean. The concentration parameter α is like an inverse-variance of the Dirichlet Process. It describes the concentration of mass around the base distribution. In a Dirichlet Process mixture model, we can show that the probability of assigning an observation to a newly drawn distribution around the base distribution is $\frac{\alpha}{T-1+\alpha}$. Therefore, the larger the α is, the higher the probability of assigning an observation to a new distribution, and thus prior mean of the number of mixture is higher.

We employ a Normal-Inverse Gamma distribution as the base distribution. This means that we use $N\left(\lambda_0, \psi_0 h_m^2\right)$ as the prior distribution for mixture mean μ_m, and we use $IG(\frac{\nu_h}{2}, \frac{\delta_h}{2})$ as the prior distribution for mixture variance h_m^2. These are the conjugate priors for the Dirichlet Process Mixture model. In the case of finite mixture, the joint distribution of $\left(\mu_m, h_m^2\right)$ is given by G_0, and thus, $G = G_0$. The α parameter can be either fixed or random. In case the α parameter is random, it is common to employ a Gamma prior. Note that, conditional on g_{W_t}, μ_{D_t}, and $h_{D_t}^2$, the first line in equation (22) implies

$$\varepsilon_t = \frac{1}{g_{W_t}}\mu_{D_t} + h_{D_t}v_t, \quad v_t \sim i.i.d.N(0,1). \tag{23}$$

To complete the model, we employ the following priors for the parameters except those associated with the mixture of normals:

Other Priors

$$\begin{aligned}
&\tilde{\phi} = \begin{bmatrix}\phi_1 & \phi_2 & \dots & \phi_p\end{bmatrix}' \sim N(A_{\tilde{\phi}}, \Sigma_{\tilde{\phi}})_{1[S_{\tilde{\phi}}]},\\
&\tilde{a} = \begin{bmatrix}a_2 & a_3 & \dots & a_K\end{bmatrix}' \sim N(A, \Sigma_{\tilde{a}})_{1[a_2>0,\dots,a_K>0]},\\
&(1+b_n) \sim IG\left(\frac{\nu_{n,0}}{2}, \frac{\delta_{n,0}}{2}\right)_{1[(1+b_n)>1]}, \quad n = 2,3,\dots,N,\\
&\tilde{p}_{S,k} = \begin{bmatrix}p_{S,k1} & p_{S,k2} & \dots & p_{S,kK}\end{bmatrix}' \sim Dir(\alpha_{S,k1}, \alpha_{S,k2}, \dots, \alpha_{S,kK})_{1[p_{S,kk}>0.5]}, \quad k = 1,2,\cdot,K,\\
&\tilde{p}_{W,m} = \begin{bmatrix}p_{W,m1} & p_{W,m2} & \dots & p_{W,mM}\end{bmatrix}' \sim Dir(\alpha_{W,m1}, \alpha_{W,m2}, \dots, \alpha_{W,mM})_{1[p_{W,mm}>0.5]},\\
&m = 1,2,\cdot,M,
\end{aligned} \tag{24}$$

where $\Sigma_{\tilde{a}}$ is diagonal; $1[.]$ is the indicator function; $S_{\tilde{\phi}}$ refers to the stationary region of $\tilde{\phi}$; $IG(.)$ refers to the inverted Gamma distribution; and $Dir(.)$ refers to the Dirichlet distribution. Following Section 3.2, we impose the constraints $p_{S,kk} > 0.5, k = 1,2,\dots,K$, and $p_{W,nn} > 0.5, n = 1,2,\dots,N$, in order to identify the Markov-switching processes S_t and W_t against the mixture indicator variable D_t.

4.2 MCMC Procedure

In this section, we present an MCMC procedure for estimating the model that consists of equations (21)–(24).

4.2.1 Drawing Variates Associated with Markov-Switching Regression Equation Conditional on the Mixture of Normals and Data

By multiplying both sides of the first equation in (21) by $\phi(L)$ and then by substituting equation (23) in the resulting equation, we obtain

9 A Dirichlet process is a probability distribution whose range is itself a set of probability distributions.

$$\phi(L)y_t = \phi(L)\left(a_2\sum_{k=2}^{K}S_{k,t} + a_3\sum_{k=3}^{K}S_{k,t} + \cdots + a_{K-1}\sum_{k=K-1}^{K}S_{k,t} + a_K S_{K,t}\right) + \mu_{D_t} + g_{W_t}h_{D_t}\upsilon_t,$$
$$\upsilon_t \sim i.i.d.N(0,1), \tag{25}$$

which can be used to draw $\tilde{\phi}$, $\tilde{a}$, $\tilde{g}^2\left(=\left[g_2^2 \quad g_3^2 \quad \ldots \quad g_N^2\right]'\right)$, $\tilde{P}_S\left(=\left[\tilde{p}'_{S,1} \quad \ldots \quad \tilde{p}'_{S,K}\right]'\right)$, $\tilde{P}_W\left(=\left[\tilde{p}'_{W,1} \quad \ldots \quad \tilde{p}'_{W,M}\right]'\right)$, $\tilde{S}_T\left(=\left[S_1 \quad S_2 \quad \ldots \quad S_T\right]'\right)$, and $\tilde{W}_T\left(=\left[W_1 \quad W_2 \quad \ldots \quad W_T\right]'\right)$ conditional on $\tilde{\mu}\left(=\left[\mu_1 \quad \ldots \quad \mu_M\right]'\right)$, $\tilde{h}^2\left(=\left[h_1^2 \quad \ldots \quad h_M^2\right]'\right)$, $\tilde{D}_T\left(=\left[D_1 \quad \ldots \quad D_T\right]'\right)$, and data $\tilde{Y}\left(=\left[y_1 \quad \ldots \quad y_T\right]'\right)$.

4.2.1.1 Drawing $\tilde{a}$ Conditional on $\tilde{\phi}$, $\tilde{g}^2$, $\tilde{S}_T$, $\tilde{W}_T$, $\tilde{\mu}$, $\tilde{h}^2$, $\tilde{D}_T$ and Data

By rearranging equation (25), we obtain

$$y_{1t} = a_2 z_{2t}^{\dagger} + a_3 z_{3t}^{\dagger} + \cdots + a_K z_{Kt}^{\dagger} + \upsilon_t, \quad \upsilon_t \sim i.i.d.N(0,1), \tag{26}$$

where $y_{1t} = \frac{\phi(L)y_t - \mu_{D_t}}{g_{W_t}h_{D_t}}$ and $z_{jt}^{\dagger} = \frac{\sum_{k=j}^{K}\phi(L)S_{k,t}}{g_{W_t}h_{D_t}}$, $j = 2, 3, \ldots, K$. Then, for given y_{1t} and $z_{jt}^{\dagger}$, $t = p+1, 2, \ldots, T$, $j = 2, 3, \ldots, K$, we can generate $a_2, a_3, \ldots, a_K$ directly from the following truncated normal distributions, without resorting to the rejection sampling:

(1) Draw a_2 from

$$a_2 \mid a_3, a_4, \ldots, a_K \sim N\left(c_{a,2}, \omega_{a,2}^2\right)_{1[a_2>0]}$$

(2) Draw a_3 from

$$a_3 \mid a_2, a_4, \ldots, a_K \sim N\left(c_{a,3}, \omega_{a,3}^2\right)_{1[a_3>0]}$$

⋮

(K-1) Draw a_K from

$$a_K \mid a_2, a_3, \ldots, a_{K-1} \sim N\left(c_{a,K}, \omega_{a,K}^2\right)_{1[a_K>0]},$$

where $c_{a,j}$ and $\omega_{a,j}^2$ refer to the posterior mean and posterior variance of the truncated full conditional distribution of a_j, $j = 2, 3, \ldots, K$. Here, as discussed in Section 3.1, a_k should be drawn based on the observations for which St $= j$, j $= k$, k + 1, …,K.

