

# Non-Markovian Regime-Switching Models

by

**Chang-Jin Kim**

*University of Washington*

and

**Jaeho Kim**<sup>1</sup>

*University of Oklahoma*

March, 2018

## Abstract

This paper revisits the non-Markovian regime switching model considered by Chib and Dueker (2004), who employ an autoregressive continuous latent variable in order to specify the dynamics of the latent regime-indicator variable. We show that, in spite of the non-Markovian nature of the regime indicator variable, the Markovian property of this continuous latent variable allows us to easily estimate the model within the Bayesian framework without any approximations. In particular, we show that the conventional Gibbs sampling is enough in generating the regime indicator variable as well as the continuous latent variable conditional on all the parameters of the model and data. For an application to business cycle modeling of postwar US real GDP, a modified version of Hamilton's (1998) Markovian switching model is slightly preferred to a non-Markovian switching model by a Bayesian model selection criterion. For an application to volatility modeling of the weekly stock return, a non-Markovian switching model with endogenous switching or the leverage effect is strongly preferred to Markovian switching models.

**Key Words:** Non-Markovian Regime Switching, Markovian Regime Switching, Exogenous Switching, Endogenous Switching.

---

<sup>1</sup> **Chang-Jin Kim:** Dept. of Economics, University of Washington, Seattle, WA (E-mail: changjin@u.washington.edu); **Jaeho Kim:** Department of Economics, University of Oklahoma, Norman, OK (E-mail: jaeho@ou.edu). Kim acknowledges financial support from the Bryan C. Cressey Professorship at the University of Washington.

## 1. Introduction

Since Hamilton (1989), almost all extensions and applications of regime switching models have been based on the assumption that the latent regime-indicator variable follows a Markov-switching process.<sup>2</sup> Moreover, the literature has mostly focused on a first-order Markov-switching process with some exceptions that include Hering et al. (2015), who estimate a regime-switching vector autoregressive model with a second-order Markov-switching process; Neale et al. (2016), who estimate a second-order Markov-switching model with time varying transition probabilities; and Siu et al. (2009), who consider higher-order Markov switching processes for modeling risk management.

In the meantime, for binary time series in which the regime-indicator variable is observed, general p-th order Markov models have been investigated by researchers such as Zeger and Qaqish(1988), Raftery (1985), and Li (1994). These binary time series models have been further extended to the case of ARMA models by Startz (2008). Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) consider estimation of non-Markovian binary processes, by introducing a Probit model in which the latent continuous variable follows an autoregressive process.

To date, however, little attention has been paid to regime switching models with non-Markovian latent regime indicator variables, except in Chib and Dueker (2004). We revisit the following version of non-Markovian regime switching model considered by them:

$$\begin{aligned}
 y_t &= x_t' \beta_{S_t} + \sigma_{S_t} \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N(0, 1), \quad S_t = \{0, 1\}, \\
 \beta_{S_t} &= \beta_0 1[S_t^* < 0] + \beta_1 1[S_t^* \geq 0], \\
 \sigma_{S_t}^2 &= \sigma_0^2 1[S_t^* < 0] + \sigma_1^2 1[S_t^* \geq 0],
 \end{aligned} \tag{1}$$

where  $1[\cdot]$  is the indicator function and the continuous latent variable  $S_t^*$  evolves according to the following autoregressive process:<sup>3</sup>

---

<sup>2</sup> For an overview of econometric analysis of time series that are subject to changes in regime, readers are referred to Hamilton (2016).

<sup>3</sup> When equation (2) below is replaced by

$$S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t, \quad \omega_t \sim i.i.d.N(0, 1), \tag{2'}$$

we have a first-order Markov switching model with endogenous switching as proposed by Kim et al. (2008).

$$S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t, \quad \omega_t \sim i.i.d.N(0, 1). \quad (2)$$

where  $|\psi| < 1$  and  $\alpha = E(S_t^*)$ . The joint distribution of  $\varepsilon_t$  and  $\omega_t$  is specified as:

$$\begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} \sim i.i.d.N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right). \quad (3)$$

As discussed in Chib and Dueker (2004), the distinction between the above model and a first-order Markov switching model can be best explained by a business cycle model in which boom or recession is represented by a particular realization of  $S_t \in 0, 1$ . For a first-order Markov-switching model, for example, conditional on knowing that last period was a recession, no other past information is relevant in predicting the business condition this period. For the above non-Markovian regime switching model, however, the severity of recession, which is determined by the level of  $S_{t-1}^*$ , carry additional information in predicting the current business condition. That is, the discrete latent variable  $S_t$  generated by equations (2) and (3) depends not only on the sign of  $S_{t-1}^*$  but also on the level of  $S_{t-1}^*$ .

Note that the continuous latent variable  $S_{t-1}^*$  in equation (2) carries all the information on the past regimes that is hidden in the data, and this makes the regime indicator variable  $S_t$  a non-Markovian process. For making inferences on  $S_t$  from the above model, a key would be in appropriately integrating out  $S_{t-1}^*$  from the joint distribution for  $S_t$  and  $S_{t-1}^*$  conditional on the parameters of the model and data.<sup>4</sup> Here, integrating out  $S_{t-1}^*$  would be equivalent to integrating out all the past regime indicator variables. This is why maximum likelihood estimation of a non-Markovian switching model may be infeasible.

To estimate the model, Chib and Dueker (2004) cast equations (1), (2) and (3) into a state-space model, in which equation (1) serves as a measurement equation and equation (2) serves as a transition equation. As the indicator function ( $1[S_t^* > 0]$ ) in equation (1) is a nonlinear function of  $S_t^*$ , they employ linear approximation. They use the latent variables  $S_t^*$ ,  $t = 1, 2, \dots, T$ , drawn by employing the extended Kalman filter as candidate values. Then, they subject these candidate values to a Metropolis-Hastings step where the candidate draws

---

<sup>4</sup> For making inferences on  $S_t$  from a conventional first-order Markov switching model, a key is in integrating out  $S_{t-1}$  from the joint distribution of  $S_t$  and  $S_{t-1}$  conditional on the parameters of the model and data. This is straightforward as  $S_{t-1}$  is discrete.

can be rejected. However, the nature of this approximation is unknown and the resulting prediction error for  $y_t$  obtained from the extended Kalman filter is non-normal, while the Kalman filter recursion is valid under the assumption of normality.

Chang et al. (2017) consider maximum likelihood estimation of a similar model, in which the contemporaneous correlation in equation (3) is replaced by  $E(\varepsilon_{t-1}\omega_t) = \rho$ . However, their estimation procedure is derived under a situation in which the model is observationally equivalent to a first-order Markov-switching model when  $\rho = 0$ . That is, when  $\rho = 0$ , the model estimated by their algorithm results in exactly the same likelihood value as does a first-order Markov-switching model. Their focus is on effective estimation of the  $\rho$  parameter in a regime switching model with endogenous switching.

In this paper, we derive a Markov Chain Monte Carlo (MCMC) procedure for estimating non-Markovian switching models without resorting to approximations. We take advantage of the Markovian property of the latent variable  $S_t^*$  in equation (2) and show that the conventional Gibbs sampling is enough in generating latent variables  $S_t$  and  $S_t^*$  conditional on all the parameters of the model. Once the  $S_t$  and  $S_t^*$  variables are generated, generating the parameters of the model is standard. We apply the non-Markovian switching models and the proposed algorithms to the business cycle modeling of postwar real GDP and the volatility modeling of weekly stock returns.

The rest of the paper is organized as follows. Section 2 provides an algorithm for estimating a non-Markovian exogenous switching model with  $\rho = 0$ . In Section 3, the algorithm in Section 2 is extended to the case of endogenous switching with  $\rho \neq 0$ . Section 4 provides a simulation study in order to show that the proposed algorithm is working well. Pitfalls of estimating a non-Markovian switching process by a Markovian switching model are also discussed. Section 5 deals with applications, and Section 6 concludes.

