Markov-Switching Models with Unknown Error Distributions

by

Shih-Tang Hwu [1]
University of Washington
and
Chang-Jin Kim
University of Washington

September, 2017

(*Job Market Paper*)

Abstract

To this day, the basic Markov-switching model has been extended in various ways ever since the seminal work of Hamilton (1989). Without exception, however, estimation of Markov-switching models in the literature has relied upon parametric assumptions on the distribution of the error term. In this paper, we first examine the pitfalls of estimating Markov-switching models by maximizing a normal log-likelihood when the normality assumption is violated. We then 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 a special attention to identification of the model. We apply the proposed model to the growth of post-war U.S. industrial production index in order to investigate its regime-switching dynamics. Our univariate model can effectively control for the irregular components that is not related to business conditions. This leads to sharp and accurate inferences on recession probabilities just like the dynamic factor models of Kim and Yoo (1995), Chauvet (1998), and Kim and Nelson (1998) do.

Key Words: Markov-switching, Dirichlet Process, Mixture of Normals, Business Cycle, Dynamic Factor Model.

---

[1] **Hwu:** Department of Economics, University of Washington, Seattle, WA 98195 [Email: hwus@uw.edu]; **Kim:** Department of Economics, University of Washington, Seattle, WA. 98195 [Email: changjin@uw.edu]. Kim acknowledges financial support from the Bryan C. Cressey Professorship at the University of Washington.

1. Introduction

Since the seminal work of Hamilton (1989), the basic Markov-switching model has been extended in various ways. For example, Diebold et al. (1994) and Filardo (1994) extend the model to allow the transition probabilities governing the Markov process to be a function of exogenous or predetermined variables. Kim (1994) extends the model to the state-space representation of general dynamic linear models, which includes autoregressive moving average processes, unobserved components models, dynamic factor models, etc. Chib (1998) introduces a structural break model with multiple change-points by constraining the transition probabilities of the Markov-switching model so that the state variable can either stay at the current value or jump to the next higher value. More recently, Fox et al. (2011), Song (2014), and Bauwens et al. (2017) introduce infinite hidden Markov models and generalize the finite-state Markov switching model of Hamilton (1989) to an infinite number of states. Their models integrate the regime switching and structural break dynamics in a unified Bayesian framework. For these models, the number of states is possibly infinite and is determined when estimating the model.

Without a single exception, estimation of the aforementioned models and the other Markov-switching models in the literature has relied upon parametric assumptions on the distribution of the error terms. Most applications of Markov-switching models in the literature assume normally distributed error terms, with rare exceptions like Dueker (1997) who proposes a model of stock returns in which the innovation comes from a Student-t distribution. The question then would be: what if a normal log-likelihood is maximized but the normality assumption is violated? Even though White (1994) shows that the quasi-maximum likelihood estimators (QMLEs) are consistent and asymptotically normally distributed under some regularity conditions, little is known about the properties of the QMLEs for Markov-switching models. We thus performed a simulation study in order to investigate the finite sample properties of the QMLEs, leading to a conclusion that quasi-maximum likelihood estimation could lead to sizable bias in the parameter estimates and poor inferences about regime probabilities, even for a sample size as large as 5,000. [2]

[2] We deal with this issue in Section 2.

In Bayesian semi-parametric econometrics, approximating an unknown distribution based on a mixture of normals is popular as surveyed in Marin et al. (2005). There are two alternative models for achieving the goal. They are: i) the finite mixture normals model in which the number of states is fixed, and ii) the Dirichlet process mixture normals model in which the number of states is treated as a random variable. Kim et al. (1998) and Omori et al. (2007) demonstrate the usefulness of the finite mixture of normals in approximating the log chi-square distribution in stochastic volatility models; and Alexander and Lazar (2006) employ it to approximate the unknown error distribution in a GARCH model. More recently, Jensen and Maheu (2013) apply the Dirichlet process mixture of normals to a multivariate GARCH model; Jensen and Maheu (2010, 2014) apply it to deal with unknown error distributions in stochastic volatility models; and Jin and Maheu (2016) apply it for Bayesian semi-parametric modeling of realized covariance matrices.

The goal of this paper is to present a Bayesian approach to estimating Markov-switching models without imposing a priori parametric assumption on the distribution of the error term. We implement the Dirichlet process mixture normals model in order to approximate the unknown and potentially non-normal error distribution. We note that, in order to allow for an asymmetric, as well as fat-tailed, error distribution within a Markov-switching model, special attention has to be paid to the identification of the model.

We apply the proposed model to the growth of postwar U.S. Industrial Production index covering the period January 1947-January 2017. We demonstrate that a model with a normality assumption performs poorly in identifying the NBER reference cycles. The null hypothesis of normality for the error term is rejected at a 5% significance level. However, the proposed univariate model can effectively control for the irregular components that are not related to business conditions. This leads to sharp and accurate inferences on recession probabilities just like the dynamic factor models of Kim and Yoo (1995), Chauvet (1998), and Kim and Nelson (1998) do. Furthermore, the null of normality is not rejected for the standardized error term that is obtained conditional on the mixing indicator variable.

The rest of this paper is organized as follows. In Section 2, we motivate our paper by exploring the finite sample properties of the quasi-maximum likelihood estimation of Markov-switching models. We discuss our model specifications with special attention to identification issues in Section 3. In Section 4, we present a Markov Chain Monte Carlo (MCMC) algorithm for estimating the proposed model. Section 5 provides an empirical application of the proposed model, and Section 6 concludes the paper.

2. Quasi-Maximum Likelihood Estimation of Markov-Switching Models: Finite Sample Properties Based on Simulation Study

In this section, we investigate finite sample properties of the quasi-maximum likelihood estimation (QMLE) of Markov-switching models when a normal log-likelihood is maximized but the normality assumption is violated. For this purpose, we consider the following model with Markov-switching mean and variance:

$$\begin{aligned} y_t = \beta_{S_t} + h_{S_t}\varepsilon_t, \quad \varepsilon_t \sim i.i.d(0,1), \quad S_t = 1,2, \\ t = 1,2,...,T, \end{aligned} \tag{1}$$

where S_t is a 2-state Markov-switching process with transition probabilities

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

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

<u>*Case #1*</u>

$$\varepsilon_t \sim i.i.d. \;\; N(0,1)$$

<u>*Case #2*</u>

$$\varepsilon_t \sim \frac{u_t}{\sqrt{\nu/(\nu-2)}}, \quad u_t \sim i.i.d. \;\; t-distribution \;\; with \;\; d.f. = \nu$$

<u>*Case #3*</u>

$$\varepsilon_t = \frac{ln(u_t^2) - E(ln(u_t^2))}{\sqrt{(var(ln(u_t^2))}}, \quad u_t \sim i.i.d. \;\; N(0,1),$$

where $E(\ln v_t^2) = -1.2704$, $var(\ln v_t^2) = \pi^2/2$.

Case #4

$$\varepsilon_t \, | D_t \sim i.i.d. \;\; N(\mu_{D_t}, \sigma^2_{D_t}), \;\; D_t = 1, 2, 3,$$
$$Pr[D_t = 1] = w_1, \;\; Pr[D_t = 2] = w_1, \;\; Pr[D_t = 1] = w_3$$

For each of the above four cases, we generate 1,000 sets of data. For each data set generated, we estimate the model in equations (1) and (2) by maximizing a normal log-likelihood. While we have exact maximum likelihood estimation for Case #1, we have QMLE for the other 3 cases. We consider two alternative sample sizes: $T = 500$ and $T = 5000$. The parameter values we assign are given below:

$$\beta_1 = -0.5, \quad \beta_2 = 1; \quad h_1 = 2, \quad h_2 = 1; \quad p_{11} = 0.9, \quad p_{22} = 0.95;$$

$$\nu = 5;$$

$$\mu_1 = 0.72, \quad \mu_2 = 0, \quad \mu_3 = -1.8; \quad \sigma_1^2 = 0.025, \quad \sigma_2^2 = 0.2, \quad \sigma_3^2 = 0.1;$$

$$w_1 = 0.5, \quad w_2 = 0.3, \quad w_3 = 0.2$$

For each of the above four cases and for each parameter, Table 1 reports the mean of 1,000 point estimates, as well as the root mean squared error (RMSE) of the estimates from the true value. For case #1, in which we have normally distributed error term, the mean parameter estimates are very close to their true values for both sample sizes. For the other cases in which the error term is not normally distributed, the mean parameter estimates are far from their true values. Note that the mean parameter estimates are almost identical when $T = 500$ or $T = 5,000$, suggesting the bias in these parameter estimates may not just be a small sample phenomenon. When we compare the results among Cases #2, #3, and #4, the bias is smallest for Case #2, in which the error term is non-normal but symmetrically distributed.

In order to investigate how inferences on regime probabilities are affected by the violation of the normality assumption and the QMLE, we conduct another simulation study. When generating data, we consider the same data generating processes as given above, except that we generate S_t, $t = 1, 2, ..., T$, 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 smoothed probabilities conditional on estimated parameters. Figure 1 plots the average smoothed probabilities of high-mean regime for each case. The shaded areas represent the true high-mean regime. Case #1 with the normal error term provides us with the sharpest regime inferences. However, as the distribution of the error term deviates from normality, inferences about regime probabilities deteriorate a lot especially for Cases #3 and #4, in which the error terms are asymmetrically distributed.

The simulation study in this section clearly demonstrates the pitfalls of estimating Markov-switching models by maximizing a normal log-likelihood when the normality assumption is violated. Quasi-maximum likelihood estimation results in inconsistent parameter estimates and poor inferences about regime probabilities. In the next two sections, we introduce a Bayesian approach to estimating Markov-switching models with unknown and potentially non-normal error distributions.