4.2.1.2 Drawing $\tilde{\phi}$ Conditional on $\tilde{a}$, $\tilde{g}^2$, $\tilde{S}_T$, $\tilde{W}_T$, $\tilde{\mu}$, $\tilde{h}^2$, $\tilde{D}_T$, and Data

By rearranging equation (25), we obtain

$$y_{2t} = z_t^{*\prime}\tilde{\phi} + \upsilon_t, \quad \upsilon_t \sim i.i.d.N(0,1), \tag{27}$$

where $y_{2t} = \frac{y_t - z_t'\tilde{a} - \mu_{D_t}}{g_{W_t}h_{D_t}}$ and $z_t^* = \left[\frac{y_{t-1} - z_{t-1}'\tilde{a}}{g_{W_t}h_{D_t}} \quad \frac{y_{t-2} - z_{t-2}'\tilde{a}}{g_{W_t}h_{D_t}} \quad \ldots \quad \frac{y_{t-p} - z_{t-p}'\tilde{a}}{g_{W_t}h_{D_t}}\right]'$. Then, based on equation (27), we can draw $\tilde{\phi}$ from an appropriate posterior distribution.

4.2.1.3 Drawing $\tilde{g}^2$ Conditional on $\tilde{a}$ $\tilde{\phi}$, $\tilde{S}_T$, $\tilde{W}_T$, $\tilde{\mu}$, $\tilde{h}^2$, $\tilde{D}_T$, and Data

By defining $\zeta_t = g_{W_t}\upsilon_t$ in equation (25), we can calculate ζ_t by

$$\zeta_t = \frac{\phi(L)\left(y_t - z_t'\tilde{a}\right) - \mu_{D_t}}{h_{D_t}}, \tag{28}$$

where $z_t'\tilde{a} = a_2\sum_{k=2}^{K}S_{k,t} + a_3\sum_{k=3}^{K}S_{k,t} + \cdots + a_{K-1}\sum_{k=K-1}^{K}S_{k,t} + a_K S_{K,t}$. Note that equation (20) implies

$$\zeta_t | W_t = n \sim N\left(0, g_n^2\right) \equiv g_{n-1}N(0, (1+b_n))_{1[(1+b_n)>1]}. \tag{29}$$

We want to draw b_n conditional on $g^2_{n-1}, (1+b_{n+1}), \ldots, (1+b_N)$ for $n = 2, 3, \ldots, N$, and then we can obtain g^2_n, $n = 2, 3, \ldots, N$, based on equation (20). It should be noted that the likelihood function for $(1 + b_n)$ depends on the values of ζ_t for which $W_t = n, n + 1, \ldots, N$, as $(1 + b_n)$ is a common element only in $g^2_{W_t}$, $W_t = n, n + 1, \ldots, N$. Thus, if we define

$$\zeta^*_{n,t} = \frac{\zeta_t}{g_{n-1}\sqrt{\prod_{i=n+1}^{N}(1+b_i W_{i,t})}}, \tag{30}$$

where $W_{it} = 1$, if $W_t = i$, and 0, otherwise, we have the following result:

$$\zeta^*_{n,t} \mid g_{n-1}, (1+b_{n+1}), \ldots, (1+b_N) \sim N(0, (1+b_n))_{1[(1+b_n)>1]}, \tag{31}$$

for $T_n = \{t: W_t = n, n + 1, \ldots, N\}$. Then, given the prior for $(1 + b_n)$ in equation (24), we can draw $(1 + b_n)$ from the following truncated inverse Gamma distribution:

$$(1+b_n) \mid \ g^2_{n-1}, (1+b_{n+1}), \ldots, (1+b_N), \tilde{u}_T \sim IG\left(\frac{\nu_{n,1}}{2}, \frac{\delta_{n,1}}{2}\right)_{1[(1+b_n)>1]}, \tag{32}$$

where $\delta_{n,1} = \delta_{n,0} + \sum_{t\in T_n} \zeta^{*2}_{n,t}$ and $\nu_{n,1} = \nu_{n,0} + c_n$, with c_n referring to the cardinality of T_n. When drawing b_n from equation (32), we draw b_n directly from the truncated Inverse Gamma distribution. Once we draw $(1 + b_n)$, $n = 2, 3, \ldots, N$, we can obtain $\tilde{g}^2$ based on equation (20).

4.2.1.4 Drawing $\tilde{S}_T$, $\tilde{p}_S$, $\tilde{W}_T$, and $\tilde{p}_W$ Conditional on $\tilde{a}$, $\tilde{\phi}$, $\tilde{g}^2$, $\tilde{\mu}$, $\tilde{h}^2$, $\tilde{D}_T$, and Data

For this step, we can rewrite equation (25) in the following way:

$$\phi(L)\left(y_t - \beta^*_{S_t}\right) = \mu_{D_t} + g_{W_t} h_{D_t} v_t, \quad v_t \sim i.i.d.N(0,1), \tag{33}$$

where $\beta^*_{S_t} = \sum_{j=2}^{S_t} a_j$ with $\beta^*_1 = 0$.

When drawing $\tilde{S}_T$ conditional on all the other variates, equation (33) serves as a usual model with a Markov-switching latent variable S_t, while D_t and W_t serve as dummy variables. Furthermore, drawing $\tilde{W}_T$ conditional on all the other variates, equation (33) serves as a usual model with a Markov-switching latent variable W_t, while D_t and S_t serve as dummy variables. Thus, drawing $\tilde{S}_T$ and $\tilde{W}_T$ is a standard procedure. Once $\tilde{S}_T$ and $\tilde{W}_T$ are drawn, we can draw $\tilde{p}_S$ and $\tilde{p}_W$ conditional on $\tilde{S}_T$ and $\tilde{W}_T$, respectively.

4.2.2 Drawing Variates Associated with the Mixture of Normals Conditional on $\tilde{\varepsilon}_T \left(= [\varepsilon_1 \quad \varepsilon_2 \quad \ldots \quad \varepsilon_T]'\right)$

Conditional on $\tilde{a}$, $\tilde{g}^2$, $\tilde{\phi}$, $\tilde{\sigma}^2$, $\tilde{S}_T$, $\tilde{W}_T$ and data, we can calculate the error term ε_t from the first two equations in in (21) as follows:

$$\varepsilon_t = \phi(L)\left(y_t - a_2\sum_{k=2}^{K} S_{k,t} + a_3\sum_{k=3}^{K} S_{k,t} + \cdots + a_{K-1}\sum_{k=K-1}^{K} S_{k,t} + a_K S_{K,t}\right)\frac{1}{g_{W_t}}. \tag{34}$$

Then, based on equations (22) and (23), we can draw the variates associated with the mixture of normals (i.e. $\tilde{\mu}$, $\tilde{h}^2$, $\tilde{D}_T$ and α) conditional on $\tilde{\varepsilon}_T$. As discussed in Section 3.1, we are not interested in the marginal distribution of $\mu's$ or $h^{2\prime}s$ or in the inferences on D_t. Thus, the label switching problem for D_t is not an issue here. We therefore draw $\tilde{\mu}$ or $\tilde{h}^2$ without any identifiability constraints. We proceed with the following procedures in drawing $\tilde{\mu}$, $\tilde{h}^2$, $\tilde{D}_T$ and α:

(i) Draw $\tilde{\mu}$ based on equation (23), conditional on $\tilde{g}$, $\tilde{h}^2$, $\tilde{D}_T$, and $\tilde{\varepsilon}_T$.
(ii) Draw $\tilde{h}^2$ based on equation (23), conditional on $\tilde{\mu}$, $\tilde{g}^2$, $\tilde{D}_T$, and $\tilde{\varepsilon}_T$.
(iii) Draw $\tilde{D}_T$ and α for the Dirichlet process mixture of normals specified in equation (22), conditional on $\tilde{\mu}$, $\tilde{h}^2$, $\tilde{g}^2$, and $\tilde{\varepsilon}_T$, The total number of mixtures (M^*) realized at a particular MCMC iteration is obtained as a byproduct of drawing $\tilde{D}_T$.