## 2. Bayesian Inference of a Non-Markovian Exogenous Switching Model: A Preliminary

The non-Markovian nature of  $S_t$  in the model can be understood by rewriting equation (2) as

$$S_t^* = \alpha + \omega_t + \psi\omega_{t-1} + \psi^2\omega_{t-2} + \dots, \quad (4)$$

where, due to equation (2), the discrete variable  $S_{t-j}$  is positively related to  $\omega_{t-j}$  for  $j = 0, 1, 2, \dots$ . Equation (4) therefore suggests that  $S_t^*$ , and thus,  $S_t$  is a function of all the past history of regimes. That is, we no longer have the Markovian property for  $S_t$ , as

$$\begin{aligned} f(S_t|S_{t-1}^*) &= f(S_t|S_{t-1}, S_{t-2}, S_{t-3}, \dots, S_0) \\ &\neq f(S_t|S_{t-1}, S_{t-2}, \dots, S_{t-p}), \end{aligned} \quad (5)$$

where  $p$  is finite.

Intractability of the maximum likelihood estimation for the above model can be easily seen by considering the likelihood function, given below for the case of  $\rho = 0$ :

$$\begin{aligned} L &= \prod_{t=1}^T f(y_t|I_{t-1}) \\ &= \prod_{t=1}^T \left( \sum_{S_t} f(y_t|S_t, I_{t-1}) f(S_t|I_{t-1}) \right), \end{aligned} \quad (6)$$

where  $I_{t-1}$  refers to information up to  $t - 1$ ; and

$$\begin{aligned} f(S_t|I_{t-1}) &= \int f(S_t, S_{t-1}^*|I_{t-1}) dS_{t-1}^* \\ &= \int f(S_t|S_{t-1}^*) f(S_{t-1}^*|I_{t-1}) dS_{t-1}^* \\ &= \sum_{S_{t-1}} \sum_{S_{t-2}} \dots \sum_{S_0} f(S_t|S_{t-1}, S_{t-2}, \dots, S_0) f(S_{t-1}, S_{t-2}, \dots, S_0|I_{t-1}). \end{aligned} \quad (7)$$

Due to the dependence of current regime ( $S_t$ ) on all the history of past regimes ( $S_{t-1}, S_{t-2}, \dots, S_0$ ), the likelihood function is a function of the entire history of the unobserved regime indicator variables ( $S_0, S_1, S_2, \dots, S_T$ ).<sup>5</sup>

---

<sup>5</sup> When  $S_t$  follows a first-order Markov process, for which latent variable  $S_t^*$  can be specified as  $S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t$ ,  $\omega_t \sim i.i.d.N(0, 1)$ , equation (7) collapses to

$$\begin{aligned} f(S_t|I_{t-1}) &= \sum_{S_{t-1}} f(S_t, S_{t-1}|I_{t-1}) \\ &= \sum_{S_{t-1}} f(S_t|S_{t-1}) f(S_{t-1}|I_{t-1}), \end{aligned} \quad (7')$$

where  $f(S_{t-1}|I_{t-1})$  can be evaluated recursively, and thus, the maximum likelihood estimation of the model is feasible as in Hamilton (1989).

In this section, we consider Bayesian inference of a non-Markovian exogenous switching model with  $\rho = 0$  in equation (3). We note that, due to the Markovian property of  $S_t^*$ , we can employ the Gibbs sampling approach for drawing  $\tilde{S}_T = [S_0 \ S_1 \ S_2 \ \dots \ S_T]'$  and  $\tilde{S}_T^* = [S_0^* \ S_1^* \ S_2^* \ \dots \ S_T^*]'$ .

To get an insight into how the Gibbs sampling approach can be implemented for an exogenous switching model, we first consider the following decomposition for the joint posterior density of  $\tilde{S}_T$ ,  $\tilde{S}_T^*$ , and the parameters of the model:

$$\begin{aligned} f(\tilde{\beta}, \tilde{\sigma}^2, \tilde{A}, \tilde{S}_T^*, \tilde{S}_T | \tilde{Y}_T) &= f(\tilde{\beta}, \tilde{\sigma}^2 | \tilde{A}, \tilde{S}_T^*, \tilde{S}_T, \tilde{Y}_T) f(\tilde{A} | \tilde{S}_T^*, \tilde{S}_T, \tilde{Y}_T) f(\tilde{S}_T^*, \tilde{S}_T | \tilde{Y}_T) \\ &= f(\tilde{\beta}, \tilde{\sigma}^2 | \tilde{S}_T, \tilde{Y}_T) f(\tilde{A} | \tilde{S}_T^*) f(\tilde{S}_T^*, \tilde{S}_T | \tilde{Y}_T), \end{aligned} \quad (8)$$

where  $\tilde{\beta} = [\beta_0 \ \beta_1]'$ ;  $\tilde{\sigma}^2 = [\sigma_0^2 \ \sigma_1^2]'$ ;  $\tilde{A} = [\alpha \ \psi]'$ ; and  $\tilde{Y}_T = [y_1 \ y_2 \ \dots \ y_T]'$ .

Here, a key to Bayesian estimation of a non-Markovian switching model is to apply the single-move Gibbs sampling to the last term in equation (8), i.e., to draw  $S_t$  and  $S_t^*$ ,  $t = 1, 2, \dots, T$ , from the following full conditional distribution:

$$f(S_t^*, S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) = f(S_t^* | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T) f(S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T), \quad t = 1, 2, \dots, T, \quad (9)$$

where  $\tilde{S}_{\neq t}$  refers to  $\tilde{S}_T$  with an exclusion of  $S_t$  and  $\tilde{S}_{\neq t}^*$  refers to  $\tilde{S}_T^*$  with an exclusion of  $S_t^*$ . The purpose of the above decomposition is to take advantage of the Markovian property of  $S_t^*$ .<sup>6</sup>

The decompositions in equations (8) and (9) lead to the following steps for the MCMC procedure, which can be repeated until convergence is achieved.

**Step 1:** Generate  $S_t$  and  $S_t^*$  conditional on all the parameters of the model,  $\tilde{S}_{\neq t}^*$ ,  $\tilde{S}_{\neq t}$  and  $\tilde{Y}_T$ , for  $t = 1, 2, \dots, T$ . For each  $t$ , we generate  $S_t$  and  $S_t^*$  sequentially, in the following way:

Step 1.A: Generate  $S_t$  from  $f(S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T)$ . Then, replace the  $t$ -th row of  $\tilde{S}_T$  by the generated  $S_t$ .

---

<sup>6</sup> Hereafter, we suppress the model parameters in the conditional densities associated with  $S_t$  and/or  $S_t^*$  for simplifying the notation.

Step 1.B: Generate  $S_t^*$  from  $f(S_t^*|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T)$ , where  $S_t$  is generated in Step 1.A. Then, replace the  $t$ -th row of  $\tilde{S}_T^*$  by the generated  $S_t^*$ .

**Step 2:** Generate  $\tilde{A}$  conditional on  $\tilde{S}_T^*$ .

**Step 3:** Generate  $\tilde{\beta}$  and  $\tilde{\sigma}^2$  conditional on  $\tilde{S}_T$  and  $\tilde{Y}_T$ .

In what follows, we focus on deriving the full conditional distributions from which  $S_t$  and  $S_t^*$  can be drawn. The derivation of the full conditional distribution for Step 2 or 3 based on equation (1) or (2) is standard.

### 2.1. Generating $S_t$ conditional on $\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{A}, \tilde{\beta}, \tilde{\sigma}^2$ and $\tilde{Y}_T$

Conditional on  $S_{t-1}^*$ , all the other past information is irrelevant in making inference about  $S_t$  or about the sign of  $S_t^*$ , due to the Markovian nature of  $S_t^*$ . Likewise, conditional on  $S_{t+1}^*$ , all the other future information is irrelevant in making inferences about  $S_t$  or about the sign of  $S_t^*$ . Keeping these in mind, consider the following derivation for the joint density of  $S_t$  and  $S_t^*$  conditional on  $\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}$ , and data  $\tilde{Y}_T$ :

$$\begin{aligned}
f(S_t^*, S_t|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) &= f(S_t^*, S_t|S_{t-1}^*, S_{t+1}^*, y_t) \\
&\propto f(y_t, S_{t+1}^*, S_t^*, S_t|S_{t-1}^*) \\
&= f(y_t|S_t)f(S_{t+1}^*|S_t^*)f(S_t^*, S_t|S_{t-1}^*) \\
&= f(y_t|S_t)f(S_{t+1}^*|S_t^*)f(S_t^*|S_{t-1}^*)f(S_t|S_t^*) \\
&\propto \left[ \frac{1}{\sigma_{S_t}} \phi \left( \frac{y_t - x_t' \beta_{S_t}}{\sigma_{S_t}} \right) \right] \left[ \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{\sqrt{V}} \right) \right] f(S_t|S_t^*),
\end{aligned} \tag{10}$$

where  $\phi(\cdot)$  is the pdf of the standard normal distribution and  $f(S_t|S_t^*) = 1[S_t^* \geq 0]S_t + 1[S_t^* < 0](1 - S_t)$ , with  $1[\cdot]$  denoting the indicator function;

$$V = \frac{1}{1 + \psi^2}; \text{ and } \mu_t = \alpha + \frac{1}{1 + \psi^2}(\psi(S_{t+1}^* - \alpha) + \psi(S_{t-1}^* - \alpha)). \tag{11}$$