3. Model Specifications and Identification Issues

3.1. Basic Model Specifications

We consider the following Markov-switching regression model:

<u>*Specification #1*</u>

$$\begin{aligned} y_t =& \beta^*_{1,S_t} + \beta_{2,S_t} x_{2t} + \ldots + \beta_{k,S_t} x_{kt} + h^*_{S_t} \varepsilon^*_t, \quad S_t = 1, 2, \ldots, N, \\ & h_1^{*2} < h_2^{*2} < \ldots < h_N^{*\,2}, \end{aligned} \tag{3}$$

$$\varepsilon^*_t \sim i.i.d.(0, 1), \tag{4}$$

where S_t is an N-state first order Markov-switching process with transition probabilities

$$Pr[S_t = j | S_{t-1} = i] = p_{ij}, \quad i, j = 1, \ldots, N. \tag{5}$$

Here, the distribution of the error term ε_t^* is unknown and potentially non-normal. We approximate the distribution of ε_t^* by the following mixture of normals: [3]

$$\begin{aligned} \varepsilon_t^*|D_t \ &\sim i.i.d. \ \ N(\mu_{D_t}^*, \sigma_{D_t}^{*2}), \quad D_t = 1, 2, ..., M, \\ &\sigma_1^{*2} < \sigma_2^{*2} < \ldots < \sigma_M^{*2}, \end{aligned} \tag{6}$$

where D_t is the mixture indicator variable which is independent of S_t. It is serially independent with the following mixture probabilities:

$$Pr[D_t = m] = w_m, \quad m = 1, 2, ..., M. \tag{7}$$

As the unconditional expectation and variance of ε_t^* are 0 and 1, respectively, we have the following restrictions on the conditional means and variances of ε_t^*:

$$\sum_{m=1}^{M} \mu_m^* w_m = 0; \quad and \quad \sum_{m=1}^{M} (\sigma_m^{*2} + \mu_m^{*2}) w_m = 1. \tag{8}$$

Bayesian inference of the above model with restrictions in equation (8) does not seem to be very straightforward. In order to circumvent the difficulties associated with imposing these restrictions, we consider the following alternative representation of the model:

<u>*Specification #2*</u>

$$\begin{aligned} y_t = \beta_{1,S_t} + \beta_{2,S_t} x_{2t} + \ldots + \beta_{k,S_t} x_{kt} + h_{S_t}\varepsilon_t, \quad S_t = 1, 2, ..., N, \\ (\Rightarrow \ \ y_t = x_t'\beta_{S_t} + h_{S_t}\varepsilon_t \ ) \\ h_1^2 < h_2^2 < \ldots < h_N^2, \end{aligned} \tag{9}$$

$$\varepsilon_t \sim i.i.d.(\bar{\mu}, \bar{\sigma}^2). \tag{10}$$

Conditional on the mixture indicator variable D_t, the distribution of ε_t is specified as:

$$\begin{aligned} \varepsilon_t|D_t \ &\sim i.i.d. \ \ N(\mu_{D_t}, \sigma_{D_t}^2), \quad D_t = 1, 2, ..., M, \\ &\sigma_1^2 < \sigma_2^2 < \ldots < \sigma_M^2, \end{aligned} \tag{11}$$

where

[3] We allow for potential asymmetry in the distribution of ε_t^*. Note that in case $\mu_m^* = 0$ for all m, the distribution is ε_t^* is symmetric.

$$\sum_{m=1}^{M} \mu_m w_m = \bar{\mu}; \quad and \quad \sum_{m=1}^{M} (\sigma_m^2 + (\mu_m - \bar{\mu})^2) w_m = \bar{\sigma}^2. \tag{12}$$

with w_m referring to the mixture probability in equation (7).

While identification of the model in Specification #1 is achieved by normalizing the unconditional expectation and variance of ε_t^* to be 0 and 1, respectively, we can achieve identification of the model in Specification #2 by imposing the following normalizations:

$$\beta_{1,1} = 0; \quad and \quad h_1^2 = 1. \tag{13}$$

The Markov Chain Monte Carlo (MCMC) algorithm presented in Section 4 is based on Specification #2, and the parameters of the original model (Specification #1) can be recovered as follows:

Relation between Parameters for Specifications #1 and #2

$$\beta_{1,1}^* = \bar{\mu}; \quad h_1^{*2} = \bar{\sigma}^2$$

$$\beta_{1,j}^* = \beta_{1,j} + \bar{\mu}; \quad h_j^{*2} = h_j^2 \times \bar{\sigma}^2, \quad j = 2, 3, ..., N$$

3.2. Bayesian Modeling of the Finite Mixture of Normals and the Dirichlet Process Mixture of Normals

The Dirichlet process mixture of normals that we employ in this paper builds on the finite mixture of normals. In order to help understand the Dirichlet process mixture of normals and its relation to the finite mixture of normals, we review both models in this section. When the total number of mixtures, M, is fixed and pre-specified, we have the following specification for finite mixture of normals:

Finite Mixture of Normals

$$
\begin{aligned}
\varepsilon_t|D_t \;&\sim\; i.i.d.\;\; N(\mu_{D_t},\sigma^2_{D_t}),\;\; D_t=1,2,...,M,\\
(w_1,w_2,\ldots,w_M) \;&\sim\; Dirichlet(\frac{\alpha}{M},\ldots,\frac{\alpha}{M}),\\
(\mu_m,\sigma^2_m) \;&\sim\; G_0, \qquad m=1,2,...,M,\\
\sigma^2_1<\sigma^2_2<\ldots<\sigma^2_M&,\\
G_0 \;&\equiv\; N(\lambda_0,\psi_0\sigma^2_m)IG(\frac{\delta_0}{2},\frac{\nu_0}{2}),
\end{aligned}
\tag{14}
$$

where w_m is the mixing probability in equation (7) and G_0, the joint prior distribution of (μ_m,σ^2_m), is assumed to be Normal-Inverse-Gamma. The α parameter can be either fixed or random.

For the above finite mixture of normals, the prior probability of D_t conditional on $\tilde{D}_{\neq t}$ can be derived as: [4]

$$
\begin{aligned}
&Pr[D_t=m|\tilde{D}_{\neq t},\alpha]=\frac{T_{m,\neq t}+\frac{\alpha}{M}}{T-1+\alpha},\quad m=1,2,...,M,\\
&(with\;\sum_{m=1}^{M}Pr[D_t=m|\tilde{D}_{\neq t},\alpha]=1)
\end{aligned}
\tag{15}
$$

where $\tilde{D}_{\neq t}=[\,D_1\quad\ldots\quad D_{t-1}\quad D_{t+1}\quad\ldots\quad D_T\,]'$ is the collection of mixing indicators excluding D_t; and $T_{m,\neq t}$ is the total number of observations that belong to class m in a sample that excludes period t. An important thing to note is that the above probabilities always add up to 1. With this background, we are now ready to discuss the Dirichlet process mixture of normals and its properties.

As suggested by Neal (2000), Gorur and Rasmussen (2010), and others, the limit of the model in equation (15) as $M\to\infty$ is equivalent to the Dirichlet process mixture of normals. A formal specification for the Dirichlet process mixture of normals is given below:

Dirichlet Process Mixture of Normals

[4] Proof of equation (15) is given in Appendix A.

$$
\begin{aligned}
\varepsilon_t|D_t \ &\sim \ i.i.d. \ \ N(\mu_{D_t}, \sigma^2_{D_t}), \ D_t = 1, 2, ..., M, \\
(\mu_m, \sigma^2_m) \ &\sim \ G, \qquad m = 1, 2, ..., M, \\
\sigma^2_1 < \sigma^2_2 &< \ldots < \sigma^2_M, \\
G \mid G_0, \alpha \ &\sim DP(\alpha, G_0) \\
G_0 \ &\equiv \ N(\lambda_0, \psi_0 \sigma^2_m) IG(\frac{\delta_0}{2}, \frac{\nu_0}{2}),
\end{aligned} \tag{16}
$$

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

Here, M is a random variable, potentially infinite, that is to be estimated. Note that in the case of the finite mixture of normals, the joint distribution of (μ_m, σ^2_m) is given by G_0, and thus, $G \equiv G_0$. In the case of the Dirichlet process mixture of normals, however, the joint distribution of (μ_m, σ^2_m) is a random distribution generated by a Dirichlet process with based distribution G_0 and the concentration parameter α. [5]

The prior probability of D_t conditional on $\tilde{D}_{\neq t}$ can be obtained by taking the limit $M \rightarrow \infty$ for equation (15), as given below:

$$
\begin{aligned}
&Pr[D_t = m|\tilde{D}_{\neq t}, \alpha] = \frac{T_{m,\neq t}}{T - 1 + \alpha}, \ m = 1, 2, ..., M^*_{\neq t}, \\
&(with \ \sum_{m=1}^{M} Pr[D_t = m|\tilde{D}_{\neq t}, \alpha] < 1)
\end{aligned} \tag{17}
$$

where $T_{m,\neq t}$ is is defined earlier and $M^*_{\neq t}$ is the total number of distinctive classes (or mixtures) realized in the sample that excludes period t.

Unlike the case of the finite mixture of normals in equation (15), the above probabilities do not add up to 1, suggesting that there always exists non-zero probability that an observation at period t belongs to a new class that does not belong to the existing $M^*_{\neq t}$ classes. This probability is given below:

[5] That is, the Dirichlet process provides a random distribution over distributions on infinite sample spaces. The hierarchical models in which the Dirichlet process is used as a prior over the distribution of the parameters are referred to as the Dirichlet process mixture model.

$$
\begin{aligned}
Pr[D_t = M^*_{\neq t} + 1 | \tilde{D}_{\neq t}, \alpha] &= 1 - \sum_{m=1}^{M^*_{\neq t}} Pr[D_t = m | \tilde{D}_{\neq t}, M^*_{\neq t}] \\
&= \frac{\alpha}{T - 1 + \alpha},
\end{aligned}
\tag{18}
$$

which suggests that, if α is larger, the prior mean of M is higher with less concentrated distribution for G in equation (16).

The α parameter can be either fixed or random. In case α is treated as random, its conjugate prior is the Gamma distribution, given below:

$$
\alpha \sim Gamma(a, b), \quad a > 0, \; b > 0. \tag{19}
$$

4. Estimation of the Model

We denote $\tilde{\theta}_1$ as a vector that contains all the parameters associated with the Markov-switching regression equation in (9), as given below:

$$
\tilde{\theta}_1 = [\, \tilde{\beta}' \quad \tilde{h}^{2\prime} \quad \tilde{p}' \,]',
$$

where $\tilde{\beta} = [\, \tilde{\beta}'_{1,\neq 1} \quad \tilde{\beta}'_2 \quad \ldots \quad \tilde{\beta}'_k \,]'$ with $\tilde{\beta}_{1,\neq 1} = [\, \beta_{1,2} \quad \ldots \quad \beta_{1,N} \,]'$ and $\tilde{\beta}_i = [\, \beta_{i,1} \quad \ldots \quad \beta_{i,N} \,]'$, $i = 2, 3, ..., k$; $\tilde{h}^2 = [\, h_2^2 \quad h_3^2 \quad \ldots \quad h_N^2 \,]'$; and $\tilde{p}$ is a vector that contains the transition probabilities of S_t.