Drawing $\tilde{\mu}$ and $\tilde{h}^2$ from their full conditional distributions derived from equation (23) is standard. The procedure for drawing $\tilde{D}_T$ and α is based on West, Muller, and Escobar (1994), Escobar and West (1995), and Neal (2000).

4.3 Simulation Study

In this section, we perform simulation studies in order to show that the proposed model-identification schemes and the proposed algorithm work properly. For this purpose, we first generate 100 sets of samples based on the following data generating process, which is the same as Case #4 of Section 2 (with $K = 2, N = 1$, and $\phi(L) = 1$ for the model presented in Section 4.1):

4.3.1 Data Generating Process #1

$$
\begin{aligned}
& y_t = \beta_{S_t} + \sigma\varepsilon_t^*, \quad \varepsilon_t^* \sim (0,1), \quad S_t = 1,2; \; t = 1,2,\ldots,T, \\
& (\; y_t = \beta_1 + a_2 S_{2,t} + \sigma\varepsilon_t^*, \quad a_2 = \beta_2 - \beta_1 > 0, \;) \\
& \varepsilon_t^* \; |D_t \sim i.i.d. \;\; N\left(\mu_{D_t}^*, h_{D_t}^{*2}\right), \; D_t = 1,2,3, \\
& T = 500; \; \beta_1 = -0.6, \; \beta_2 = 0.7; \quad \sigma^2 = 1.1; \quad p_{S,11} = 0.9, \; p_{S,22} = 0.95,
\end{aligned}
$$

where $S_{2,t} = 1$ if $S_t = 2$ and $S_{2,t} = 0$, otherwise; S_t and D_t are independent of each other and $p_{ij} = \Pr[S_t = j|S_{t-1} = i]$. The parameter values associated with the mixture of normals for ε_t^* are also the same as those for Case #4 in Section 2.

Based on the discussions on the identification issues in Section 3, we consider the following representation of the model for estimation:

$$
y_t = a_2 S_{2,t} + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.\left(\beta_1, \sigma^2\right), \quad S_{2,t} = 0,1, \quad a_2 > 0,
$$

where we approximate the distribution of ε_t by the Dirichlet process mixture of normals in equation (22). The priors we employ are:[10]

$$
G_0 \;\equiv\; N\left(-0.6, 5h_m^2\right) IG\left(\frac{130}{2}, \frac{30}{2}\right); \quad \alpha \sim Gamma(10,3); \quad a_2 \sim N(1.3, 0.45)_{1[a_2>0]}
$$

$$
\begin{bmatrix} p_{S,11} & p_{S,12} \end{bmatrix}' \sim Dirichlet(9,1)_{1[\,p_{S,11}>0.5]}; \;\; \begin{bmatrix} p_{S,21} & p_{S,22} \end{bmatrix}' \sim Dirichlet(0.5, 9.5)_{1[\,p_{S,22}>0.5]}.
$$

When we estimate the model under a normality assumption for the error term, we employ the following priors for β_1 and σ^2:

$$
\beta_1 \sim N(-0.6, 0.45); \quad \sigma^2 \sim IG(3.4, 2.7),
$$

which are the same as the unconditional distributions for β_1 and σ^2 implied by our specification of the based distribution G_0 for the Dirichlet process mixture of normals.

To show that the proposed identification schemes also work properly for a model with Markov-switching variances, we additionally generate 100 sets of samples based on the following data-generating process:

4.3.2 Data Generating Process #2

$$
\begin{aligned}
& y_t = \beta + \sigma_{W_t}\varepsilon_t^*, \quad \varepsilon_t^* \sim (0,1), \quad W_t = 1,2; \; t = 1,2,\ldots,T, \\
& \varepsilon_t^* \; |D_t \sim i.i.d. \;\; N\left(\mu_{D_t}^*, h_{D_t}^{*2}\right), \; D_t = 1,2,3, \\
& T = 500; \; \beta = 1, \; \sigma_1^2 = 0.5; \quad \sigma_2^2 = 2; \quad p_{W,11} = 0.9, \; p_{W,22} = 0.95,
\end{aligned}
$$

where the parameter values associated with the mixture of normals for ε_t^* are the same as those for Case #4 in Section 2.

10 The specified prior distribution of the concentration parameter α implies that the prior mean for the number of mixtures is 3.32 when sample size equals 500.

Based on the discussions in Section 3, we consider the following representation of the model for estimation:

$$y_t = g_{W_t}\varepsilon_t, \quad \varepsilon_t \sim i.i.d.\left(\beta_1/g_{W_t}, \sigma_1^2\right), \quad g_1 = 1, \quad g_2 = \sigma_2^2/\sigma_1^2$$

where we approximate the distribution of ε_t by the Dirichlet process mixture of normals in equation (22). The priors we employ are:

$$G_0 \equiv N\left(1, 5h_m^2\right)IG\left(\frac{130}{2}, \frac{15}{2}\right); \quad \alpha \sim Gamma(10,3); \quad g_2 \sim IG(11,40)_{1[g_2>1]}$$

$$\left[p_{W,11} \quad p_{W,12}\right]' \sim Dirichlet(9,1)_{1[p_{W,11}>0.5]}; \; \left[p_{W,21} \quad p_{W,22}\right]' \sim Dirichlet(0.5, 9.5)_{1[p_{W,22}>0.5]}.$$

When we estimate the model under a normality assumption for the error term, we employ the following priors for β_1 and σ^2:

$$\beta \sim N(1, 0.45); \quad \sigma_1^2 \sim IG(3,1),$$

which are the same as the unconditional distributions for β_1 and σ^2 implied by our specification of the based distribution G_0 for the Dirichlet process mixture of normals.

For both data-generating processes, we obtain the posterior mean of each parameter conditional on each of the 100 generated samples. We then calculate the mean and the standard deviation of 100 posterior means for each parameter obtained from these 100 samples. This is equivalent to investigating the sampling moments of the posterior mean for each parameter.

The third column of Table 2 reports the sample mean and standard deviation of the posterior means when the distribution of the error term is erroneously assumed to be normal. For data generating process #1, the results reported in the upper panel of Table 2 are almost the same as those based on the maximum likelihood approach as shown in the 6th column of Table 1 for $T = 500$. We have large biases in the parameter estimates. However, the fourth column of Table 2 shows that, when the non-normality of the error distribution is appropriately taken care of as outlined in Section 4.2, these biases almost disappear. In summary, we find strong simulation evidence that the Markov switching component of the conditional mean is well identified from the mixture of normals specification of the error innovation.

We reach at the same conclusion for data generating #2. However, the evidence seems to be less compelling for the identification of the Markov switching component of the volatility process, as the results reported in the lower panel of Table 2 suggest.

Table 2: Performance of the proposed algorithm [simulation Studies].