The second term in the last line in equation (10) is obtained from the following derivation for the  $f(S_{t+1}^*|S_t^*)f(S_t^*|S_{t-1}^*)$  term in the fourth line of equation (10) conditional on  $S_{t-1}^*$  and  $S_{t+1}^*$ :

$$\begin{aligned}
f(S_{t+1}^*|S_t^*)f(S_t^*|S_{t-1}^*) &\propto \exp \left[ -\frac{1}{2} \left( (\eta_{t+1} - \psi\eta_t)^2 + (\eta_t - \psi\eta_{t-1})^2 \right) \right] \\
&\propto \exp \left[ -\frac{1}{2\frac{1}{1+\psi^2}} \left( \eta_t - \frac{1}{1+\psi^2}(\psi\eta_{t+1} + \psi\eta_{t-1}) \right)^2 \right], \tag{12}
\end{aligned}$$

where  $\eta_t = S_t^* - \alpha$ .

Finally, by integrating  $S_t^*$  out of equation (10) we obtain the following results:

$$\begin{aligned}
f(S_t = 0|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) &\propto \int_{-\infty}^{\infty} \left[ \frac{1}{\sigma_0} \phi \left( \frac{y_t - x_t' \beta_0}{\sigma_0} \right) \right] \left[ \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{\sqrt{V}} \right) \right] f(S_t = 0|S_t^*) dS_t^* \\
&= \int_{-\infty}^0 \left[ \frac{1}{\sigma_0} \phi \left( \frac{y_t - x_t' \beta_0}{\sigma_0} \right) \right] \left[ \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{\sqrt{V}} \right) \right] dS_t^* \\
&= \frac{1}{\sigma_0} \phi \left( \frac{y_t - x_t' \beta_0}{\sigma_0} \right) \Phi \left( -\frac{\mu_t}{\sqrt{V}} \right), \tag{13}
\end{aligned}$$

$$\begin{aligned}
f(S_t = 1|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) &\propto \int_{-\infty}^{\infty} \left[ \frac{1}{\sigma_1} \phi \left( \frac{y_t - x_t' \beta_1}{\sigma_1} \right) \right] \left[ \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{\sqrt{V}} \right) \right] f(S_t = 1|S_t^*) dS_t^* \\
&= \int_0^{\infty} \left[ \frac{1}{\sigma_1} \phi \left( \frac{y_t - x_t' \beta_1}{\sigma_1} \right) \right] \left[ \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{\sqrt{V}} \right) \right] dS_t^* \\
&= \frac{1}{\sigma_1} \phi \left( \frac{y_t - x_t' \beta_1}{\sigma_1} \right) \Phi \left( \frac{\mu_t}{\sqrt{V}} \right), \tag{14}
\end{aligned}$$

where  $\Phi(\cdot)$  is the CDF of the standard normal distribution. Thus, we can generate  $S_t$  based on the following probabilities:

$$P(S_t = i|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) = \frac{f(S_t = i|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T)}{f(S_t = 0|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) + f(S_t = 1|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T)}, \quad i = 0, 1. \tag{15}$$

## 2.2. Generating $S_t^*$ conditional $\tilde{S}_{\neq t}^*$ , $S_t$ , $\tilde{S}_{\neq t}$ , $\tilde{A}$ , $\tilde{\beta}$ , $\tilde{\sigma}^2$ , and $\tilde{Y}_T$

The full conditional density  $f(S_t^*|\tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T)$ , from which  $S_t^*$  is to be drawn, can be derived based on equations (10). As the first term on the last line of equation (10) is a part of the normalizing constant conditional on  $S_t$ , we have the following results:



$$\begin{aligned}
f(S_t^* | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T) &\propto f(S_t^*, S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) \\
&\propto \frac{1}{\sqrt{V}} \phi\left(\frac{S_t^* - \mu_t}{\sqrt{V}}\right) f(S_t | S_t^*),
\end{aligned} \tag{16}$$

which suggests that we can generate  $S_t^*$  from the following truncated normal distribution:

$$S_t^* | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T \sim N(\mu_t, V)_{(1[S_t^* \geq 0]S_t + 1[S_t^* < 0](1 - S_t))}, \tag{17}$$

where  $\mu_t$  and  $V$  are given in equation (11).

### 3. Bayesian Inference of a Non-Markovian Endogenous Switching Model

In this section, we consider Bayesian inference of a non-Markovian endogenous switching model, in which  $\rho \neq 0$  in equation (3). A key to appropriate derivation of the MCMC algorithm lies in the fact that we can rewrite  $\omega_t$  as a function  $\varepsilon_t$  (i.e.,  $\omega_t = \rho\varepsilon_t + \sqrt{(1 - \rho^2)}\omega_t^*$ ), so that we can rewrite equation (2) as:

$$S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \rho\varepsilon_t + \sqrt{1 - \rho^2}\omega_t^*, \quad \omega_t^* \sim i.i.d.N(0, 1), \tag{18}$$

where  $E(\omega_t^* \varepsilon_t) = 0$  and  $\varepsilon_t = \frac{y_t - x_t' \beta_{S_t}}{\sigma_{S_t}}$ .

We note that, unlike in the case of exogenous regime switching, derivation of the full conditionals for  $\tilde{A} = [\alpha \quad \psi \quad \rho]'$ ,  $\tilde{\beta} = [\beta_0 \quad \beta_1]'$ , and  $\tilde{\sigma}^2 = [\sigma_0^2 \quad \sigma_1^2]'$  in Steps 2 and 3 are not standard any more. We therefore explain each step in detail in what follows.

#### 3.1. Generating $S_t$ conditional on $\tilde{S}_{\neq t}^*$ , $\tilde{S}_{\neq t}$ , $\tilde{A}$ , $\tilde{\beta}$ , $\tilde{\sigma}^2$ , and $\tilde{Y}_T$

In deriving the full conditional density  $f(S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T)$ , we first derive the joint density  $f(S_t^*, S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T)$ , and then  $S_t^*$  is integrated out of this joint density.<sup>7</sup> For an exogenous switching model,  $y_{t+1}$  and  $S_{t+1}$  are irrelevant in deriving this joint density. However, equation (18) suggests that  $S_{t+1}^*$  is a function of  $\varepsilon_{t+1}$ , which is a function of  $y_{t+1}$  and  $S_{t+1}$ , as well

---

<sup>7</sup> All the densities in Sections 3.1 and 3.2 are conditional on  $\tilde{A}$ ,  $\tilde{\beta}$ , and  $\tilde{\sigma}^2$ . However, we suppress them for the sake of notational brevity.

as  $S_t^*$ . Thus, unlike in the case of an exogenous switching model,  $y_{t+1}$  and  $S_{t+1}$  terms play important roles when deriving the joint density  $f(S_t^*, S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T)$  in the case of endogenous switching. Keeping this in mind, let us consider the following derivation:

$$\begin{aligned}
& f(S_t^*, S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) \\
&= f(S_t^*, S_t | S_{t-1}^*, S_{t+1}^*, S_{t+1}, y_{t+1}, y_t) \\
&\propto f(S_{t+1}^*, S_{t+1}, y_{t+1}, S_t^*, S_t, y_t | S_{t-1}^*) \\
&= f(S_{t+1}^*, S_{t+1}, y_{t+1} | S_t^*) f(S_t^*, S_t, y_t | S_{t-1}^*) \\
&= f(S_{t+1}^* | S_t^*, S_{t+1}, y_{t+1}) f(S_{t+1}, y_{t+1} | S_t^*) f(S_t^* | S_{t-1}^*, S_t, y_t) f(S_t, y_t | S_{t-1}^*) \\
&= f(S_{t+1}^* | S_t^*, S_{t+1}, \varepsilon_{t+1}) f(y_{t+1}, S_{t+1} | S_t^*) f(S_t^* | S_{t-1}^*, S_t, \varepsilon_t) f(y_t, S_t | S_{t-1}^*),
\end{aligned} \tag{19}$$

where the first and the third terms in the last line are given as

$$f(S_{t+j}^* | S_{t+j-1}^*, S_{t+j}, \varepsilon_{t+j}) = \frac{f(S_{t+j}^* | S_{t+j-1}^*, \varepsilon_{t+j}) f(S_{t+j} | S_{t+j}^*)}{f(S_{t+j} | S_{t+j-1}^*, \varepsilon_{t+j})}, \quad j = 0, 1; \tag{20}$$

and, as shown in Appendix A, product of the second and the fourth terms in the last line can be derived as

$$f(y_{t+j}, S_{t+j} | S_{t+j-1}^*) = \frac{1}{\sigma_{S_{t+j}}} \phi \left( \frac{y_{t+j} - x'_{t+j} \beta_{S_{t+j}}}{\sigma_{S_{t+j}}} \right) f(S_{t+j} | S_{t+j-1}^*, \varepsilon_{t+j}), \quad j = 0, 1. \tag{21}$$