For the parameters associated with the Dirichlet process mixture of normals for ε_t, we define

$$
\tilde{\theta}_2 = [\, \tilde{\mu}' \quad \tilde{\sigma}^{2\prime} \quad \alpha \quad M \,]',
$$

where $\tilde{\mu} = [\, \mu_1 \quad \ldots \quad \mu_M \,]'$; $\tilde{\sigma}^{2\prime} = [\, \sigma_1^2 \quad \ldots \quad \sigma_M^2 \,]'$; and $\tilde{D}_T = [\, D_1 \quad \ldots \quad D_T \,]'$.

Then, the hierarchical nature of our model allows us to decompose the posterior distribution of our interest as follows:

$$
\begin{aligned}
f(\tilde{\theta}_1, \tilde{\theta}_2, \tilde{S}_T, \tilde{D}_T | \tilde{Y}_T) &\propto f(\tilde{\theta}_2, \tilde{D}_T | \tilde{\theta}_1, \tilde{S}_T, \tilde{Y}_T) f(\tilde{\theta}_1, \tilde{S}_T | \tilde{Y}_T) \\
&= f(\tilde{\theta}_2, \tilde{D}_T | \tilde{\varepsilon}_T) f(\tilde{\theta}_1, \tilde{S}_T | \tilde{Y}_T),
\end{aligned}
\tag{20}
$$

where $\tilde{S}_T = [\,S_1 \quad \ldots \quad S_T\,]'$, $\tilde{Y}_T = [\,Y_1 \quad \ldots \quad Y_T\,]'$, and $\tilde{\varepsilon}_T = [\,\varepsilon_1 \quad \ldots \quad \varepsilon_T\,]'$. Equation (20) suggests that the MCMC algorithm consists of the following two steps:

Step 1: Draw the variates for the Markov-switching regression model conditional on mixture of Normals and data $\tilde{Y}_T$. That is, draw $\tilde{\theta}_1$ and $\tilde{S}_T$ conditional on $\tilde{\theta}_2$, $\tilde{D}_T$, and data.

Step 2: Draw the variates associated with the mixture of normals conditional on the error term for the Markov-switching regression equation in (9). That is, draw $\tilde{\theta}_2$ and $\tilde{D}_T$ conditional on $\tilde{\varepsilon}_T$.

4.1. Drawing Variates Associated with Markov-switching Regression Equation Conditional on the Mixture of Normals

Equation (11) implies that

$$\varepsilon_t = \mu_{D_t} + \sigma_{D_t} u_t, \quad u_t \sim i.i.d. \;\; N(0,1), \tag{21}$$

and thus, by substituting this into equation (9) and rearranging terms, we obtain

$$y_t^* = x_t^{*\prime}\beta_{S_t} + h_{S_t} z_t + h_{S_t} u_t, \quad u_t \sim i.i.d. \;\; N(0,1), \tag{22}$$

where $y_t^* = \frac{y_t}{\sigma_{D_t}}$; $x_t^* = \frac{x_t}{\sigma_{D_t}}$; and $z_t = \frac{\mu_{D_t}}{\sigma_{D_t}}$ serve as data.

Based on equation (22), we can draw $\tilde{\theta}_1 = [\,\tilde{\beta}' \quad \tilde{h}^{2\prime} \quad \tilde{p}'\,]'$ and $\tilde{S}_T$ in the following sequence:

i) Draw $\tilde{S}_T$ conditional on $\tilde{\beta}$, $\tilde{h}^2$, $\tilde{Y}_T^* = [\,y_1^* \quad \ldots \quad y_T^*\,]'$, $\tilde{X}_T^* = [\,x_1^* \quad \ldots \quad x_T^*\,]'$, and $\tilde{Z}_T = [\,z_1 \quad \ldots z_T\,]'$.

ii) Draw $\tilde{\beta}$ conditional on $\tilde{h}^2$, $\tilde{S}_T$, $\tilde{Y}_T^*$, $\tilde{X}_T^*$, and $\tilde{Z}_T$.

iii) Draw $\tilde{h}^2$ conditional on $\tilde{\beta}$, $\tilde{S}_T$, $\tilde{Y}_T^*$, $\tilde{X}_T^*$ and $\tilde{Z}_T$.

Equation (22) is a standard Markov-switching model. Thus, drawing $\tilde{S}_T$ and $\tilde{\mu}$ from the appropriate full conditional distributions is standard. However, as the standard deviation (h_{S_t}) of the error term enters in the mean function of equation (22), we need to employ the Metropolis Hasting algorithm for drawing $\tilde{h}^2$. We explain how we can draw $\tilde{h}^2$ in what follows.

Drawing $\tilde{h}^2 = [h_2^2, h_3^2, ..., h_N^2]'$ Conditional on $\tilde{S}_T$, $\tilde{\beta}$, $\tilde{D}_T$, $\tilde{\mu}$, $\tilde{\sigma}^2$, and Data

Under the assumption that h_2^2, h_3^2,..., h_N^2 are independent of one another, the full conditional distribution of $\tilde{h}^2$ can be derived as:

$$\begin{aligned} f(\tilde{h}^2|\tilde{\beta}, \tilde{S}_T, \tilde{Y}_T^*) &\propto f(\tilde{Y}_T^*|\tilde{h}^2, \tilde{\beta}, \tilde{S}_T) f(\tilde{h}^2) \\ &= \prod_{i=2}^{N} \left[\prod_{J_i} f(y_t^*|h_i^2, \beta_i) f(h_i^2) \right], \end{aligned} \tag{23}$$

where $J_i = \{t : S_t = i\}$; $\tilde{Y}_t^* = [y_1^*, ..., y_t^*]'$; $f(h_i^2)$ is the prior density for h_i^2 and $f(y_t^*|h_i^2, \beta_i)$ is the density function of y_t given $S_t = i$, $i = 2, \ldots, N$. Equation (23) suggests that each h_i^2 can be drawn separately, and what follows explains how h_i^2, $i = 2, 3, ..., N$, can be generated sequentially.

We employ an inverse Gamma distribution as the prior distribution for each h_i^2, i.e., $h_i^2 \sim IG\left(\frac{d_i}{2}, \frac{v_i}{2}\right)$, $i = 2, \ldots, N$. Combining the prior density with the likelihood function for observations associated with $S_t = i$, we obtain the following posterior distribution for h_i^2:

$$\begin{aligned} f(h_i^2|\tilde{\beta}, \tilde{S}_T, \tilde{Y}_T^*) &\propto \prod_{J_i} f(y_t^*|h_i^2, \beta_i) f(h_i^2) \\ &\propto \prod_{J_i} \frac{1}{h_i^2}^{\frac{d_i-1}{2}} \exp\left[-\frac{(y_t^* - x_t^{*\prime}\beta_i - h_i z_t)^2}{2h_i^2} - \frac{v_i}{2h_i^2} \right], \quad i = 2, ..., N, \end{aligned} \tag{24}$$

The above density function does not belong to any known family of distributions. We thus employ a Metropolis-Hastings (MH) algorithm in order to draw h_i^2, $i = 2, \ldots, N$ from the target density in equation (24). By denoting $h_{i,old}^2$ as the accepted h_i^2 from the previous MCMC iteration and $h_{i,new}^2$ as a candidate for h_i^2, the following steps are sequentially repeated for $i = 2, ..., N$, starting with $i = 2$:

i) Generate $h_{i,new}^2$ from the following random walk candidate generating distribution:

$$h_{i,new}^2 = h_{i,old}^2 + \eta_i, \quad \eta_i \sim N(0, c_i), \quad h_{i,new}^2 > h_{i-1}^2, \tag{25}$$

where $h_1^2 = 1$. The parameter c_i is chosen to get an acceptance probability between 0.2 to 0.5, as suggested by Koop (2003).

ii) Calculate the acceptance probability in the following way:

$$\alpha\left(h_{i,new}^2, h_{i,old}^2\right) = min\left[1, \prod_{J_i} f(h_i^2 = h_{i,new}^2|\tilde{\beta}, \tilde{S}_T, \tilde{Y}_T^*)/\prod_{J_i} f(h_i^2 = h_{i,old}^2|\tilde{\beta}, \tilde{S}_T, \tilde{Y}_T^*)\right], \quad (26)$$

where $f(h_i^2|\tilde{\beta}, \tilde{S}_T, \tilde{Y}_T^*)$ is given in equation (24).

iii) Set $h_i^2 = h_{i,new}^2$ with probability $\alpha\left(h_{i,new}^2, h_{i,old}^2\right)$ and set $h_i^2 = h_{i,old}^2$ with probability $1 - \alpha\left(h_{i,new}^2, h_{i,old}^2\right)$.

iv) Set $i = i + 1$ and go to i).

4.2. Drawing Variates Associated with the Mixture of Normals Conditional on $\tilde{\varepsilon}_T$

Conditional on $\tilde{\beta}$, $\tilde{h}^2$, $\tilde{S}_T$, and data $\tilde{Y}_T$, we can calculate the error term ε_t in equation (9) as

$$\varepsilon_t = \frac{(y_t - x_t'\beta_{S_t})}{h_{S_t}}, \quad t = 1, 2, ..., T. \quad (27)$$

Then, based on equation (21), we can draw the variates associated with mixture of normals in the following sequence:

i) Conditional on $\tilde{\mu}$, $\tilde{\sigma}^2$, and $\tilde{\varepsilon}_T$, draw $\tilde{D}_T$ and α for the Dirichlet process mixture of normals. The total number of mixtures M is generated as a byproduct of generating $\tilde{D}_T$.

ii) Conditional on $\tilde{\sigma}^2$, $\tilde{D}_T$, M, and $\tilde{\varepsilon}_T$, draw $\tilde{\mu}$.

iii) Conditional on $\tilde{\mu}$, $\tilde{D}_T$, M, and $\tilde{\varepsilon}_T$, draw $\tilde{\sigma}^{2'}$

Drawing $\tilde{\mu}$ and $\tilde{\sigma}^2$ from appropriate full conditional distributions derived based on equation (21) is standard. We thus focus on drawing $\tilde{D}_T$ and α in what follows. [6]

[6] This section is largely based on the works of West et al. (1994), Escobar and West (1995), and Neal (2000).