Parameter	**True value**	**Average of posterior mean (RMSE)**	
		Normality assumption	**Mixture of normals**
		Data generating process #1	
β_1	−0.6	−1.113 (0.856)	−0.601 (0.088)
β_2	0.7	0.682 (0.132)	0.703 (0.069)
σ^2	1.1	0.923 (0.231)	1.156 (0.168)
$P_{S,11}$	0.9	0.774 (0.172)	0.897 (0.032)
$P_{S,22}$	0.95	0.924 (0.034)	0.944 (0.018)
		Data generating process #2	
β	1	1.154 (0.216)	0.993 (0.094)
σ_1^2	0.5	0.635 (0.472)	0.519 (0.322)
σ_2^2	2	2.719 (1.973)	2.223 (0.531)
$P_{W,11}$	0.9	0.834 (0.135)	0.863 (0.104)
$P_{W,22}$	0.95	0.801 (0.220)	0.890 (0.121)

5 An Application to the Growth of Postwar U.S. Industrial Production Index [1947:M1–2019:M9]

5.1 Specification for an Empirical Model

We consider the following univariate Markov-switching model for the growth of industrial production index (Δy_t), with a two-state Markov-switching mean ($S_t = 1, 2$) and a three-state Markov-switching variance ($W_t = 1, 2, 3$):[11]

$$\begin{aligned} &\Delta y_t = \beta_{1,C_t} + a_{2,C_t} S_{2,t} + u_t, \quad C_t = 1, 2, 3, \\ &a_{2,C_t} > 0, \quad \forall\ t, \\ &u_t = \phi u_{t-1} + g_{W_t} u_t^*, \quad u_t^* \sim i.i.d.\left(0, \sigma_1^2\right), \quad |\phi| < 1, \end{aligned} \tag{35}$$

where u_t is independently distributed; $S_{2,t} = 1$ if $S_t = 2$ and $S_{2,t} = 0$, otherwise; β_{1,C_t} is the mean growth rate during recession and $\beta_{1,C_t} + a_{2,C_t}$ is the mean growth rate during boom; $g_{W_t}^2$ is specified in equation (20) with $N = 3$; S_t and W_t are independent. The transitional dynamics of S_t and W_t are specified in equations (4) and (17) with $K = 2$ and $N = 3$. In the above model, we introduce a latent discrete variable C_t to allow for structural breaks in the mean growth rates for boom recession.

Kim and Nelson (1999) show empirical evidence of a narrowing gap between growth rates of real GDP during recessions and booms. They argue that this narrowing gap is as important as the reduction in the volatility of the shocks as a feature of the Great Moderation. More recently, by specifying the regime-specific mean growth rates of real GDP to follow random walks, Eo and Kim (2016) also show that the mean growth rate during the boom has been steadily decreasing along with the long-run mean growth rate since 1947. To incorporate these particular features of the business cycle discussed in Kim and Nelson (1999) and Eo and Kim (2016), we incorporate two structural breaks with unknown break points in the mean growth rates for boom and recession. We specify β_{1,C_t} and a_{2,C_t} in the following way:[12]

$$\begin{aligned} &\beta_{1,C_t} = \gamma_1 + \gamma_2 C_{2,t} + (\gamma_2 + \gamma_3) C_{3,t}, \\ &a_{2,C_t} = (\eta_1 + \eta_2 + \eta_3) C_{1,t} + (\eta_2 + \eta_3) C_{2,t} + \eta_3 C_{3,t}, \\ &\gamma_2 > 0,\ \gamma_3 > 0; \quad \eta_1 > 0,\ \eta_2 > 0,\ \eta_3 > 0, \end{aligned} \tag{36}$$

where

$$C_{k,t} = \begin{cases} 1, & \text{if} \quad C_t = k; \quad k = 1, 2, 3 \\ 0, & \text{otherwise,} \end{cases} \tag{37}$$

and C_t follows a three-state Markov-switching process with absorbing states, as specified below:

11 We allow for a 3-state Markov-switching process for the variance of the shocks to capture the unusually high volatility during the Financial Crisis period.

12 Incorporating structural breaks in the mean growth rates for boom or recession such that their gap narrows is based on the prior belief that the Great Moderation is not over with the onset of the Financial Crisis. In his recent study on whether the Great Moderation is over, Clark (2009) concludes that, over time, macroeconomic volatility will likely undergo occasional shifts between high and low levels with low volatility being the norm, suggesting that the Great Moderation is not over. Gadea-Rivas, Gomez-Lscos, and Perez-Quiros (2014) also provide empirical evidence suggesting that output volatility remains subdued despite the turmoil created by the Financial Crisis of 2008.

$$p_{C,11} > 0.5;\ p_{C,12} = 1 - p_{C,11};\ p_{C,13} = 0,\ p_{C,21} = 0,\quad p_{C,22} > 0.5;\ p_{C,23} = 1 - p_{C,22};$$
$$p_{C,31} = 0,\ p_{C,32} = 0,\ p_{C,33} = 1, \tag{38}$$

where $p_{C,ij} = \Pr[C_t = j|C_{t-1} = i]$.

Note that the existence of the absorbing states in C_t allows us to identify C_t from the Markov-switching process S_t in our model. The ordering constraints in the last line of equation (36) guarantee a narrowing gap between mean growth rates for booms and recessions. At the same time, they also guarantee that $a_{2,C_t} > 0,\ \forall\ t$, thereby allowing us to identify regime 2 (i.e. $S_t = 2$) as a boom. A graphical illustration of the implied priors for the mean growth rates during recessions and booms is provided in Figure 2.

By substituting equation (36) into equation (35), we obtain

$$\Delta y_t = \gamma_1 + \gamma_2 C_{2,t} + (\gamma_2 + \gamma_3) C_{3,t}$$
$$+ ((\eta_1 + \eta_2 + \eta_3) C_{1,t} + (\eta_2 + \eta_3) C_{2,t} + \eta_3 C_{3,t}) S_{2,t} + u_t. \tag{39}$$

Then, by defining $e_t = \gamma_1 + u_t$ and rearranging the terms in equation (39), we obtain:
Empirical Model with Transformed Parameters

$$\Delta y_t = \gamma_2 \sum_{j=2}^{3} C_{j,t} + \gamma_3 C_{3,t} + \left(\eta_1 C_{1,t} + \eta_2 \sum_{j=1}^{2} C_{j,t} + \eta_3 \sum_{j=1}^{3} C_{j,t} \right) S_{2,t} + e_t,$$
$$e_t = \phi e_{t-1} + g_{W_t} \varepsilon_t,\quad |\phi| < 1,\quad \varepsilon_t | W_t \sim i.i.d.\left(\frac{1}{g_{W_t}} (1-\phi)\gamma_1, \sigma_1^2 \right),$$
$$g_n^2 = g_{n-1}^2 (1 + b_n),\quad g_1^2 = 1,\ (1 + b_n) > 1,\ n = 2, 3, \tag{40}$$

where $\varepsilon_t = (1-\phi)\gamma_1 / g_{W_t} + u_t^*$, with $u_t^* \sim i.i.d.N(0, \sigma_1^2)$. The unknown distribution of the error term ε_t conditional on W_t can be approximated by the Dirichlet Process mixture of normals specified in equation (22). Given the truncated normal priors, each of the γ and η parameters can be sequentially drawn from appropriate truncated normal distributions as explained in Section 3.1, without resorting to the rejection sampling.