By substituting equations (20) and (21) into equation (19), we have the following intermediate result:

$$f(S_t, S_t^* | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) \propto \frac{1}{\sigma_{S_t}} \phi \left( \frac{y_t - x'_t \beta_{S_t}}{\sigma_{S_t}} \right) f(S_{t+1}^* | S_t^*, \varepsilon_{t+1}) f(S_t^* | S_{t-1}^*, \varepsilon_t) f(S_t | S_t^*). \tag{22}$$

Here, as shown in Appendix B,  $f(S_{t+1}^* | S_t^*, \varepsilon_{t+1}) f(S_t^* | S_{t-1}^*, \varepsilon_t)$  term can be derived as:

$$f(S_{t+1}^* | S_t^*, \varepsilon_{t+1}) f(S_t^* | S_{t-1}^*, \varepsilon_t) \propto g(\varepsilon_t(S_t)) \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{\sqrt{V}} \right), \tag{23}$$

where

$$V = \frac{1 - \rho^2}{1 + \psi^2}; \quad \text{and} \quad \mu_t = \alpha + \frac{1}{1 + \psi^2} (\psi(S_{t+1}^* - \alpha) - \rho\psi\varepsilon_{t+1} + \psi(S_{t-1}^* - \alpha) + \rho\varepsilon_t). \tag{24}$$

$$g(\varepsilon_t(S_t)) = \exp\left\{-\frac{\rho^2\varepsilon_t^2 + 2\rho\psi(S_{t-1}^* - \alpha)\varepsilon_t - (1 + \psi^2)(\mu_t - \alpha)^2}{2(1 - \rho^2)}\right\}, \quad (25)$$

where  $\varepsilon_t = \frac{y_t - x_t'\beta_{S_t}}{\sigma_{S_t}}$ .

By substituting equation (23) in equation (22), we obtain the following final derivation for the full conditional joint distribution for  $S_t$  and  $S_t^*$ :

$$f(S_t, S_t^* | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) \propto \left[ \frac{1}{\sigma_{S_t}} \phi\left(\frac{y_t - x_t'\beta_{S_t}}{\sigma_{S_t}}\right) \right] \left[ g(\varepsilon_t(S_t)) \frac{1}{\sqrt{V}} \phi\left(\frac{S_t^* - \mu_t}{\sqrt{V}}\right) \right] f(S_t | S_t^*), \quad (26)$$

where  $f(S_t | S_t^*) = 1[S_t^* \geq 0]S_t + 1[S_t^* < 0](1 - S_t)$ . Here, unlike in the case of exogenous switching model,  $g(\varepsilon_t(S_t))$  term is not a part of the normalizing constant as it is a function of  $S_t$ .

Finally, by integrating  $S_t^*$  out of equation (22) we obtain

$$\begin{aligned} f(S_t = 0 | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) & \propto \int_{-\infty}^{\infty} \left[ \frac{1}{\sigma_0} \phi\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right) \right] \left[ g(\varepsilon_t(S_t = 0)) \frac{1}{\sqrt{V}} \phi\left(\frac{S_t^* - \mu_t}{\sqrt{V}}\right) \right] f(S_t = 0 | S_t^*) dS_t^* \\ & = \int_{-\infty}^0 \left[ \frac{1}{\sigma_0} \phi\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right) \right] \left[ g(\varepsilon_t(S_t = 0)) \frac{1}{\sqrt{V}} \phi\left(\frac{S_t^* - \mu_t}{\sqrt{V}}\right) \right] dS_t^* \\ & = \left[ \frac{1}{\sigma_0} \phi\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right) \right] \left[ g(\varepsilon_t(S_t = 0)) \Phi\left(-\frac{\mu_t}{\sqrt{V}}\right) \right], \end{aligned} \quad (27)$$

$$\begin{aligned} f(S_t = 1 | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) & \propto \int_{-\infty}^{\infty} \left[ \frac{1}{\sigma_1} \phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right) \right] \left[ g(\varepsilon_t(S_t = 1)) \frac{1}{\sqrt{V}} \phi\left(\frac{S_t^* - \mu_t}{\sqrt{V}}\right) \right] f(S_t = 1 | S_t^*) dS_t^* \\ & = \int_0^{\infty} \left[ \frac{1}{\sigma_1} \phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right) \right] \left[ g(\varepsilon_t(S_t = 1)) \frac{1}{\sqrt{V}} \phi\left(\frac{S_t^* - \mu_t}{\sqrt{V}}\right) \right] dS_t^* \\ & = \left[ \frac{1}{\sigma_1} \phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right) \right] \left[ g(\varepsilon_t(S_t = 1)) \Phi\left(\frac{\mu_t}{\sqrt{V}}\right) \right], \end{aligned} \quad (28)$$

where  $\Phi(\cdot)$  is the CDF of the standard normal distribution. Thus, we can generate  $S_t$  based on the following probabilities:

$$P(S_t = i | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) = \frac{f(S_t = i | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T)}{f(S_t = 0 | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) + f(S_t = 1 | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T)}, \quad i = 0, 1. \quad (29)$$

### 3.2. Generating $S_t^*$ conditional on $\tilde{A}$ , $\tilde{\beta}$ , $\tilde{\sigma}^2$ , $\tilde{S}_{\neq t}^*$ , $S_t$ , $\tilde{S}_{\neq t}$ , and $\tilde{Y}_T$

The full conditional density  $f(S_t^* | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T)$ , from which  $S_t^*$  is to be drawn, can be derived based on equations (26). As the first term on the right-hand-side of equation (26) and the  $g(\varepsilon_t(S_t))$  term are normalizing constants conditional on  $S_t$ , we have the following results:

$$\begin{aligned} & f(S_t^* | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T) \\ & \propto f(S_t^*, S_t | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, \tilde{Y}_T) \\ & \propto \left[ \frac{1}{\sigma_{S_t}} \phi \left( \frac{y_t - x_t' \beta_{S_t}}{\sigma_{S_t}} \right) \right] \left[ g(\varepsilon_t(S_t)) \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{\sqrt{V}} \right) \right] f(S_t | S_t^*) \\ & \propto \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{\sqrt{V}} \right) f(S_t | S_t^*), \end{aligned} \quad (30)$$

which indicates that we can generate  $S_t^*$  from the following truncated normal distribution:

$$S_t^* | \tilde{S}_{\neq t}^*, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T \sim N(\mu_t, V)_{(1[S_t^* \geq 0]S_t + 1[S_t^* < 0](1 - S_t))}, \quad (31)$$

where  $\mu_t$  and  $V$  are given in equation (24).