4.2.1. Drawing $\tilde{D}_T$ Conditional on α

If the total number of mixtures, M, were fixed as in the case of the finite mixture of normals, it would be straightforward to generate D_t based on the following full conditional distribution of D_t:

$$f(D_t|\tilde{\mu}, \tilde{\sigma}^2, \tilde{D}_{\neq t}, \varepsilon_t) \propto f(D_t|\tilde{D}_{\neq t}, \alpha) f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t), \quad D_t = 1, 2, ..., M, \tag{28}$$

where $\tilde{D}_{\neq t}$ is the collection of mixing indicators in the sample excluding D_t; $f(D_t|\tilde{D}_{\neq t}, \alpha)$ is the prior probability in equation (15); and $f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t = m) = \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left[-\frac{(\varepsilon_t - \mu_m)^2}{2\sigma_m^2}\right]$. That is, we could draw D_t based on the following probabilities:

$$Pr[D_t = m|\varepsilon_t, \tilde{\mu}, \tilde{\sigma}^2, \tilde{D}_{\neq t}] = \frac{Pr[D_t = m|\tilde{D}_{\neq t}] f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t = m)}{\sum_{m=1}^{M} Pr[D_t = m|\tilde{D}_{\neq t}] f(\varepsilon_t|\mu_m, \sigma_m^2, D_t = m)}, \ m = 1, 2, ..., M. \tag{29}$$

For the Dirichlet process mixture of normals, in which M is a random variable, Neal (2000) suggests that equation (28) should be replaced by:

$$f(D_t|\tilde{\mu}, \tilde{\sigma}^2, D_{\neq t}, \alpha, \varepsilon_t) \propto f(D_t|\alpha, \tilde{D}_{\neq t}) f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t), \quad D_t = 1, \dots, M^*_{\neq t}, M^*_{\neq t} + 1, \tag{30}$$

where $M^*_{\neq t}$ is the number of distinctive classes (or mixtures) in the sample that exclude period t; and $f(D_t|\tilde{D}_{\neq t}, \alpha)$ is the prior probability given in equation (17) or (18). Here, when $D_t = M^*_{\neq t} + 1$, it means that period t belongs to a new class that does not exist in $\tilde{D}_{\neq t}$. Given equation (30), we can then generate D_t using the following probabilities:

$$Pr[D_t = m|\tilde{\mu}, \tilde{\sigma}^2, D_{\neq t}, \alpha, \varepsilon_t] = \frac{Pr[D_t = m|\tilde{D}_{\neq t}, \alpha] f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t)}{\sum_{m=1}^{M^*_{\neq t}+1} Pr[D_t = m|\tilde{D}_{\neq t}, \alpha] f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t)}, \tag{31}$$

$$m = 1, 2, ..., M^*_{\neq t}, M^*_{\neq t} + 1.$$

Depending on whether D_t belongs to the existing class ($m = 1, 2, ..., \ or \ M^*_{\neq t}$) or a new class ($m = M^*_{\neq t} + 1$), we have the following two conditional densities for ε_t:

$$f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t = m) = f_N(\varepsilon_t|\mu_m, \sigma_m^2), \quad for \ m = 1, 2, ..., M^*_{\neq t}; \tag{32}$$

$$f(\varepsilon_t|\tilde{\mu}, \tilde{\sigma}^2, D_t = M^*_{\neq t} + 1) = \int f_N(\varepsilon_t|\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1}) dG_0(\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1}), \tag{33}$$

where $f_N(\cdot|\mu_j, \sigma^2_j)$ refers to a normal density function with mean μ_j and variance σ^2_j. The intuition for the integral in equation (33) is that, when period t belongs to a new class of normal with unknown mean and variance, we evaluate the density of ε_t by taking average of the densities for all possible values of mean and variance generated from the base distribution G_0. This integral can be evaluated by Monte Carlo simulation as suggested by West et al. (1994). [7]

By denoting $\tilde{D}_T$ as a collection of the mixing indicators (or class indicators) generated from the previous iteration of the MCMC, we can generate D_t by repeating the following steps sequentially for $t = 1, 2, ..., T$, starting with $t = 1$:

i) Count the total number of distinctive classes in $\tilde{D}_{\neq t}$ and set it as $M^*_{\neq t}$.

ii) Generate D_t according to the probabilities in equation (31), and replace the $t-$th element of $\tilde{D}_T$ with the generated D_t.

iii) If D_t is generated to be $M^*_{\neq t} + 1$, it means that period t belongs to a new class that does not exists in $\tilde{D}_{\neq t}$. In this case, we have to generate intermediate values for the mean ($\mu_{M^*_{\neq t}+1}$) and variance ($\sigma^2_{M^*_{\neq t}+1}$) that are associated with this new class. They can be generated from the following posterior distributions:

$$\sigma^2_{M^*_{\neq t}+1} \,|\varepsilon_t \;\sim IG\left(\frac{1+d_0}{2}, \frac{v_0 + (\varepsilon_t - \lambda_0)^2/(1+\psi_0)}{2}\right), \tag{34}$$

$$\mu_{M^*_{\neq t}+1}|\sigma^2_{M^*_{\neq t}+1}, \varepsilon_t \sim N\left(\frac{\lambda_0 + \psi_0\varepsilon_t}{1+\psi_0}, \frac{\psi_0}{1+\psi_0}\sigma^2_{M^*_{\neq t}+1}\right), \tag{35}$$

which can be easily derived given the joint prior G_0 for $(\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1})$ in equation (16) and a single observation ε_t.

[7] The integral in equation (33) can be approximated by

$$\int f_N(\varepsilon_t|\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1}) dG_0(\mu_{M^*_{\neq t}+1}, \sigma^2_{M^*_{\neq t}+1}) \approx \frac{1}{R}\sum_{i=1}^{R} f_N(\varepsilon_t|\mu_i, \sigma^2_i),$$

where μ_i and σ^2_i are drawn from the base distribution G_0 in equation (16) and R is large enough. Alternatively, Escobar and West (1995) analytically derive that this integral results in a density function for a scaled and shifted Student's t-distribution.

iv) In order to impose the inequality constraints for the variances in equation (11), we reorder the distinctive classes according to an ascending order of $\sigma_m^{2'}s$. That is, we assign $m = 1$ to the class with the lowest variance and $m = 2$ to the class with the second largest variance, etc.

v) Set t=t+1, and go to i).

At the end of the iteration, we have a new set of $\tilde{D}_T$. The number of distinctive classes in $\tilde{D}_T$ is the realized M or the realized total number of mixtures.

4.2.2. Drawing α conditional on $\tilde{D}_T$, and thus, on M

Drawing α conditional on $\tilde{D}_T$ is equivalent to drawing α conditional on M, the total number of mixtures or classes in the sample. [8] In this section, we explain an algorithm for generating α as proposed by Escobar and West (1995).

Given the prior distribution of α in equation (19), the prior density is:

$$f(\alpha) \propto \alpha^{a-1} \exp(-\alpha b), \tag{36}$$

and as derived by Antoniak (1974), the likelihood for M is

$$f(M|\alpha) \propto \alpha^M \frac{\Gamma(\alpha)}{\Gamma(\alpha + T)}, \tag{37}$$

where $\Gamma(.)$ refers to the Gamma function and T is the sample size. Thus, Escobar and West (1995) derive the posterior density of α as: [9]

---

[8] Note that the posterior distribution of α depends only on M, for given $\tilde{D}_T$.

[9] Note that gamma functions in equation (37) can be written as

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha + T)} = \frac{(\alpha + T)\beta(\alpha + 1, T)}{\alpha\Gamma(T)},$$

where $\beta(.,.)$ refers to the beta function, and

$$\beta(\alpha + 1, T) = \int_0^1 x^\alpha (1 - x)^{T-1} dx$$

$$
\begin{aligned}
f(\alpha|M) &\propto f(\alpha)f(M|\alpha) \\
&\propto \alpha^{a+M-2}\exp(-\alpha b)(\alpha+T)\int_0^1 x^{\alpha}(1-x)^{T-1}dx,
\end{aligned}
\tag{38}
$$

which implies that the posterior distribution of α is the marginal distribution obtained from a joint distribution of α and a continuous quantity η such that

$$
f(\alpha,\eta|M) \propto \alpha^{a+M-1}\exp(-\alpha b)(\alpha+T)\eta^{\alpha}(1-\eta)^{T-1},\ 0<\eta<1. \tag{39}
$$

As shown in Appendix B, Escobar and West (1995) further derive the conditional posterior densities $f(\eta|\alpha,M)$ and $f(\alpha|\eta,M)$, and show that

$$
\eta|\alpha, M \ \sim Beta(\alpha+1,T) \tag{40}
$$

and

$$
\alpha|\eta, M \ \sim r_\eta G(a+M, b-\ln(\eta)) + (1-r_\eta)G(a+M-1, b-\ln(\eta)), \tag{41}
$$

where the latter is a mixture of two Gamma distributions with $r_\eta/(1-r_\eta) = (a+M-1)/\{T[b-\ln(\eta)]\}$.

Thus, the following two-step algorithm can be employed to draw α:

i) Conditional on α generated in the previous iteration of the Gibbs sampling, draw an intermediate random variable η from the distribution given in equation (40).

ii) Conditional on η, draw α from the distribution given in equation (41).

5. An Application to the Growth of Postwar U.S. Industrial Production Index [1947M1-2017M1]

5.1. Specification for an Empirical Model

We consider the following univariate Markov-switching model for the growth of industrial production index (Δy_t):

$$
\begin{aligned}
\Delta y_t =& \beta_{1,t} + \beta_{2,t} 1[S_t = 2] + \Delta y_t^* \\
& \Delta y_t^* = \phi \Delta y_{t-1}^* + h_t \varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0,1), \\
& S_t = 1, 2, \\
& \beta_{2,t} < 0, \quad \forall t,
\end{aligned}
\tag{42}
$$

where 1[.] is the indicator function; $\beta_{1,t}$ is the mean growth rate during boom and $\beta_{1,t} + \beta_{2,t}$ is the mean growth rate during recession. We have a boom when $S_t = 1$ and we have a recession when $S_t = 2$. The distribution of the error term ε_t is potentially non-normal, and it is approximated by the Dirichlet process mixture of normals in equation (16).