Lastly, note that structural breaks in the mean growth rates for boom and recession imply structural breaks in the long-run mean growth rate. Based on equation (39), the long-run mean growth rate (τ_t) at each iteration

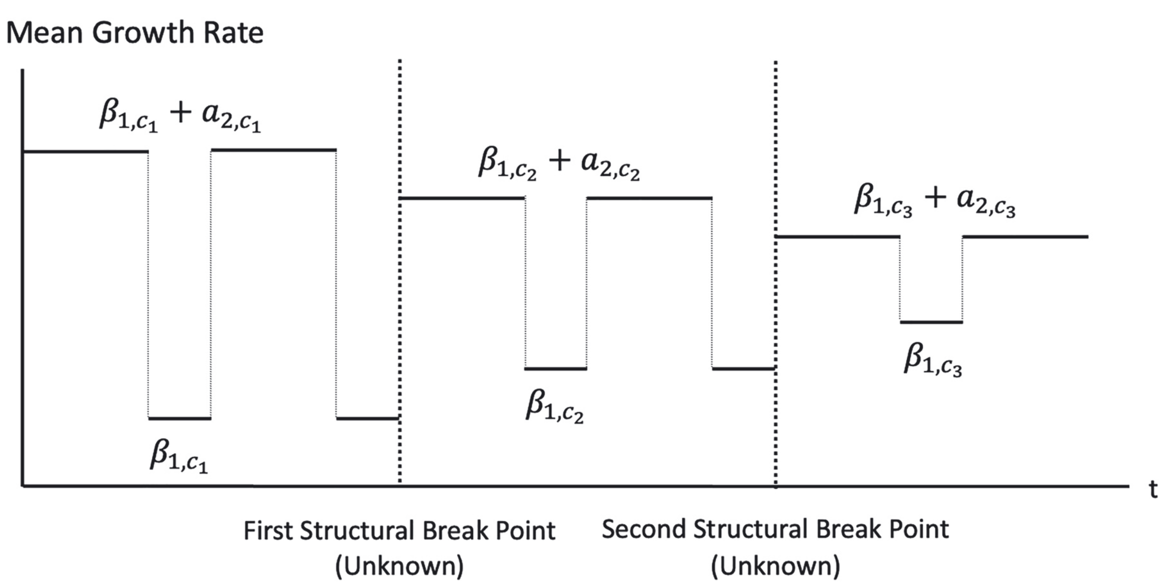


Figure 2: Graphical illustration of the priors for the narrowing gap between mean growth rates during boom and recession.

of the MCMC can be obtained by:

$$\tau_t = \gamma_1 + \gamma_2 \Pr[C_t = 2|I_T] + (\gamma_2 + \gamma_3) \Pr[C_t = 3|I_T] \\ + ((\eta_1 + \eta_2 + \eta_3) \Pr[C_t = 1|I_T] + (\eta_2 + \eta_3) \Pr[C_t = 2|I_T] + \eta_3 \Pr[C_t = 3|I_T]) \Pr[S_t = 2], \quad (41)$$

where γ_1 can be recovered in the same way as the β_1 coefficient is recovered in footnote 7, with M in footnote 7 now referring to the realized number of mixtures at a particular iteration of the MCMC; I_T refers to information up to T; and $\Pr[S_t = 2]$ refers to the steady-state probability that $S_t = 2$, which is given by $\Pr[S_t = 2] = (1 - p_{S,11})/(2 - p_{S,11} - p_{S,22})$.

5.2 Empirical Results

Data employed is the seasonally-adjusted postwar U.S. industrial production index, which is obtained from the Federal Reserve Bank of St. Louis economic database (FRED), and the sample covers the period 1947:M1–2019:M9. Figure 3 depicts the growth rate of the industrial production index. We estimate both the proposed model and the model with a normality assumption for the error term. We obtain 500,000 MCMC draws and discard the first 100,000 to guarantee the convergence of the sampler and to avoid the effect of the initial values. All the inferences are based on the remaining 400,000 draws. We first consider the following tight priors:

Priors #1: Tight Priors

$$\gamma_2 \sim N(0.1, 0.1)_{[\gamma_2>0]}, \quad \gamma_3 \sim N(0.2, 0.2)_{[\gamma_3>0]},$$

$$\eta_1 \sim N(1.5, 0.1)_{[\eta_1>0]}, \quad \eta_2 \sim N(0.5, 0.2)_{[\eta_2>0]}, \quad \eta_3 \sim N(0.2, 0.5)_{[\eta_3>0]},$$

$$\phi \sim N(0.5, 0.5)_{[|\phi|<1]}, \quad (1 + b_2) \sim IG(4, 4), \quad (1 + b_3) \sim IG(4, 8),$$

$$[p_{S,11}, P_{S,12}]' \sim Dir(0.45, 0.05)_{P_{S,11}>0.5}, \quad [P_{S,21}, P_{S,22}]' \sim Dir(0.05, 0.45)_{P_{S,22}>0.5},$$

$$[P_{W,11}, P_{W,12}, P_{W,13}]' \sim Dir(0.9, 0.05, 0.05)_{P_{W,11}>0.5},$$

$$[P_{W,21}, P_{W,22}, P_{W,23}]' \sim Dir(0.05, 0.9, 0.05)_{P_{W,22}>0.5},$$

$$[P_{W,31}, P_{W,32}, P_{W,33}]' \sim Dir(0.05, 0.05, 0.9)_{P_{W,33}>0.5},$$

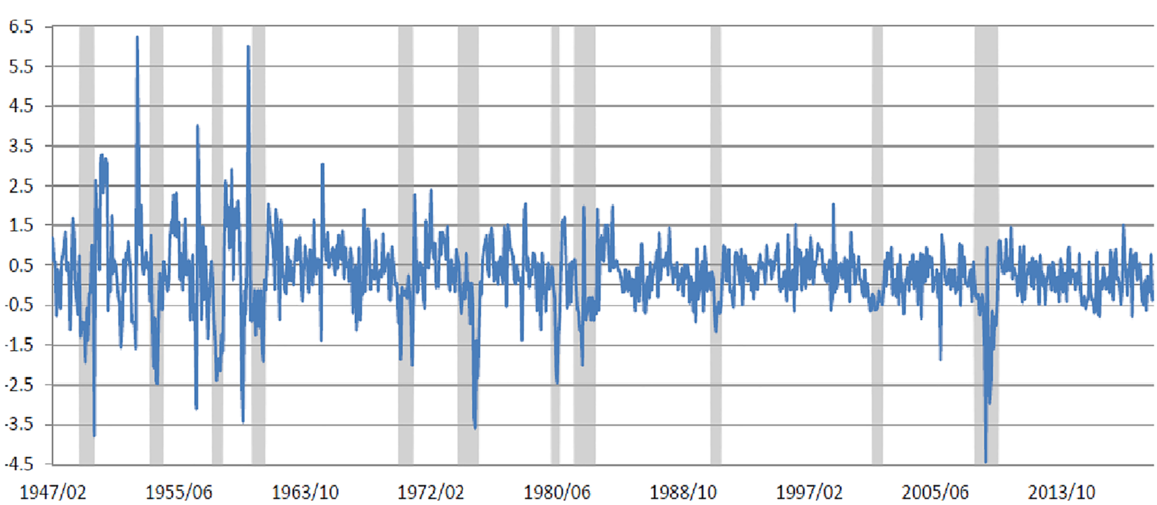


Figure 3: U.S. Industrial production (IP) index growth [1947:M1–2019:M9].

$$P_{C,11} \sim Dir(9.9, 0.1)_{P_{C,11}>0.5}, \quad P_{C,22} \sim Dir(9.9, 0.1)_{[P_{C,22}>0.5]},$$

$$\left(\mu_m, h_m^2\right) \sim G_0 \equiv N\left(-0.5, 3h_m^2\right)IG(17, 4),$$

where the base distribution G_0 specified for the Dirichlet process implies the following unconditional distributions for γ_1 and σ_1^2:

$$\gamma_1 \sim N(-0.5, 0.2) \quad and \ \sigma_1^2 \sim IG(4, 2),$$

which are used as the priors for γ_1 and σ_1^2 for the model with a normality assumption.

When we apply a normality test to the posterior means of the standardized errors from the model with a normality assumption, the null is rejected at a 5 % significance level. This provides us with a justification for employing the proposed model. For the proposed model, the posterior mean for the number of mixtures turns out to be slightly higher than 3, and the null hypothesis of normality is not rejected (at a 5 % significance level) for the posterior means of the standardized errors.[13] These results suggest that the Dirichlet process mixture normals model reasonably well approximates the unknown distribution of the error term. Furthermore, a Bayesian model selection criterion (Watanabe-Akaike information criterion or WAIC by Watanabe (2010)) very strongly prefers the proposed model to the model with a normality assumption.