### 3.3. Generating $\tilde{A} = [\alpha \quad \psi \quad \rho]'$ conditional on $\tilde{\beta}$ , $\tilde{\sigma}^2$ , $\tilde{S}_T^*$ , $\tilde{S}_T$ , and $\tilde{Y}_T$

Generating  $\rho$  conditional on  $\alpha$ ,  $\psi$ ,  $\tilde{\beta}$ ,  $\tilde{\sigma}^2$ ,  $\tilde{S}_T^*$ ,  $\tilde{S}_T$ , and  $\tilde{Y}_T$

The full conditional density from which  $\rho$  is drawn can be derived as:<sup>8</sup>

---

<sup>8</sup> For the sake of notational brevity, we suppress  $\alpha$ ,  $\psi$ ,  $\tilde{\beta}$ , and  $\tilde{\sigma}^2$  in the conditional density of  $\rho$ .

$$\begin{aligned}
f(\rho|\tilde{S}_T^*, \tilde{S}_T, \tilde{Y}_T) &\propto f(\tilde{S}_T^*, \tilde{S}_T, \tilde{Y}_T|\rho)f(\rho) \\
&= \prod_{t=1}^T [f(S_t^*, S_t, y_t|\tilde{S}_{t-1}^*, \tilde{S}_{t-1}, \tilde{Y}_{t-1}, \rho)]f(S_0^*, S_0)f(\rho) \\
&\propto \prod_{t=1}^T [f(S_t^*|S_{t-1}^*, S_t, y_t, \rho)f(S_t, y_t|S_{t-1}^*, \rho)]f(\rho) \\
&\propto \prod_{t=1}^T [f(S_t^*|S_{t-1}^*, S_t, \varepsilon_t, \rho)f(S_t, y_t|S_{t-1}^*, \rho)]f(\rho),
\end{aligned} \tag{32}$$

where  $f(\rho)$  is the prior density of  $\rho$ ; and  $f(S_t^*|S_{t-1}^*, S_t, \varepsilon_t, \rho)$  and  $f(S_t, y_t|S_{t-1}^*, \rho)$  are given in equations (20) and (21), respectively, with the  $\rho$  parameter suppressed. By substituting equations (20) and (21) into equation (32), we obtain the following target density of  $\rho$ :

$$f(\rho|\tilde{S}_T^*, \tilde{Y}_T, \tilde{S}_T) \propto \prod_{t=1}^T [f(S_t^*|S_{t-1}^*, \varepsilon_t, \rho)]f(\rho), \tag{33}$$

where the density  $f(S_t^*|S_{t-1}^*, \varepsilon_t, \rho)$  can be derived from equation (18), as given below:

$$f(S_t^*|S_{t-1}^*, \varepsilon_t, \rho) \propto \frac{1}{\sqrt{(1-\rho^2)}} \phi \left( \frac{S_t^* - \alpha(1-\psi) - \psi S_{t-1}^* - \rho \varepsilon_t}{\sqrt{(1-\rho^2)}} \right). \tag{34}$$

The intuition is that, conditional on  $\tilde{S}_T^*$  and  $\tilde{\varepsilon}_T$ , all that matters for the derivation of the likelihood function for  $\rho$  is equation (18).

For the Metropolis-Hastings algorithm, let  $\rho_o$  denote the accepted  $\rho$  at the previous MCMC iteration and  $\rho_n$  is a newly generated candidate from the following random walk candidate generating distribution:

$$\rho_n = \rho_o + \epsilon, \quad \epsilon \sim N(0, c)_{1[-\rho_o-1 < \epsilon < -\rho_o+1]}, \tag{35}$$

so that  $\rho_n$  is constrained to be between -1 and 1. Based on the target posterior density of  $\rho$  in equation (33), we employ the following acceptance probability to decide whether to accept or reject  $\rho_n$ :

$$\alpha(\rho_n, \rho_o) = \min \left\{ 1, \frac{\prod_{t=1}^T [f(S_t^*|S_{t-1}^*, \varepsilon_t, \rho_n)]f(\rho_n)}{\prod_{t=1}^T [f(S_t^*|S_{t-1}^*, \varepsilon_t, \rho_o)]f(\rho_o)} \right\}. \tag{36}$$

Generating  $\alpha$  conditional on  $\rho, \psi, \tilde{\beta}, \tilde{\sigma}^2, \tilde{S}_T, \tilde{S}_T^*$ , and  $\tilde{Y}_T$

Rearranging equation (18), we have

$$\left( \frac{S_t^* - \psi S_{t-1}^* - \rho \varepsilon_t}{\sqrt{1 - \rho^2}} \right) = \alpha \left( \frac{1 - \psi}{\sqrt{1 - \rho^2}} \right) + \omega_t^*, \quad \omega_t^* \sim i.i.d.N(0, 1), \quad (37)$$

and drawing  $\alpha$  from the full conditional distribution derived based on this equation is standard.

Generating  $\psi$  conditional on  $\rho, \alpha, \tilde{\beta}, \tilde{\sigma}^2, \tilde{S}_T, \tilde{S}_T^*$ , and  $\tilde{Y}_T$

Rearranging equation (18), we have

$$\hat{S}_t^* = \psi z_t + \omega_t^{**}, \quad \omega_t^{**} \sim i.i.d.N(0, 1), \quad (38)$$

where  $\hat{S}_t^* = (S_t^* - \alpha - \rho \varepsilon_t) / \sqrt{1 - \rho^2}$  and  $z_t = (S_{t-1}^* - \alpha) / \sqrt{1 - \rho^2}$ . Drawing  $\psi$  from an appropriate full conditional distribution derived based on this equation is standard.

**3.4. Generating  $\tilde{\beta} = [\beta_0 \ \beta_1]'$  and  $\tilde{\sigma}^2 = [\sigma_0^2 \ \sigma_1^2]'$  conditional on  $\tilde{A}, \tilde{S}_T, \tilde{S}_T^*$ , and data  $\tilde{Y}_T$**

Generating  $\tilde{\beta}$  conditional on  $\tilde{\sigma}^2, \tilde{A}, \tilde{S}_T, \tilde{S}_T^*$ , and data  $\tilde{Y}_T$

From the joint normality of  $\varepsilon_t$  and  $\omega_t$  in equation (3), we can write  $\varepsilon_t$  as a function of  $\omega_t$  (i.e.,  $\varepsilon_t = \rho \omega_t + \sqrt{1 - \rho^2} \varepsilon_t^*$ ), which allows us to rewrite the regression equation in equation (1) as:

$$y_t = x_t' \beta_{S_t} + \sigma_{S_t} \rho \omega_t + \sigma_{S_t} \sqrt{1 - \rho^2} \varepsilon_t^*, \quad \varepsilon_t^* \sim i.i.d.N(0, 1), \quad (39)$$

where  $\omega_t = S_t^* - \alpha(1 - \psi) - \psi S_{t-1}^*$ . Thus, the full conditional distribution from which  $\tilde{\beta} = [\beta_0 \ \beta_1]'$  is to be drawn can be easily derived from the following regression equation that is obtained by rearranging equation (39):

$$y_t^* = x_t \beta_{S_t} + e_t, \quad e_t \sim i.i.d.N(0, \sigma_{S_t}^2 (1 - \rho^2)), \quad (40)$$

$$y_t^* = y_t - \sigma_{S_t} \rho \omega_t.$$

Generating  $\tilde{\sigma}^2$  conditional on  $\tilde{\beta}$ ,  $\tilde{A}$ ,  $\tilde{S}_T$ , and data  $\tilde{Y}_T$

As  $\rho$  is irrelevant in making inference on  $\tilde{\sigma}^2$ , this step is based on equation (1). Thus, conditional on  $\tilde{\beta}$ , it is straightforward to generate  $\tilde{\sigma}^2$  from an appropriate full conditional distribution.

#### 4. Simulation Study: Performance of the Proposed Algorithm and Pitfalls of Estimating Non-Markovian processes with Markovian Switching Models

In this section, we conduct a simulation study in order to evaluate the performance of the proposed algorithm for Bayesian estimation of non-Markovian regime switching models. Our simulation study is based on the following non-Markovian switching model:

$$\begin{aligned} y_t &= \beta_0 1[S_t^* < 0] + \beta_1 1[S_t^* \geq 0] + (\sigma_0 1[S_t^* < 0] + \sigma_1 1[S_t^* \geq 0])\varepsilon_t, \\ S_t^* &= \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t, \\ \begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} &\sim i.i.d.N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \end{aligned} \tag{41}$$

where  $1[\cdot]$  is the indicator function.

When generating data, we consider the following two alternative cases that differ in the parameter values assigned:

Case #1:  $\rho = 0$  [Exogenous Switching]

$$\rho = 0; \beta_0 = 1; \beta_1 = 1; \sigma_0^2 = 0.5; \sigma_1^2 = 1; \psi = 0.9; \alpha = -1$$

Case #2:  $\rho = 0.9$  [Endogenous Switching]

$$\rho = 0.9; \beta_0 = 1; \beta_1 = 1; \sigma_0^2 = 0.5; \sigma_1^2 = 1; \psi = 0.9; \alpha = 0$$

For each case, we generate 1000 samples of 500 or 3,000 observations. For estimation of Case #1, we apply both the non-Markovian exogenous switching model in Section 2 and the Markovian exogenous switching model. For estimation of Case #2, we apply both the non-Markovian endogenous switching model in Section 3 and the Markovian endogenous

switching model. For each model and for each case, we obtain the sampling distribution of the posterior mean for each parameter.