In the above model, S_t follows a first-order Markov-switching process with the following transition probabilities as in Hamilton (1989):

$$
Pr[S_t = j | S_{t-1} = i] = p_{ij}, \quad \sum_{j=1}^{2} p_{ij} = 2, \quad i, j = 1, 2. \tag{43}
$$

It would be unreasonable to assume that the regime-specific mean growth rates during boom or recession are constant in a sample that covers the entire postwar period. While Eo and Kim (2016) propose to specify the regime-specific mean growth rates of real GDP to be random walks, we assume that there are two structural breaks in the regime-specific mean growth rates. We thus specify $\beta_{0,t}$ and $\beta_{1,t}$ as follows: [10]

$$
\begin{aligned}
\beta_{1,t} &= \beta_1 + \delta_{1,1}\ 1[S_{\beta,t} = 1] + \delta_{1,2}\ 1[S_{\beta,t} = 2], \\
\beta_{2,t} &= \beta_2 + \delta_{2,1}\ 1[S_{\beta,t} = 1] + \delta_{2,2}\ 1[S_{\beta,t} = 2], \\
& \delta_{1,1} > \delta_{1,2} > 0, \quad \delta_{2,1} < \delta_{2,2} < 0,
\end{aligned}
\tag{44}
$$

where $S_{\beta,t}$ is a latent discrete variable that evolves according to the following transition probabilities:

$$
\begin{gathered}
Pr[S_{\beta,t} = j | S_{\beta,t-1} = i] = p_{\beta,ij}, \quad \sum_{j=1}^{2} p_{\beta,ij} = 1, \quad i, j = 1, 2, 3, \\
with \ \ p_{\beta,13} = p_{\beta,21} = p_{\beta,31} = p_{\beta,32} = 0 \ and \ p_{\beta,33} = 1.
\end{gathered}
\tag{45}
$$

[10] The inequalities in equation (44) is based on Kim and Nelson (1999), who suggest that the difference between the mean growth rates during recessions and booms has been decreasing.

Note that β_1 is the mean growth rate during boom after the second structural break and $\beta_1 + \beta_2$ is the mean growth rate during recession after the second structural break.

We further assume that the volatility evolves according to a 3-state Markov-switching process, as specified below: [11]

$$\begin{aligned} h_t^2 =& h_1^2\ 1[S_{h,t} = 1] + h_2^2\ 1[S_{h,t} = 2] + h_3^2\ 1[S_{h,t} = 3], \quad h_1^2 < h_2^2 < h_3^2, \\ & Pr[S_{h,t} = j | S_{h,t-1} = i] = p_{h,ij}, \quad \sum_{j=1}^{3} p_{h,ij} = 1, \quad i, j = 1, 2, 3. \end{aligned} \tag{46}$$

Note that the model specified above implies a time-varying long-run mean growth rate, which can be estimated by:

$$\begin{aligned} g_t &= \beta_1 + \delta_{1,1} Pr[S_{\beta,t} = 1 | \Delta\tilde{Y}_T] + \delta_{1,2} Pr[S_{\beta,t} = 2 | \Delta\tilde{Y}_T] \\ &+ \Big(\beta_2 + \delta_{2,1} Pr[S_{\beta,t} = 1 | \Delta\tilde{Y}_T] + \delta_{2,2} Pr[S_{\beta,t} = 2 | \Delta\tilde{Y}_T]\Big) \times Pr[S_t = 2], \end{aligned} \tag{47}$$

where $Pr[S_t = 2]$ is the steady-state probability that $S_t = 2$, which is given below:

$$Pr[S_t = 2] = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}.$$

5.2. Priors

We estimate the model by employing the normalization introduced in *Specification #2* of Section 3, and set $\beta_1 = 0$ and $h_1^2 = 1$. We then employ the estimation procedure in Section 4 and the original parameters β_1, β_2, and h_i^2, $i = 1, 2, 3$, are recovered as discussed in Section 3. The priors that we employ are described below:

$$[\beta_2 \ \ \delta_{1,1} \ \ \delta_{1.2} \ \ \delta_{2,1} \ \ \delta_{2,2}]' \sim N\left([-0.5 \ \ 1 \ \ 0.5 \ \ -1 \ \ -0.5]', 0.5 I_5\right),$$

$$\phi \sim N(0.3, 0.5)_{1[|\phi|<1]}$$

$$\frac{h_2^2}{h_1^2} \sim IG(1, 2), \quad \frac{h_3^2}{h_1^2} \sim IG(1, 4),$$

$$p_{11} \sim Beta(8, 2), \quad p_{22} \sim Beta(8, 2),$$

[11] We assume that S_t, $S_{\beta,t}$ and $S_{h,t}$ are independent of one another.

$$p_{\beta,11} \sim Beta(24.9, 0.1), \quad p_{\beta,22} \sim Beta(19.9, 0.1),$$

$$[p_{h,11}, p_{h,12}, p_{h,13}] \sim Dirichlet(9, 0.5, 0.5),$$

$$[p_{h,21}, p_{h,22}, p_{h,23}] \sim Dirichlet(0.5, 9, 0.5),$$

$$[p_{h,31}, p_{h,32}, p_{h,33}] \sim Dirichlet(0.5, 0.5, 9),$$

$$(\mu_m, \sigma_m^2) \sim \; G_0 \; \equiv \; N(1, \sigma_m^2) IG\,(4, 1)\,, \; m = 1, 2, \ldots,$$

$$\alpha \sim Gamma(1, 2).$$

When ε_t is assumed to be normally distributed, β_1 and h_1^2 are estimated directly. In this case, we employ the following priors for these parameters:

$$\beta_1 \sim N(1, 0.5), \quad h_1^2 \sim IG(4, 1).$$

5.3. Empirical Results

Data employed are seasonally-adjusted postwar U.S. industrial production index. Data are obtained from the Federal Reserve Bank of St. Louis economic database (FRED), and the sample covers the period 1947M1-2017M1. Figure 2 depicts the data. We obtain 60,000 MCMC draws and discard the first 10,000 to avoid the effect of the initial values. All the inferences are based on the remaining 50,000 draws.

Table 2 reports the posterior moments of the parameters obtained under the normality assumption for the error term. When we performed a normality test for the error term (ε_t) for this case, however, the null was rejected at a 5% significance level. This provides a justification for employing the proposed model, in which we approximate the unknown error term with the Dirichlet process mixture of normals.

For the proposed model, the posterior mean for the total number of mixtures is slightly higher than 3, as shown in Table 3. The null hypothesis of normality is not rejected for the standardized error term estimated conditional on the mixing indicator variable. [12] 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)) strongly prefers the proposed model.

Figures 3.A and 3.B depict the posterior probabilities of recession for the two models. The shaded areas represent the NBER recessions. Estimates of turning points from the proposed model are much sharper and agree much more closely with the NBER reference cycles than the estimates from a model with normally distributed errors do.

Figure 4 depicts the time-varying volatility for the IP growth rate estimated from the proposed model. Note that we model the volatility process as a 3-state Markov-switching process. It seems that the high and the medium volatility regimes are mostly focused on the period prior to the mid 1980s. However, in most of the post-1984 period, the low volatility regime dominates except for a few episodes of high or medium volatility. Finally, Figure 5 depicts the posterior mean of the long-run mean growth rates estimated based on equation (45). It demonstrates a pattern for steadily decreasing long-run mean growth rate, which is consistent with Stock and Watson (2012) and Eo and Kim (2016).

6. Concluding Remarks

In their dynamic factor models of business cycle, Kim and Yoo (1996), Chauvet (1998), and Kim and Nelson (1998) assume that each individual coincident variable consists of an idiosyncratic component and a common factor component, which is subject to Markov switching mean. They estimate their models either by the QMLE method or by the Bayesian method, under the assumption of normally distributed shocks. They all show that their estimates of turning points are much sharper and agree much more closely with the NBER reference cycles than the estimates from a univariate Markov switching model do. The intuition is that the idiosyncratic components in these multivariate models, which consist of

[12] To calculate the Jarque-Bera test statistic for the normality test, we use the posterior mean of the standardized error term ($\frac{\varepsilon_t - \mu_{D_t}}{\sigma_{D_t}}$) from equation (11), for $t = 1, 2, ..., T$.

irregular components and outliers, are averaged out across individual series.

However, even in case the common factor component is normally distributed, the existence of irregular components and outliers in individual series makes the error term in a univariate model to deviate from normality. This is the main reason why our univariate Markov-switching model of the postwar industrial production index results in poor inferences on recession probabilities under a normality assumption. By modeling the error term as the Dirichlet process mixture of normals, we can effectively control for the irregular component that is not related to the business conditions. This leads to sharp and accurate inferences on recession probabilities just like the dynamic factor models do.

Appendix A. Derivation of Equation (15)

Given the prior for $(w_1, w_2, \ldots, w_M)$ in equation (14), the marginal distribution of w_m is

$$w_m \sim Beta(\frac{\alpha}{M}, \frac{\alpha}{M}(M-1)), \tag{A.1}$$

with the following density function:

$$f(w_m) \propto w_m^{\frac{\alpha}{M}-1}(1-w_m)^{\frac{\alpha}{M}(M-1)-1}. \tag{A.2}$$

The likelihood of $\tilde{D}_{\neq t}$ given w_m can be expressed as:

$$f(\tilde{D}_{\neq t}|w_m) \propto w_m^{T_{m,\neq t}}(1-w_m)^{T-1-T_{m,\neq t}}, \tag{A.3}$$

where $T_{m,\neq t}$ denotes the total number of observations that belong to the $m-th$ class in a sample that excludes period t.

By combining equations (A.2) and (A.3), we have:

$$\begin{aligned} f(w_m|\tilde{D}_{\neq t}) &\propto f(w_m)Pr(\tilde{D}_{\neq t}|w_m) \\ &= w_m^{T_{m,\neq t}+\frac{\alpha}{M}-1}(1-w_m)^{T-1-T_{m,\neq t}+\frac{\alpha}{M}(M-1)-1}, \end{aligned} \tag{A.4}$$

which suggests that

$$w_m|\tilde{D}_{\neq t} \;\sim\; Beta(T_{\neq t,m} + \frac{\alpha}{M}, T-1-T_{m,\neq t} + \frac{\alpha}{M}(M-1)). \tag{A.5}$$

From equation (A.5), we can derive the following probability of interest in equation (13):

$$\begin{aligned} Pr[D_t = m|\tilde{D}_{\neq t}) &= E(w_m|\tilde{D}_{\neq t}) \\ &= \frac{T_{m,\neq t} + \frac{\alpha}{M}}{T-1+\alpha}. \end{aligned} \tag{A.6}$$

Appendix B. Derivation of Equations (40) and (41)

Conditional on α, equation (39) results in

$$f(\eta|\alpha, M) \propto \eta^{\alpha}(1-\eta)^{T-1}, \quad 0 < \eta < 1, \tag{B.1}$$

which suggests that

$$\eta \mid \alpha, M \ \sim Beta(\alpha+1, T). \tag{B.2}$$

Thus, as in Escobar and West (1995), the conditional density of α given η and M can be derived as:

$$\begin{aligned}
&f(\alpha|\eta, M) \\
&\quad \propto \alpha^{a+M-1} \exp\{-\alpha[b-\ln(\eta)]\} + T\alpha^{a+M-2} \exp\{-\alpha[b-\ln(\eta)]\} \\
&\quad = \frac{\Gamma(a+M)}{[b-\ln(\eta)]^{a+M}} G(a+M, b-\ln(\eta)) + T\frac{\Gamma(a+M-1)}{[b-\ln(\eta)]^{a+M-1}} G(a+M-1, b-\ln(\eta)) \\
&\quad \propto (a+M-1)G(a+M, b-\ln(\eta)) + T[b-\ln(\eta)]G(a+M-1, b-\ln(\eta)),
\end{aligned} \tag{B.3}$$

which can be written as the following mixture of two Gamma distributions:

$$(\alpha|\eta, M) \sim r_\eta G(a+M, b-\ln(\eta)) + (1-r_\eta)G(a+M-1, b-\ln(\eta)), \tag{B.4}$$

where $r_\eta/(1-r_\eta) = (a+M-1)/\{T[b-\ln(\eta)]\}$.