Figure 4 depicts the posterior probabilities of recession from the two models under the tight priors. The shaded areas represent the NBER recessions. Estimates of the recession probabilities from the proposed model are much sharper and agree much more closely with the NBER reference cycles than those from a model with a normality assumption.

To examine the robustness of the results to the priors employed, we also consider the following loose priors for some of the parameters by keeping the priors for the rest of the parameters unchanged:

Prior #2: Loose Priors

$$\gamma_2 \sim N(0.1, 1)_{[\gamma_2>0]}, \quad \gamma_3 \sim N(0.2, 2)_{[\gamma_3>0]},$$

$$\eta_1 \sim N(1.5, 1)_{[\eta_1>0]}, \quad \eta_2 \sim N(0.5, 2)_{[\eta_2>0]}, \quad \eta_3 \sim N(0.2, 4)_{[\eta_3>0]},$$

$$[P_{S,11}, P_{S,12}]' \sim Dir(0.09, 0.01)_{P_{S,11}>0.5}, \quad [P_{S,21}, P_{S,22}]' \sim Dir(0.01, 0.09)_{P_{S,22}>0.5},$$

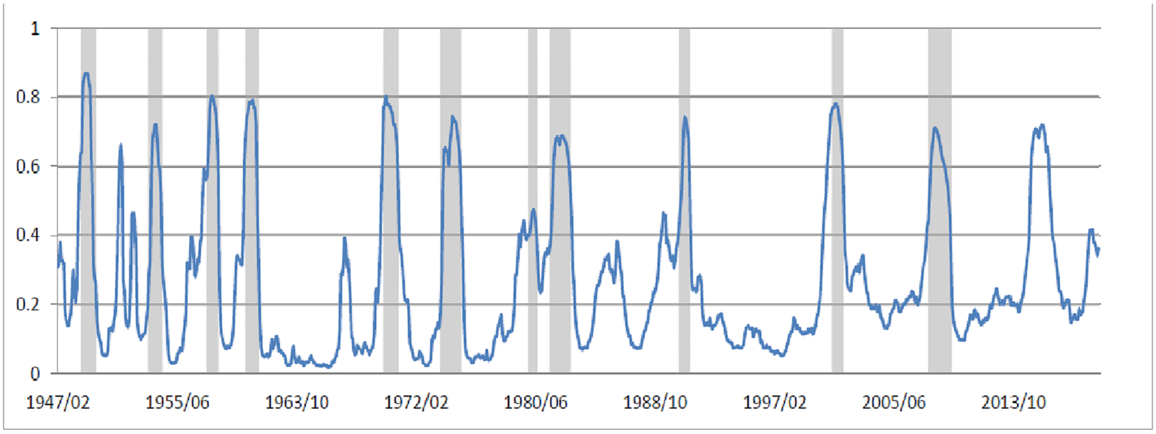
$$P_{C,11} \sim Dir(0.99, 0.01)_{P_{C,11}>0.5}, \quad P_{C,22} \sim Dir(0.99, 0.01)_{P_{C,22}>0.5}.$$

For the case of the loose priors, the prior variances of the parameters are set to be much larger than those for the case of the tight priors. We set the prior means of the parameters to be identical for the two cases. Figure 5 compares the posterior probabilities of recession from the two competing models under the loose priors. For the model with a normality assumption, the inference on the recession probabilities deteriorates considerably with the loose priors. For the proposed model, however, the recession probabilities under the loose priors are almost the same as those under the tight priors, and we continue to have sharp inferences on the recession probabilities. That is, the proposed model is robust to the priors employed, while the model with a normality assumption is very sensitive to the priors.

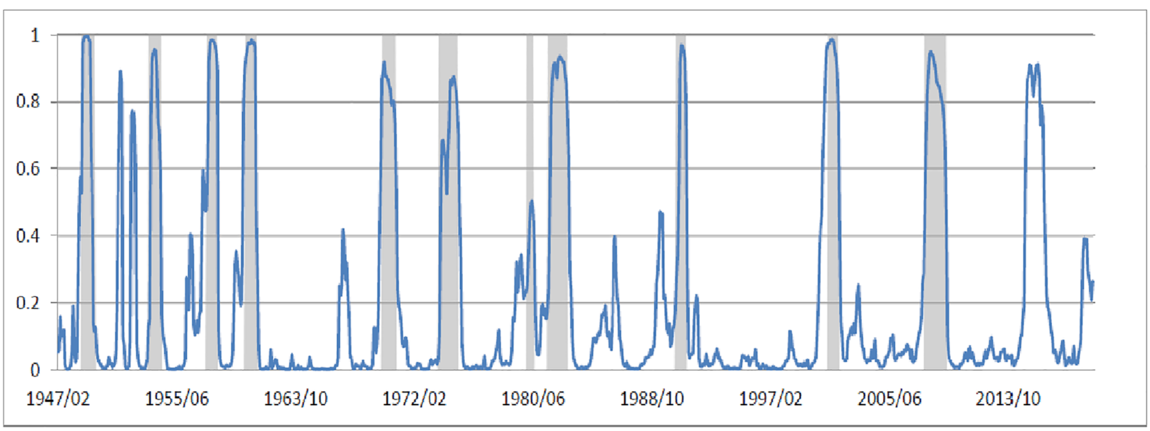
Figure 6 depicts the posterior means of the error volatilities and those of the long-run mean growth rates obtained based on equation (41). These are obtained from the proposed model under the tight priors.[14] As shown in the upper panel of Figure 6, the high and medium volatility regimes are mostly focused on the period before the mid-1980s. In most of the post-1984 period, the low volatility regime dominates except for a few episodes of medium or high volatility that include the Great Recession. The lower panel of Figure 6 demonstrates a pattern for a steadily decreasing long-run mean growth rate, which is consistent with Stock and Watson (2012) and Eo and Kim (2016).

13 To calculate the Jarque-Bera test statistic for the normality test, we use the posterior means of the standardized errors ($\upsilon_t = \frac{1}{h_{D_t}}(\varepsilon_t - \frac{1}{g_{W_t}}\mu_{D_t})$) obtained based on equation (23), for $t = 1, 2, \ldots, T$.