Table 1.A reports the results for Case #1. Regardless of the sample size, estimation results for the non-Markovian exogenous switching model based on the proposed algorithm show in little bias in the parameter estimates. The only difference for different sample sizes is that the standard deviations decrease when the sample size increases. However, estimation results for the Markovian exogenous switching model show bias in some parameter estimates when  $T = 500$ , and this bias does not disappear even when  $T = 3000$ .

Table 1.B reports the results for Case #2. Again, regardless of the sample size, estimation results for a correctly specified model based on the proposed algorithm show little bias in the parameter estimates. However, estimation results for a misspecified model show considerable bias in some parameter estimates regardless of the sample size. For example, the estimates of the  $\rho$  parameter considerably underestimate the true value when a Markovian switching model is employed. While the true value of  $\rho$  is 0.9, the sample mean of the posterior mean is around 0.5. Note that, for Case #1, we estimated the two models under the maintained assumption that the true value  $\rho = 0$  is known. For Case #2, however,  $\rho$  is assumed unknown and estimated, and a misspecified model leads to considerable bias in its estimates. This is the reason why the problem of employing a mis-specified model (i.e., a Markovian switching model) may be more severe for the case of endogenous switching than for that of the exogenous switching.

In sum, the proposed algorithms for non-Markovian switching models result in little bias in the parameter estimates. On the contrary, estimating a non-Markovian switching process by a Markovian switching model results in considerable bias in the parameter estimates, especially when the data generating process involves endogenous switching.

## 5. Applications

### 5.1. Markovian versus Non-Markovian Switching Models of Real GDP [1952Q1 - 2007Q2]

Consider the following model specification for the log of real GDP covering the period



1952Q1 - 2007Q2:

$$\begin{aligned}
y_t &= \beta_{0,M_t}1[S_t^* < 0] + \beta_{1,M_t}1[S_t^* \geq 0] + \sigma_{M_t}\varepsilon_t, \\
\beta_{0,M_t} &> \beta_{1,M_t}, \\
M_t &= \begin{cases} 0, & \text{if } t < 1984Q4; \\ 1, & \text{otherwise,} \end{cases} \\
\begin{cases} \text{Non - Markovian Switching :} \\ \text{or Markovian Switching :} \end{cases} & \begin{cases} S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t, \\ S_t^* = \alpha_0 + \alpha_1 S_{t-1}^* + \omega_t, \end{cases} \\
\begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} &\sim i.i.d.N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),
\end{aligned}$$

where  $y_t$  is the log difference of real GDP plotted in Figure 1.  $\beta_{0,M_t}$  and  $\beta_{1,M_t}$  represent the mean growth rate of real GDP during boom and recession, respectively. Following Kim and Nelson (1999), we assume that these mean growth rates as well as the standard deviation of the shocks ( $\sigma_{M_t}$ ) underwent a structural break with the onset of the Great Moderation in 1984Q4. Note that the above model can be considered as an extension of Kim and Nelson's (1999) model, in which the regime-indicator variable follows a first-order Markovian exogenous switching process, to the case of non-Markovian endogenous switching.

We estimate four competing models, i.e., non-Markovian and Markovian switching models with endogenous switching ( $\rho \neq 0$ ) and those with exogenous switching ( $\rho = 0$ ). If we focus on the results for endogenous switching models reported in Table 2.B, a non-Markovian switching model seems to be preferred to a Markovian switching model by a Bayesian model selection criteria (i.e., Watanabe-Akaike Information Criterion or WAIC, Watanabe(2010)). However, the results in Table 2.A show that models with exogenous switching have much lower WAIC's than those with endogenous switching. We thus focus on the discussion of the results for exogenous switching models in Table 2.A.

Estimation results for the two models with exogenous switching seem to be very close. Plots of recession probabilities from both the Markovian and non-Markovian exogenous switching models are presented in Figure 2.A, and they are almost identical. Plots of the latent variable  $S_t^*$  from the two models presented in Figure 2.B are also very close. Besides, the WAIC's for these two models are very close, even though the Bayesian information criterion for the Markovian switching model (WAIC=253.29) is slightly lower than that for

the non-Markovian switching model (WAIC=252.65).

We thus conclude that the non-Markovian exogenous switching model and Hamilton's (1998) conventional Markovian switching model performs equally well in describing the regime switching dynamics of the business cycle dynamics in postwar US real GDP, even though the latter is slightly preferred. Besides, the evidence of endogenous switching does not seem to be very compelling.

## 5.2. Non-Markovian versus Markovian Switching Models of Weekly Stock Return Volatility [January 1990 –May 2017]

In the literature on stochastic volatility, asymmetry in the stock return is typically modeled by introducing the leverage effect. By denoting  $\zeta_t$  as the innovation to the stock return volatility and  $\varepsilon_t$  as the innovation to the stock return, one approach is to assume  $E(\zeta_t \varepsilon_t) \neq 0$  and the other is to assume  $E(\zeta_t \varepsilon_{t-1}) \neq 0$ . For example, the former is adopted by Jacquier et al. (2004) and the latter is adopted by Harvey and Shephard (1996), among others. Yu (2005) provides a discussion on some of the issues related to these two alternative approaches to modeling the leverage effect.

In this section, we consider the leverage effect within a regime-switching model of the stock return volatility. By noting that  $\omega_t$  in equation (2) is equivalent to  $\zeta_t$  in a stochastic volatility, we adopt the assumption that  $E(\varepsilon_t \omega_t) \neq 0$  in line with Jacquier et al. (2004). An empirical model that we employ is given below:

$$\begin{aligned}
 r_t &= \beta_0 + \beta_1 r_{t-1} + (\sigma_0^2 1[S_t^* < 0] + \sigma_1^2 1[S_t^* \geq 0]) \varepsilon_t, \\
 &\quad \sigma_0^2 < \sigma_1^2, \\
 &\quad \left\{ \begin{array}{l} \text{Non - Markovian Switching : } S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t, \\ \text{or Markovian Switching : } S_t^* = \alpha_0 + \alpha_1 S_{t-1}^* + \omega_t, \end{array} \right. \\
 &\quad \begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} \sim i.i.d.N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),
 \end{aligned}$$

where  $r_t$  is the excess stock return.

Data we use are the weekly excess stock return for value-weighted portfolio of all CRSP firms listed on the NYSE, AMEX, or NASDAQ. The sample period is from the first week of

January 1990 to the fourth week of May 2017.<sup>9</sup> The data are plotted in Figure 3.

Tables 3.A and 3.B report Bayesian estimates for four competing models. Unlike the example of postwar real GDP in Section 5.1, endogenous switching or the leverage effect seems to be an important feature of the stock return volatility. That is, the Bayesian model selection criteria favors models with endogenous switching or the leverage effect to exogenous switching models. We thus focus our attention on comparing these two models with endogenous switching.

Table 3.B reports that the posterior mean of  $\rho$  is -0.45 with a small posterior standard deviation (0.09) for the Markovian switching model, while it is -0.68 with a smaller posterior standard deviation (0.06) for the non-Markovian switching model. As discussed in Section 4, this result is what we would expect if the true data generating process is non-Markovian. Actually, between these two endogenous switching models, the non-Markovian switching model (WAIC= 2981.47) is strongly preferred to the Markovian switching model (WAIC=3014.45).

Figure 4.A plots and compares the posterior means of time-varying volatility obtained from these models. Figure 4.B plots the posterior means of  $S_t^*$  obtained from the two competing models. These figures show how inferences based on potentially misspecified model (a Markovian switching model) can be different from those based on a non-Markovian switching model, which is preferred by the Bayesian model selection criterion.