References

[1] Alexander, C., E. Lazar, 2006, "Normal Mixture GARCH(1,1): Applications to Exchange Rate Modeling," *Journal of Applied Econometrics*, 21, 307-336.

[2] Antoniak, C. E., 1974, "Mixtures of Dirichlet Processes With Applications to Nonparametric Problems," *The Annals of Statistics*, 2, 1152-1174.

[3] Bauwens, L., J.-F. Carpantier, A. Dufays, 2017, "Autoregressive Moving Average Infinite Hidden Markov-Switching Models," *Journal of Business and Economic Statistics*, 35, 162-182.

[4] Chauvet, M., 1998, "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switches," *International Economic Review*, 39, 969-996.

[5] Chib, S., 1998, "Estimation and Comparison of Multiple Change-Point Models," *Journal of Econometrics*, 86, 221-241.

[6] Diebold, F. X., J.-H. Lee, and G. C. Weinbach, 1994, "Regime Switching with Time-Varying Transition Probabilities," In C. Hargreaves, Ed.: *Nonstationary Time Series Analysis and Cointegration*, 283-302. Oxford University Press.

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

[8] Eo, Y. and C-J. Kim, 2016, "Markov-Switching Models with Evolving Regime-Specific Parameters: Are Postwar Booms and Recessions All Alike?," *Review of Economics and Statistics*, 98, 940-949.

[9] Escobar, M. D., and M. West, 1995, "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577-588.

[10] Filardo, A. J., 1994, "Business Cycle Phases and Their Transitional Dynamics," *Journal of Business and Economic Statistics*, 12, 299-308.

[11] Fox, E., E. Sudderth, M. Jordan, and A. Willsky, 2011, "A Sticky HDP-HMM with

Application to Speaker Diarization," *Annals of Applied Statistics*, 5, 1020-1056.

[12] Gorur, D., and C. E. Rasmussen, 2010, "Dirichlet Process Gaussian Mixture Models: Choice of the Base Distribution," *Journal of Computer Science and Technology*, 25, 615-626.

[13] Hamilton, J. D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and The Business Cycle," *Econometrica*, 57, 357-384.

[14] Jensen, M. and J. Maheu, 2010, "Bayesian Semiparametric Stochastic Volatility Modeling," *Journal of Econometrics*, 157, 306-316.

[15] Jensen, M. and J. Maheu, 2013, "Bayesian Semiparametric Multivariate GARCH Modeling," *Journal of Econometrics*, 176, 3-17.

[16] Jensen, M, and J. Maheu, 2014, "Estimating a Semiparametric Asymmetric Stochastic Volatility Model with a Dirichlet Process Mixture," *Journal of Econometrics*, 178, 523-538.

[17] Jin, X. and J. Maheu, 2016, "Bayesian Semiparametric Modeling of Realized Covariance Matrices," *Journal of Econometrics*, 192, 19-39.

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

[19] Kim, C-J. and C. Nelson, 1998, "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on A Dynamic Factor Model with Regime-Switching," *Review of Economics and Statistics*, 80, 188-201.

[20] Kim, C-J. and C. 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, 608-616.

[21] Kim, S., N. Shephard, and S. Chib, 1998, "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361-393.

[22] Kim, M.-J. and J.-S. Yoo, 1995, "New Index of Coincident Indicators: A Multivariate

Markov Switching Factor Model Approach," *Journal of Monetary Economics*, 36, 607-630.

[23] Koop, G., 2003, *Bayesian Econometrics*, New York: Wiley

[24] Marin, J.-M., K. Mengersen, C.P. Robert, 2005, "Bayesian Modeling and Inference on Mixtures of Distributions," *Handbook of Statistics*, 25, 459-507.

[25] Neal, R. M., 2000, "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249-265.

[26] Omori, Y., S. Chib, N. Shephard, and J. Nakajima, 2007, "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, 140, 425-449.

[27] Song, Y., 2014, "Modeling Regime Switching and Structural Breaks with an Infinite Hidden Markov Model," *Journal of Applied Econometrics*, 29, 825-842.

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

[29] Watanabe, S., 2010, "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory," *Journal of Machine Learning Research*, 11, 3571-3594.

[30] West, M., P. Muller, and M. D. Escobar, 1994, "Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation," in *Aspects of Uncertainty: A Tribute to D.V. Lindley*, eds. P. R. Freeman and A. F. M. Smith, New York: Wiley, 363-386.

[31] White, H., 1994, *Estimation, Inference, and Specification Analysis*, New York: Cambridge University Press.

Table 1. Quasi Maximum Likelihood Estimation of Markov-switching Models: Monte Carlo Experiment

$T = 500$

	True	Case #1	Case #2	Case #3	Case #4
β_1	-0.5	-0.519 (0.232)	-0.360 (0.288)	-1.031 (0.634)	-1.148 (1.092)
β_2	1	0.999 (0.067)	1.044 (0.676)	1.141 (0.175)	1.324 (0.397)
h_1	2	1.990 (0.128)	2.089 (0.330)	1.982 (0.256)	1.679 (0.451)
h_2	1	1.004 (0.004)	0.891 (0.109)	0.726 (0.275)	0.495 (0.505)
p_{11}	0.9	0.900 (0.042)	0.884 (0.057)	0.651 (0.279)	0.497 (0.426)
p_{22}	0.95	0.950 (0.019)	0.933 (0.045)	0.854 (0.110)	0.702 (0.274)

$T = 5000$

	True	Case #1	Case #2	Case #3	Case #4
β_1	-0.5	-0.505 (0.069)	-0.339 (0.176)	-1.042 (0.551)	-0.907 (0.415)
β_2	1	1.000 (0.020)	0.989 (0.023)	1.148 (0.151)	1.391 (0.392)
h_1	2	1.997 (0.039)	2.081 (0.107)	1.958 (0.086)	1.707 (0.295)
h_2	1	1.000 (0.001)	0.872 (0.128)	0.718 (0.282)	0.431 (0.569)
p_{11}	0.9	0.899 (0.012)	0.882 (0.023)	0.601 (0.303)	0.462 (0.439)
p_{22}	0.95	0.950 (0.006)	0.930 (0.022)	0.833 (0.119)	0.658 (0.293)

Note:

1. This table reports quasi maximum likelihood estimation results under different error distributions. Each cell contains the average of the 1,000 point estimates for each parameter and the root mean squared error of the estimates from the true value (in parentheses).
2. Case #1: normal distribution; Case #2: t-distribution; Case #3: χ^2 distribution; Case #4: mixture of 3 normals.

Table 2. Bayesian Inference of a Model under Normality Assumption [Log Difference of the U.S. Industrial Production Index, 1947M1-2017M1]

Parameter	Mean	SD	Median	90% HPDI
β_1	0.303	0.094	0.295	[0.173,0.475]
β_2	-0.357	0.184	-0.363	[-0.660,-0.048]
$\delta_{1,1}$	0.769	0.475	0.628	[0.227,1.692]
$\delta_{1,2}$	0.318	0.204	0.278	[0.097,0.745]
$\delta_{2,1}$	-0.758	0.489	-0.732	[-1.613,0.007]
$\delta_{2,2}$	-0.497	0.367	-0.477	[-1.144,0.054]
ϕ	0.204	0.064	0.201	[0.103,0.313]
h_1	0.792	0.121	0.719	[0.680,1.002]
h_2	0.854	0.128	0.898	[0.680,1.028]
h_3	1.450	0.100	1.440	[1.305,1.628]
p_{11}	0.927	0.074	0.950	[0.763,0.977]
p_{22}	0.878	0.062	0.886	[0.771,0.958]
$p_{h,11}$	0.980	0.008	0.980	[0.965,0.991]
$p_{h,12}$	0.009	0.007	0.008	[0.001,0.023]
$p_{h,21}$	0.022	0.014	0.020	[0.004,0.049]
$p_{h,22}$	0.961	0.015	0.963	[0.933,0.982]
$p_{h,32}$	0.037	0.019	0.034	[0.009,0.073]
$p_{h,33}$	0.934	0.022	0.934	[0.895,0.966]
$p_{\beta,11}$	0.975	0.031	0.986	[0.913,0.999]
$p_{\beta,22}$	0.995	0.009	0.998	[0.982,0.999]

Table 2. (Continued).

WAIC	1775
JB	8.119 (0.017)

Note:

1. Out of 60,000 MCMC draws, the first 10,000 are discarded and inferences are based on the remaining 50,000 draws.
2. SD refers to standard deviation.
3. HPDI refers to a highest posterior density interval.
4. WAIC refers to the Watanabe-Akaike Information Criterion.
5. JB refers to the Jarque-Bera test statistic for a normality test. In the parenthesis is the p-value.