14 The results are almost the same when we employ the loose priors.



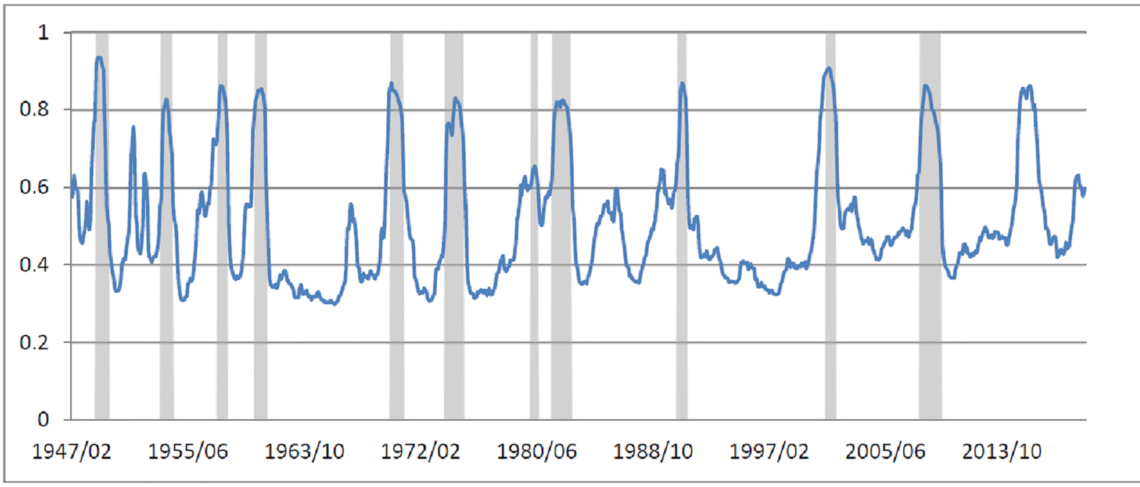
(i) Model with Normally Assumption



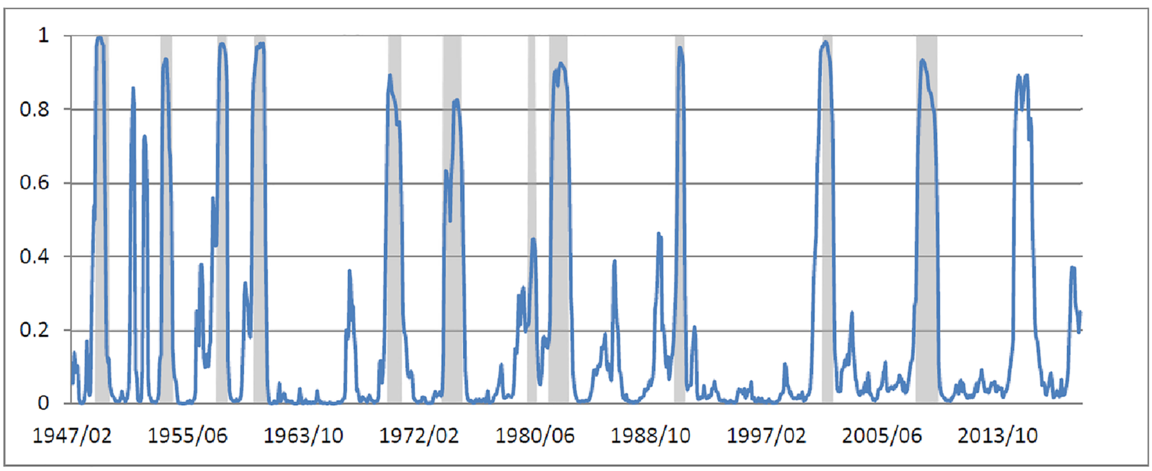
(ii) Model with Mixture of Normals: Proposed Mode

Figure 4: Posterior probabilities of recession based on the two competing models: Tight priors.

Lastly, the upper panel of Figure 7 shows that the posterior distribution of the error term is bimodal before the mixture of normals is controlled for. However, the lower panel of Figure 7 shows that, once the mixture of normals is controlled for, the distribution of the error term is very close to the normal distribution.

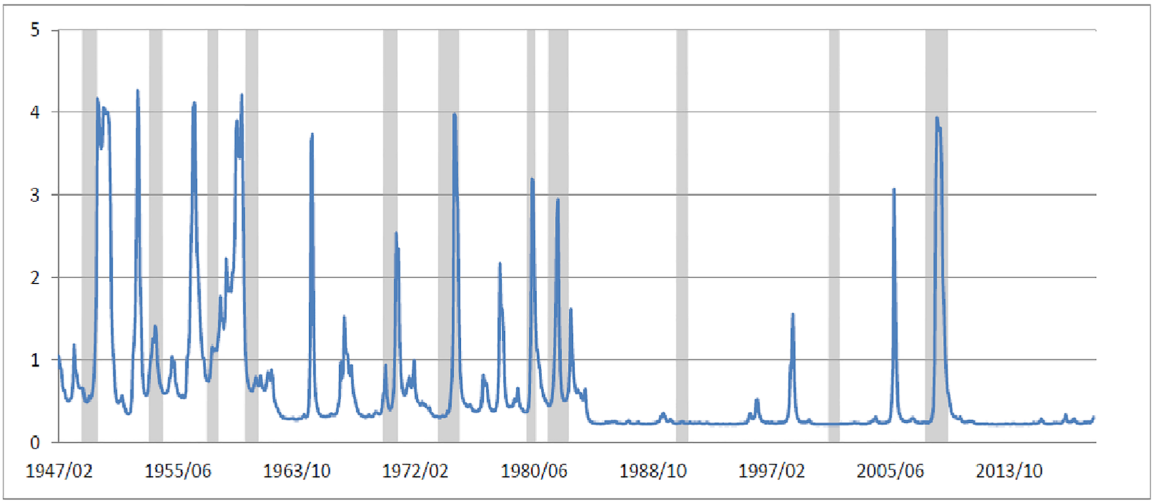


(i) Model with Normally Assumption

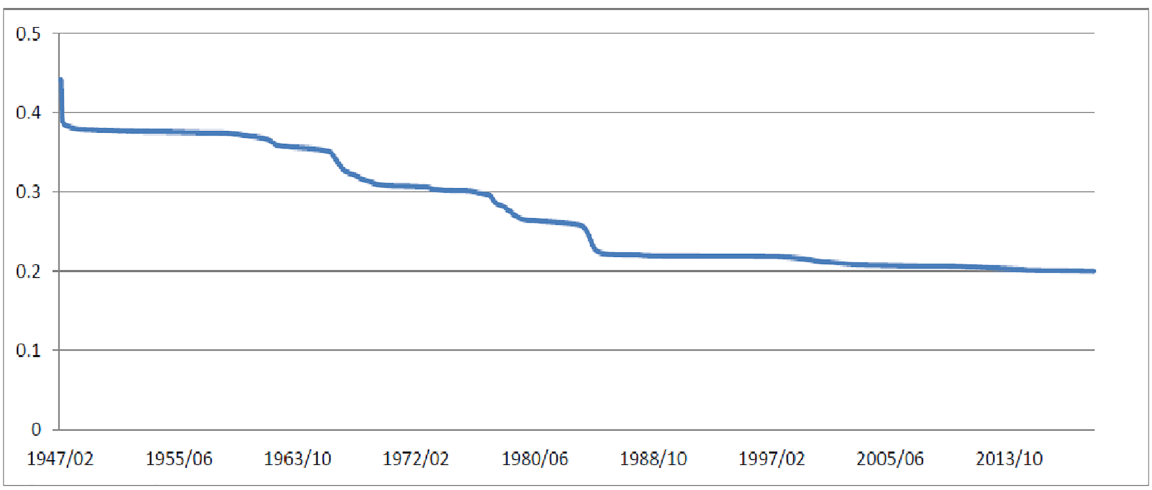


(ii) Model with Mixture of Normals: Proposed Model

Figure 5: Posterior probabilities of recession based on the two competing models: Loose priors.