## 6. Summary and Concluding Remarks

In this paper, we present algorithms for Bayesian estimation of non-Markovian switching models. Our simulation study shows that the proposed algorithms work well and that estimating a non-Markovian process by a Markovian switching model may be problematic, especially in the presence of endogenous switching. When the model and the proposed algorithm are applied to postwar US real GDP, evidence of endogenous switching is not very compelling and the Markovian exogenous switching model is slightly preferred to the non-Markovian switching model. For the volatility model of weekly stock returns, however, evi-

---

<sup>9</sup> The excess return data are freely available at the data library of Kenneth R. French's home page at [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

dence of endogenous switching or the leverage effect is much more compelling. Furthermore, the non-Markovian endogenous switching model is very strongly preferred to the Markovian endogenous switching by the Bayesian model selection criterion.

Extending two-state non-Markovian switching models to general  $N$ -state models is non-trivial. One reason is because it is not very straightforward to solve the the labeling problem that is potentially associated with the specification of the transition probabilities based on a multinomial extension of equation (2). This project is currently in progress.

## Appendix A. Derivation of Equation (21)

In this appendix, we derive equation (21) for  $j = 0$ . For this purpose, by denoting  $y_t^*$  as a realization of  $y_t$  for which we want to compute  $f(y_t^*, S_t | S_{t-1}^*)$ , consider the following CDF based on  $f(y_t, S_t | S_{t-1}^*)$ :

$$\begin{aligned} \int_{-\infty}^{y_t^*} f(y_t, S_t | S_{t-1}^*) dy_t &= \int_{-\infty}^{y_t^*} f(y_t | S_{t-1}^*) f(S_t | S_{t-1}^*, y_t) dy_t \\ &= \int_{-\infty}^{\frac{y_t^* - x_t' \beta_{S_t}}{\sigma_{S_t}}} f(\varepsilon_t | S_{t-1}^*) f(S_t | S_{t-1}^*, \varepsilon_t) d\varepsilon_t \\ &= \int_{-\infty}^{\frac{y_t^* - x_t' \beta_{S_t}}{\sigma_{S_t}}} \phi(\varepsilon_t) f(S_t | S_{t-1}^*, \varepsilon_t) d\varepsilon_t, \end{aligned} \quad (A.1)$$

where  $\phi(\cdot)$  refers to the p.d.f. of the standard normal distribution and the second line holds due to the variable change  $\varepsilon_t = \frac{y_t - x_t' \beta_{S_t}}{\sigma_{S_t}}$ .

By differentiating equation (A.1) with respect to  $y_t^*$ , we obtain

$$f(y_t^*, S_t | S_{t-1}^*) = \frac{1}{\sigma_{S_t}} \phi\left(\frac{y_t^* - x_t' \beta_{S_t}}{\sigma_{S_t}}\right) f(S_t | S_{t-1}^*, \varepsilon_t), \quad (A.2)$$

which implies that equation (21) holds.

## Appendix B. Derivation of Equation (23)

First, note that equation (18) can be rewritten as

$$\eta_t = \psi \eta_{t-1} + \rho \varepsilon_t + \sqrt{1 - \rho^2} \omega_t^*, \quad \omega_t \sim i.i.d. N(0, 1), \quad (B.1)$$

where  $\eta_t = S_t^* - \alpha$ . Then, as

$$f(S_{t+j}^* | S_{t+j-1}^*, \varepsilon_{t+j}) = \frac{1}{\sqrt{1 - \rho^2}} \phi\left(\frac{S_{t+j}^* - \alpha(1 - \psi) - \psi S_{t+j-1}^* - \rho \varepsilon_{t+j}}{\sqrt{1 - \rho^2}}\right), \quad j = 0, 1, \quad (B.2)$$

where  $\varepsilon_t = \frac{y_t - x_t' \beta_{S_t}}{\sigma_{S_t}}$ , equation (23) can be derived as follows:

$$\begin{aligned}
& f(S_{t+1}^* | S_t^*, \varepsilon_{t+1}) f(S_t^* | S_{t-1}^*, \varepsilon_t) \\
& \propto \exp \left[ -\frac{1}{2} \left( \frac{(\eta_{t+1} - \psi\eta_t - \rho\varepsilon_{t+1})^2}{1 - \rho^2} + \frac{(\eta_t - \psi\eta_{t-1} - \rho\varepsilon_t)^2}{1 - \rho^2} \right) \right] \\
& \propto \exp \left[ -\frac{\rho^2\varepsilon_t^2 + 2\rho\psi\eta_{t-1}\varepsilon_t - (1 + \psi^2) \left( \frac{1}{1+\psi^2} (\psi\eta_{t+1} - \rho\psi\varepsilon_{t+1} + \psi\eta_{t-1} + \rho\varepsilon_t) \right)^2}{2(1 - \rho^2)} \right] \\
& \quad \times \exp \left[ -\frac{1}{2\frac{1-\rho^2}{1+\psi^2}} \left( \eta_t - \frac{1}{1+\psi^2} (\psi\eta_{t+1} - \rho\psi\varepsilon_{t+1} + \psi\eta_{t-1} + \rho\varepsilon_t) \right)^2 \right] \\
& \propto g(\varepsilon_t(S_t)) \frac{1}{\sqrt{V}} \phi \left( \frac{S_t^* - \mu_t}{V} \right),
\end{aligned} \tag{B.3}$$

where  $\phi(\cdot)$  is the p.d.f. of the standard normal distribution and

$$\mu_t = \alpha + \frac{1}{1 + \psi^2} (\psi(S_{t+1}^* - \alpha) - \rho\psi\varepsilon_{t+1} + \psi(S_{t-1}^* - \alpha) + \rho\varepsilon_t); \quad \text{and } V = \frac{1 - \rho^2}{1 + \psi^2},$$

and

$$g(\varepsilon_t(S_t)) = \exp \left\{ -\frac{\rho^2\varepsilon_t^2 + 2\rho\psi(S_{t-1}^* - \alpha)\varepsilon_t - (1 + \psi^2)(\mu_t - \alpha)^2}{2(1 - \rho^2)} \right\},$$

which is a function of  $S_t$  as  $\varepsilon_t = \frac{y_t - x_t'\beta_{S_t}}{\sigma_{S_t}}$ .

## References

- [1] Chang, Y., Y. Choi, and J.Y. Park, 2017, “A New Approach to Model Regime Switching,” *Journal of Econometrics*, 196, 127-143.
- [2] Chauvet, M. and S. Potter, “Forecasting Recessions Using the Yield Curve,” *Journal of Forecasting*, 24, 2, 77-103, 2005.
- [3] Chib, S. and M. Dueker, 2004, “Non-Markovian Regime Switching with Endogenous States and Time-Varying State Strengths,” unpublished manuscript, Washington University.
- [4] Hamilton, James D., 1989, “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica*, Vol. 57, No. 2, 357-384.
- [5] Hamilton, James D., 2016, “Macroeconomic Regimes and Regime Shifts,” *Handbook of Macroeconomics*, Volume 2A, 163-201, edited by Harald Uhlig and John Taylor, Amsterdam: Elsevier.
- [6] Harvey, A. and N. Shephard, 1996, “Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns,” *Journal of Business and Economic Statistics*, Vol. 14, No. 4, 429-434.
- [7] Hering, A.S., K. Kazor, and W. Kleiber, 2015, “A Markov-Switching Vector Autoregressive Stochastic Wind Generator for Multiple Spatial and Temporal Sales,” *Resources*, 4, 70-92.
- [8] Jacquier, E., N.G. Polson, and P.E. Rossi, 2004, “Bayesian Analysis of Stochastic Volatility Models with Fat Tails and Correlated Errors,” *Journal of Econometrics*, 122, 185-212.
- [9] Kauppi, H. and Saikkonen, P., 2008. “Predicting US recessions with dynamic binary response models,” *The Review of Economics and Statistics*, 90(4), pp.777-791.
- [10] 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 Business Cycle,” *Review of Economics and Statistics*, 81(4): 608-616.