Table 3. Bayesian Inference of a Model with Unknown Error Distribution [Log Difference of the U.S. Industrial Production Index, 1947M1-2017M1]

Parameter	Mean	SD	Median	90% HPDI
β_1	0.196	0.068	0.199	[0.078,0.295]
β_2	-0.537	0.232	-0.565	[-0.878,-0.117]
$\delta_{1,1}$	0.528	0.396	0.405	[0.240,1.552]
$\delta_{1,2}$	0.188	0.080	0.180	[0.076,0.326]
$\delta_{2,1}$	-1.075	0.291	-1.055	[-1.566,-0.638]
$\delta_{2,2}$	-0.468	0.249	-0.456	[-0.903,-0.084]
ϕ	0.105	0.043	0.104	[0.034,0.176]
h_1	0.454	0.024	0.453	[0.417,0.494]
h_2	0.745	0.055	0.740	[0.662,0.841]
h_3	2.097	0.278	2.067	[1.697,2.590]
p_{11}	0.964	0.011	0.966	[0.944,0.980]
p_{22}	0.853	0.038	0.857	[0.785,0.909]
$p_{h,11}$	0.982	0.008	0.983	[0.967,0.994]
$p_{h,12}$	0.007	0.006	0.005	[0.001,0.019]
$p_{h,21}$	0.017	0.012	0.015	[0.003,0.039]
$p_{h,22}$	0.971	0.012	0.972	[0.948,0.987]
$p_{h,32}$	0.045	0.021	0.043	[0.015,0.084]
$p_{h,33}$	0.931	0.023	0.933	[0.891,0.964]
$p_{\beta,11}$	0.989	0.020	0.996	[0.949,0.999]
$p_{\beta,22}$	0.998	0.003	0.999	[0.993,0.999]

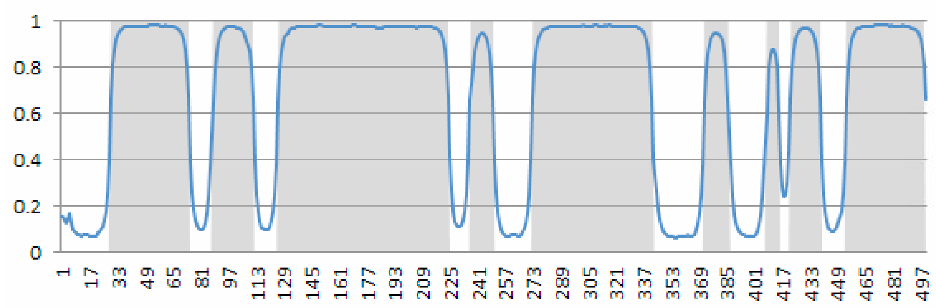
Table 3. (Continued).

M	3.007 (1.329)
WAIC	1671
JB	1.218 (0.544)
Acceptance Probability 1	0.315
Acceptance Probability 2	0.474

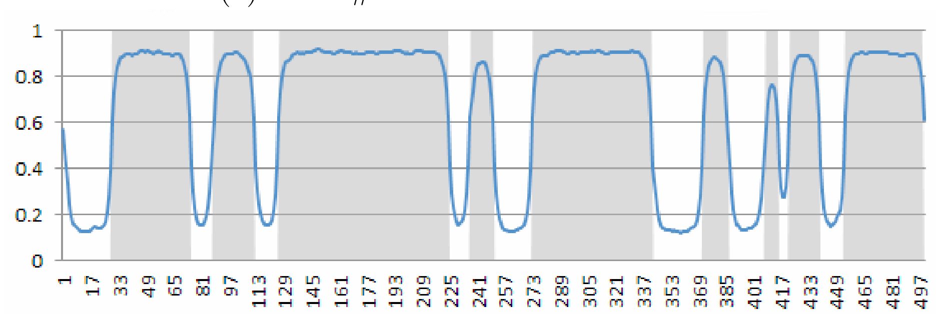
Notes:

1. Out of 60,000 MCMC draws, the first 10,000 are discarded and inferences are based on the remaining 50,000 draws.
2. SD refers to standard deviation.
3. HPDI refers to a highest posterior density interval.
4. WAIC refers to the Watanabe-Akaike Information Criterion.
5. JB refers to the Jarque-Bera test statistic for a normality test. In the parenthesis is the p-value.
6. Acceptance Probability 1 refers to the acceptance probability of Metropolis Hasting algorithm for h_1^2; Acceptance Probability 2 refers to the acceptance probability of the Metropolis Hasting algorithm for h_2^2.

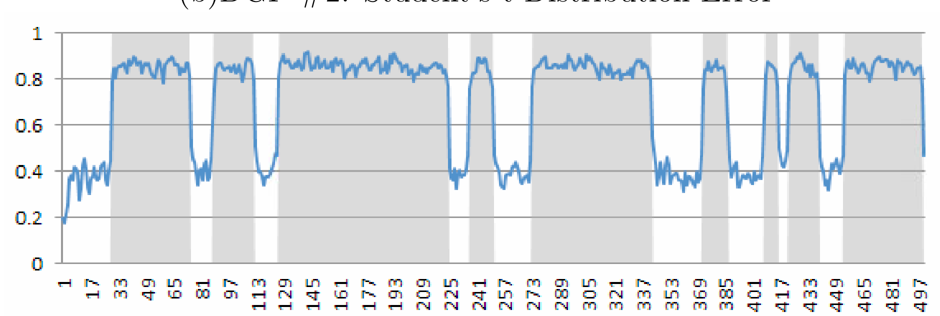
Figure 1. Smoothed Probabilities of Regime 2 based on Quasi-Maximum Likelihood Estimation under Different Error Distributions [T=500].



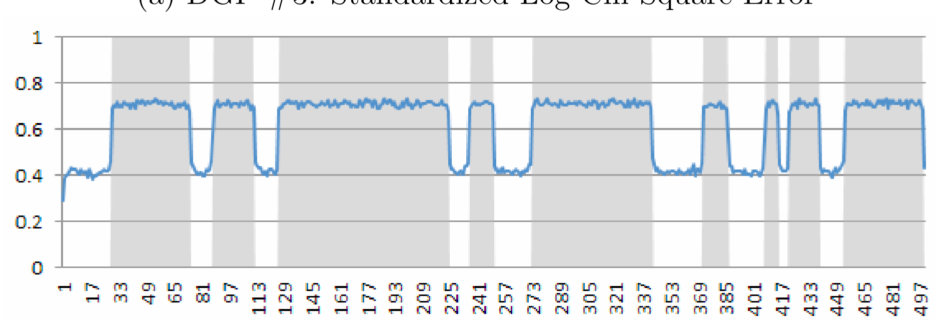
(a) DGP #1: Standard Normal Error



(b)DGP #2: Student's t Distribution Error



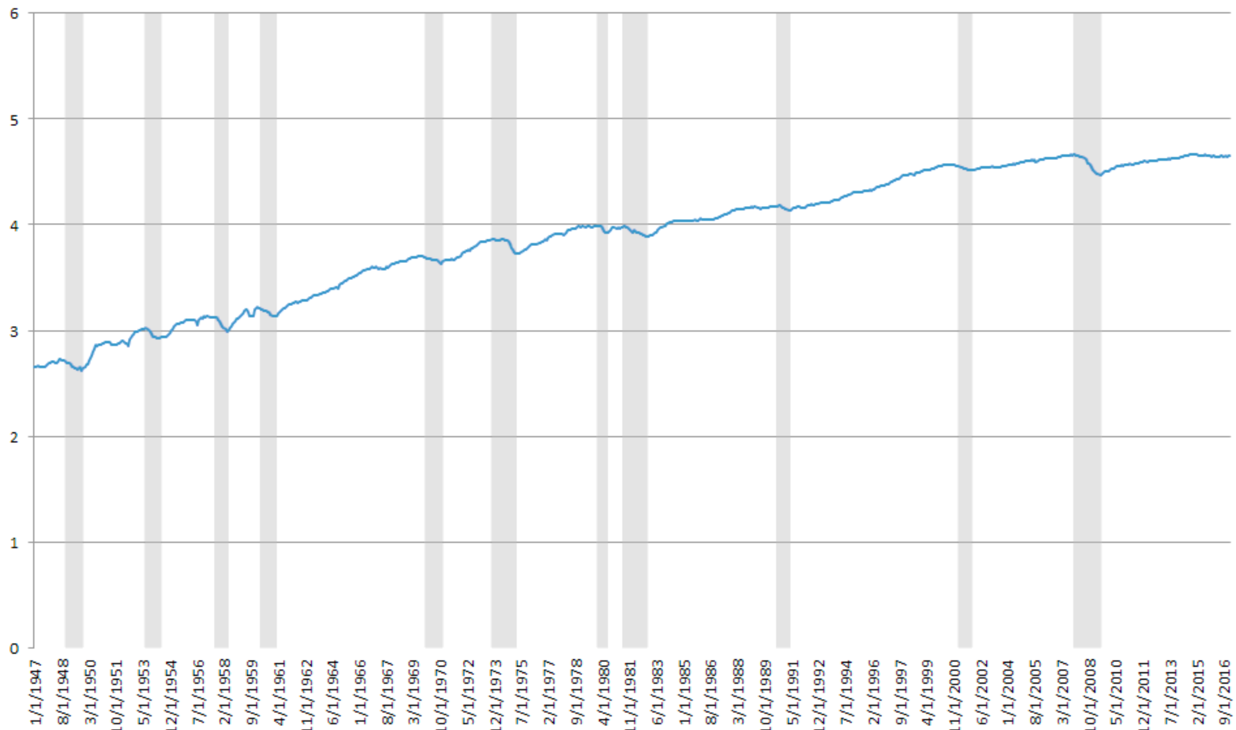
(a) DGP #3: Standardized Log Chi-Square Error



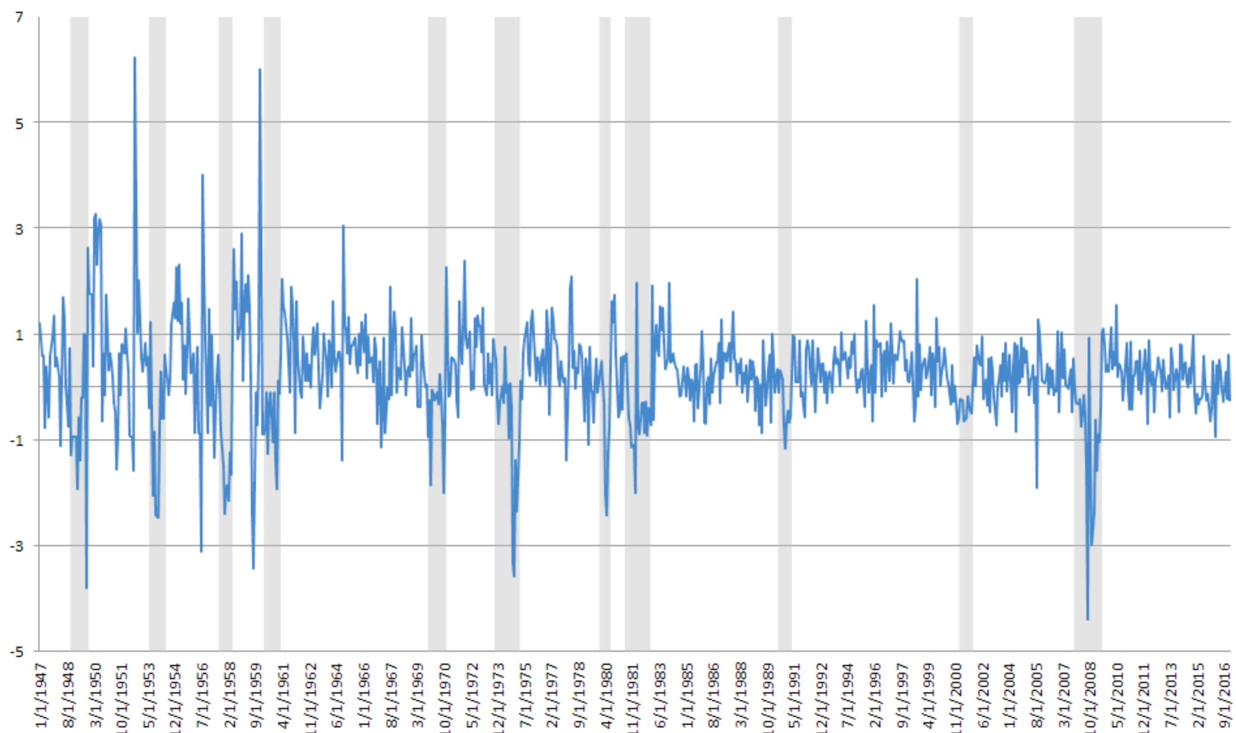
(b)DGP #4: Mixture Normal Error

Note: The shaded area denotes the data periods associated with regime 2.