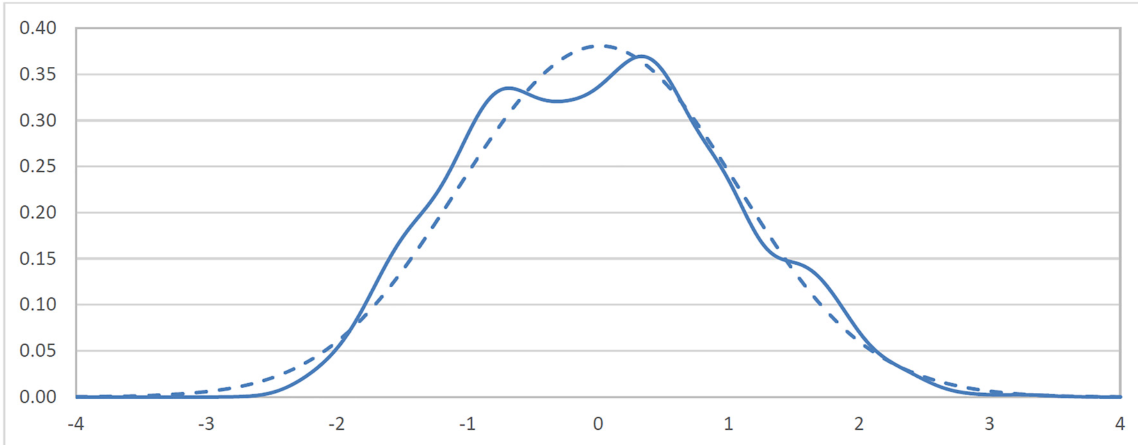


(i) Volatility of Error Term

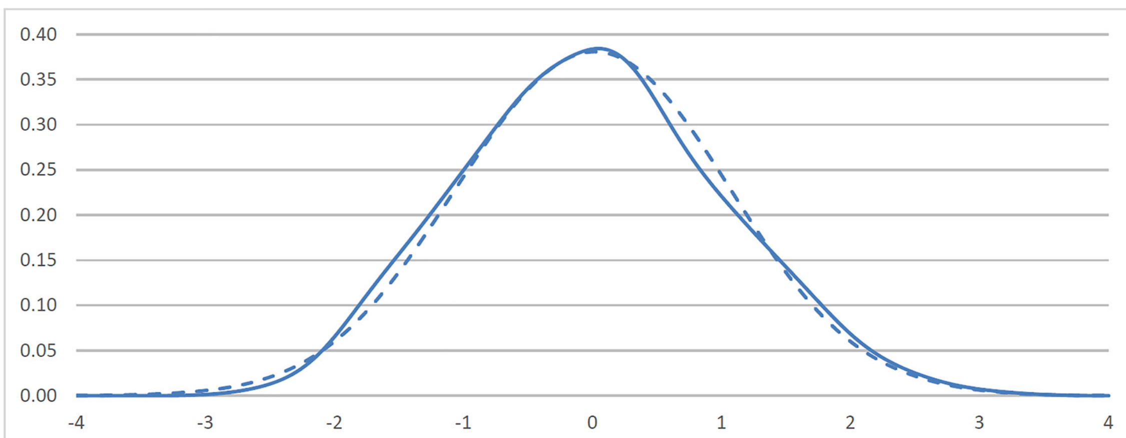


(ii) Long-Run Mean Growth Rate

Figure 6: Time-varying volatility and long-run mean growth rate of IP: Proposed model [Tight Prior].



(a) Standardized Errors before Controlling for Mixture of Normals



(b) Standardized Errors after Controlling for Mixture of Normals

Figure 7: Distribution of the standardized errors (solid line: Standardized errors; broken line: Standard Normal).

6 Summary

We provide solutions to the identification problems that are associated with the estimation of a Markov-switching model in which the unknown error distribution is approximated by the Dirichlet process mixture of normals: (i) the problem of label switching for the Markov-switching regime indicator variable; and (ii) the problem of disentangling the Markov-switching regime indicator variable from the serially independent mixture indicator variable. These solutions are very easy to implement in actual applications, and our Monte Carlo experiments show that the proposed identification schemes and MCMC procedure work well.

When the proposed model and the MCMC procedure are applied to the monthly index of industrial production (1947:M1–2019:M9), they provide us with considerably sharper and more accurate inferences on the business cycle turning points than the model with an assumption of the normally distributed error term. In our model, the irregular components that are not related to business conditions are effectively controlled for.

Acknowledgment: Kim acknowledges financial support from the Bryan C. Cressey Professorship at the University of Washington.

References

Angelis, L. D., and V. Cinzia. 2017. "A Markov-Switching Regression Model with Nongaussian Innovations: Estimation and Testing." *Studies in Nonlinear Dynamics and Econometrics* 21: 1—22.

Bauwens, L., J.-F. Carpantier, and A. Dufays. 2017. "Autoregressive Moving Average Infinite Hidden Markov-Switching Models." *Journal of Business & Economic Statistics* 35: 162—82.

Bulla, J., S. Mergner, I. Bulla, A. Sesboüé, and C. Chesneau. 2011. "Markov-Switching Asset allocation: Do Profitable Strategies Exist?" *Journal of Asset Management* 12: 310—21.

Chib, S. 1998. "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics* 86: 221—41.

Clark, T. E. 2009. "Is the Great Moderation over? an Empirical Analysis." *Economic Review*: 5—42. Fourth Quarter 2009, Federal Reserve Bank of Kansas City.

Diebold, F. X., J.-H. Lee, and G. C. Weinbach. 1994. "Regime Switching with Time-Varying Transition Probabilities." In *Nonstationary Time Series Analysis and Cointegration, Advanced Texts and Econometrics*, edited by C. Hargreaves, 259—305. Oxford and New York: Oxford University Press.

Dueker, M. 1997. "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility." *Journal of Business & Economic Statistics* 15: 26—34.

Eo, Y., and C.-J. Kim. 2016. "Markov-Switching Models with Evolving Regime-specific Parameters: Are Booms or Recessions All Alike?" *The Review of Economics and Statistics* 98: 940—9.

Escobar, M., and M. West. 1995. "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association* 90: 577—88.

Filardo, A. J. 1994. "Business Cycle Phases and Their Transitional Dynamics." *Journal of Business & Economic Statistics* 12: 299—308.

Fox, E., E. Sudderth, M. Jordan, and A. Willsky. 2011. "A Sticky HDP-HMM with Application to Speaker Diarization." *Annals of Applied Statistics* 5: 1020—56.

Frühwirth-Schnatter, S. 2001. "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models." *Journal of the American Statistical Association* 96: 194—209.

Frühwirth-Schnatter, S. 2006. *Finite Mixture and Markov Switching Models*. New York: Springer. Springer Series in Statistics.

Gadea-Rivas, M., A. Gomez-Lscos, and G. Perez-Quiros. 2014. *The Two Greatest Recession Vs. Great Moderation*. Banco de Espana Working Paper No. 1423.

Hamilton, J. D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57: 357—84.

Hamilton, J. D. 2016. "Macroeconomic Regimes and Regime Shifts." *Handbook of Macroeconomics* 2: 163—201.

Kaufmann, S. 2015. "K-state Switching Models with Time-Varying Transition Distributions — Does Loan Growth Signal Stronger Effects of Variables on Inflation?" *Journal of Econometrics* 187: 82—94.

Kim, C.-J. 1994. "Dynamic Linear Models with Markov Switching." *Journal of Econometrics* 60: 1—22.

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." *The Review of Economics and Statistics* 81: 608—16.

Neal, R. 2000. "Markov Chain Sampling Methods for Dirichlet Process Mixture Models." *Journal of Computational & Graphical Statistics* 9: 249—65.

Song, Y. 2014. "Modeling Regime Switching and Structural Breaks with an Infinite Hidden Markov Model." *Journal of Applied Econometrics* 29: 825—42.

Stephens, M. 2000. "Dealing with Label Switching in Mixture Models." *Journal of the Royal Statistical Society: Series B* 62: 795—809.

Stock, J., and M. Watson. 2012. "Disentangling the Channels of the 2007-09 Recession." *Brookings Papers on Economic Activity* 43: 81—156.

Taddy, M., and A. Kottas. 2009. "Markov Switching Dirichlet Process Mixture Regression." *Bayesian Analysis* 4: 793—816.

Watanabe, S. 2010. "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory." *Journal of Machine Learning Research* 11: 3571—94.

West, M., P. Muller, and M. Escobar. 1994. "Hierarchical Priors and Mixture Models with Application in Regression and Density Estimation." In *Aspects of Uncertainty: A Tribute to D. V. Lindley*, edited by A. F. M. Smith and P. R. Freeman, 363—86. London: John Wiley and Sons.