Figure 2. U.S. Industrial Production (IP) Index and Its Growth Rate [1947M1 - 2017M1]



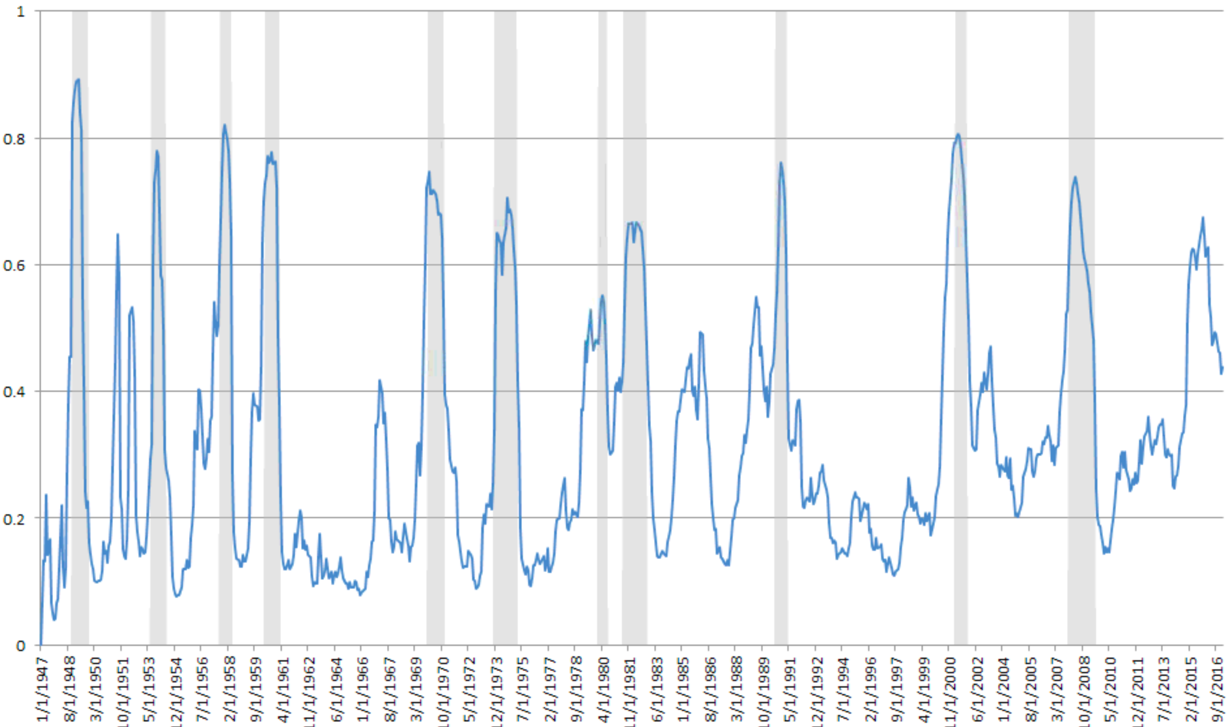
(a) Logarithm of U.S. Industrial Production (IP)



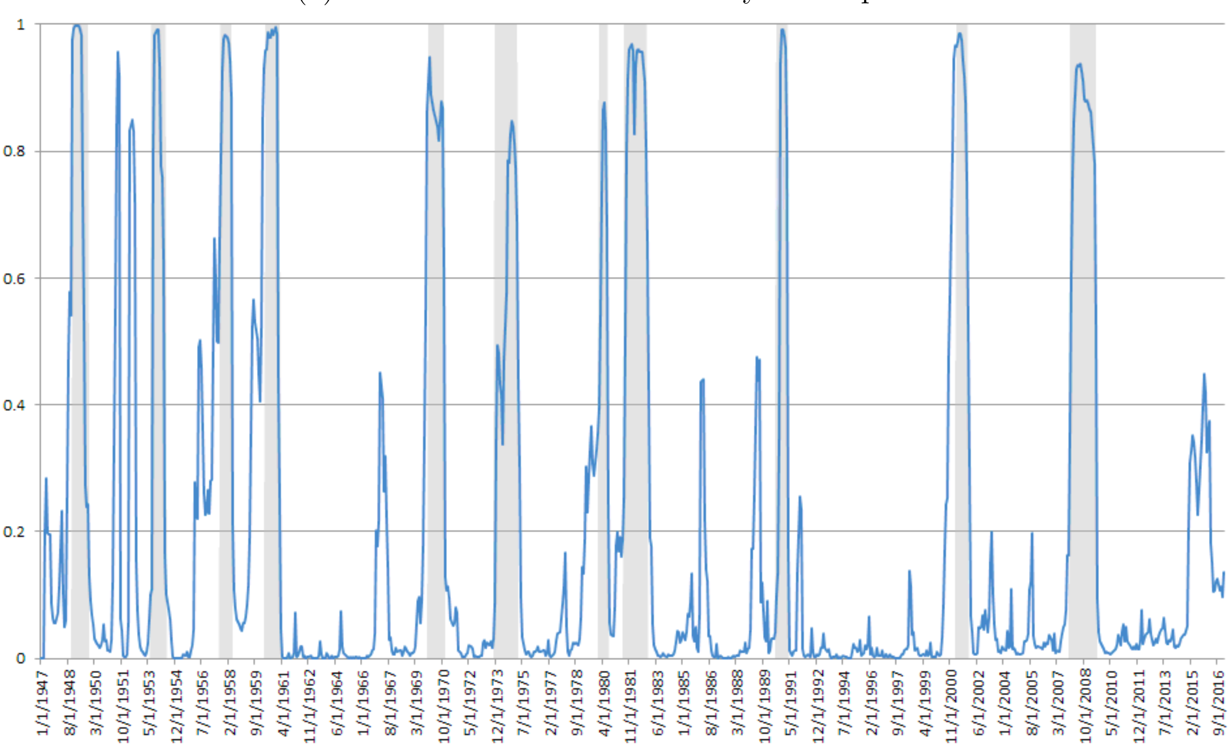
(b) IP growth

Note: The shaded area denotes the NBER recession date.

Figure 3. Posterior Probabilities of Recession: Model under Normality vs. Proposed Model.



(a) For the Model with Normality Assumption.



(b) For the Model with Dirichlet Process Mixture of Normals.

Figure 4. Time-Varying Volatility for the IP Series: Proposed Model.

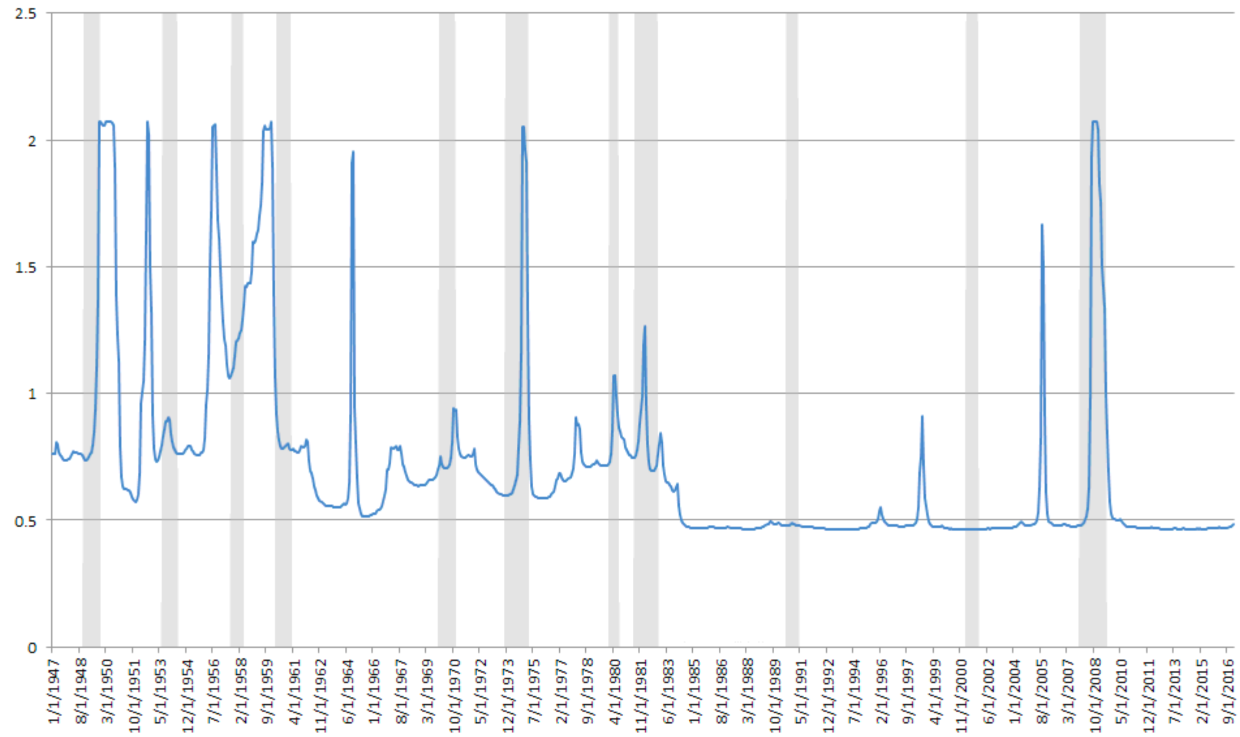


Figure 5. Time-varying Long-Run Mean Growth Rate: Proposed Model.