- [11] Kim, C.-J., J. Piger, and R. Startz, 2008, "Estimation of Markov Regime-Switching Regressions with Endogenous Switching," *Journal of Econometrics*, 143 (2), 263-273.
- [12] Li, W.K., 1994, "Time Series Models Based on Generalized Linear Models: Some Further Results," *Biometrics*, 50, 506-511.
- [13] Neale, M.C., S.L. Clark, C.V. Dolan, and M.D. Hunter, 2016, "Regime Switching Modeling of Substance Use: Time-Varying and Second-Order Markov Models of Individual Probability Plots," *Struct Equ Modeling*, 23(2), 221-233.
- [14] Raftery, A. E., 1985, "A Model for High-Order Markov Chains," *Journal of Royal Statistical Society*, Series B, 47, 528-539.
- [15] Siu, T.K., W.K. Ching, E. Fung, M. Ng, and X. Li, 2009, "A Higher-Order Markov-Switching Model for Risk Measurement," *Computers and Mathematics with Applications*, 58, 1-10.
- [16] Startz, R., 2008, "Binomial Autoregressive Moving Average Models with an Application to US Recessions," *Journal of Business and Economic Statistics*, 26 (1), 1-8.
- [17] 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.
- [18] Yu, J., 2005, "On Leverage in a Stochastic Volatility Model," *Journal of Econometrics*, 127, 165-178.
- [19] Zeger, S.L. and Qaqish, B., 1988, "Markov Regression Models for time series: A quasi-likelihood approach," *Biometrics*, 44, 1019-1031.



**Table 1.A. Simulation Study I: Sampling Distributions for Posterior Means based on Markovian and Non-Markovian Switching Models [DGP: Non-Markovian Exogenous Switching]**

$$y_t = \beta_{S_t} + \sigma_{S_t} \varepsilon_t,$$

$$\text{Non-Markovian Switching: } S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t,$$

$$\text{Markovian Switching: } S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t,$$

$$\begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} \sim i.i.d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

		<u>Estimated Model</u>			
		<u>Non-Markovian Switching</u>		<u>Markovian Switching</u>	
		<u>Sample Size</u>		<u>Sample Size</u>	
<u>Parameter</u>	<u>True Value</u>	<u>T = 500</u>	<u>T = 3,000</u>	<u>T = 500</u>	<u>T = 3,000</u>
$\beta_0$	1	0.99 (0.04)	0.99 (0.02)	0.99 (0.04)	0.99 (0.01)
$\beta_1$	-1	-0.97 (0.12)	-0.99 (0.04)	-0.85 (0.13)	-0.86 (0.05)
$\sigma_0^2$	0.5	0.52 (0.05)	0.51 (0.01)	0.52 (0.05)	0.51 (0.02)
$\sigma_1^2$	1	1.06 (0.17)	1.02 (0.07)	1.24 (0.18)	1.21 (0.08)
$\alpha$	-1	-1.03 (0.38)	-1.06 (0.22)	-	-
$\psi$	0.9	0.90 (0.02)	0.90 (0.01)	-	-
$\alpha_0$	-	-	-	-1.52 (0.18)	-1.50 (0.06)
$\alpha_1$	-	-	-	1.19 (0.20)	1.16 (0.08)

- Note:
1. This table reports mean and standard deviation (in the parenthesis) for the sampling distribution of the posterior mean.
  2. Number of Simulation = 1,000; Total iterations / Burn-in (for each sample) = 30,000/5,000
  3. Non-informative priors are employed.

**Table 1.B. Simulation Study II: Sampling Distributions for Posterior Means based on Markovian and Non-Markovian Switching Models [DGP: Non-Markovian Endogenous Switching]**

$$y_t = \beta_{S_t} + \sigma_{S_t} \varepsilon_t,$$

$$\text{Non-Markovian Switching: } S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t,$$

$$\text{Markovian Switching: } S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t,$$

$$\begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} \sim i.i.d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

<u>Estimated Model</u>					
		<u>Non-Markovian Switching</u>		<u>Markovian Switching</u>	
		<u>Sample Size</u>		<u>Sample Size</u>	
<u>Parameter</u>	<u>True Value</u>	<u>T = 500</u>	<u>T = 3,000</u>	<u>T = 500</u>	<u>T = 3,000</u>
$\beta_0$	1	0.97 (0.05)	0.98 (0.02)	0.86 (0.07)	0.87 (0.02)
$\beta_1$	-1	-0.96 (0.08)	-0.98 (0.03)	-0.77 (0.11)	-0.78 (0.03)
$\sigma_0^2$	0.5	0.52 (0.06)	0.52 (0.02)	0.49 (0.06)	0.47 (0.02)
$\sigma_1^2$	1	1.01 (0.10)	1.01 (0.04)	1.01 (0.11)	0.99 (0.04)
$\alpha$	0	0.00 (0.38)	0.00 (0.20)	-	-
$\psi$	0.9	0.90 (0.02)	0.90 (0.01)	-	-
$\rho$	0.9	0.86 (0.05)	0.90 (0.01)	0.47 (0.13)	0.52 (0.05)
$\alpha_0$	-	-	-	-1.40 (0.21)	-1.38 (0.07)
$\alpha_1$	-	-	-	1.43 (0.27)	1.41 (0.08)

- Note:
1. This table reports mean and standard deviation (in the parenthesis) for the sampling distribution of the posterior mean.
  2. Number of Simulation = 1,000; Total iterations / Burn-in (for each sample) = 30,000/5,000
  3. Non-informative priors are employed.

**Table 2.A. Posterior Moments from Non-Markovian and Markovian Exogenous Switching Models of Business Cycle ]Quarterly Real GDP Growth Rate: 1952:Q1~2007:Q2]**

$$y_t = \beta_{S_t, M_t} + \sigma_{M_t} \varepsilon_t$$

$$M_t = 0 \text{ if } t \leq 1984: Q4$$

$$M_t = 1 \text{ otherwise}$$

$$\text{Non-Markovian Switching: } S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t$$

$$\text{Markovian Switching: } S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t,$$

$$\begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} \sim i.i.d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

		<i>Non-Markov Switching</i>					<i>Markov Switching</i>				
Prior		Posterior					Posterior				
		Mean	Mode	(SD)	90 % HPDI		Mean	Mode	(SD)	90 % HPDI	
$\beta_{0,0}$	$N(1,2^2)$	1.29	1.29	(0.15)	1.03	1.54	1.31	1.34	(0.15)	1.06	1.57
$\beta_{1,0}$	$N(-1,2^2)$	-0.35	-0.18	(0.37)	-0.75	0.16	-0.26	-0.18	(0.29)	-0.69	0.18
$\beta_{0,1}$	$N(0.5,2^2)$	0.92	0.91	(0.07)	0.80	1.03	0.92	0.92	(0.07)	0.81	1.04
$\beta_{1,1}$	$N(-0.5,2^2)$	0.20	0.31	(0.23)	-0.16	0.57	0.24	0.32	(0.21)	-0.09	0.60
$\sigma_0^2$	$IG(1,1)$	0.87	0.83	(0.15)	0.62	1.13	0.87	0.82	(0.14)	0.64	1.11
$\sigma_1^2$	$IG(2,1)$	0.20	0.20	(0.03)	0.14	0.27	0.21	0.20	(0.03)	0.15	0.27
$\alpha$	$N(-1, 2^2)$	-1.17	-1.03	(0.61)	-2.07	-0.18	-	-	-	-	-
$\psi$	$N(0.9,1^2)$	0.75	0.80	(0.10)	0.60	0.92	-	-	-	-	-
$\alpha_0$	$N(-1.5, 2^2)$	-	-	-	-	-	-1.41	-1.41	(0.25)	-1.82	-1.00
$\alpha_1$	$N(1, 2^2)$	-	-	-	-	-	0.68	0.74	(0.31)	0.21	1.21
WAIC		253.29					252.65				

Note: 1. Total iterations / Burn-in = 20,000 / 70,000  
 2. TN(.) represents a truncated normal distribution defined between -1 and 1.  
 3. WAIC =  $-\sum_{t=1}^T \log(\overline{pd}_t) + \sum_{t=1}^T V_R(\log(pd_t))$  where  $\overline{pd}_t$  is the MCMC sample average of predictive density  $p(y_t | \theta^{(r)}, S_t^{(r)})$  and  $V_R(\log(pd_t))$  is the MCMC sample variance of  $\log(p(y_t | \theta^{(r)}, S_t^{(r)}))$ , where  $\theta^{(r)}$  is the vector of the parameters drawn at the r-th iteration. A model with a lower WAIC is preferred in a model comparison.

**Table 2.B. Posterior Moments from Non-Markovian and Markovian Endogenous Switching Models Business Cycle [Quarterly Real GDP Growth Rate: 1952:Q1~2007:Q2]**

$$y_t = \beta_{S_t, M_t} + \sigma_{M_t} \varepsilon_t$$

$$M_t = 0 \text{ if } t \leq 1984:Q4$$

$$M_t = 1 \text{ otherwise}$$

$$\text{Non-Markovian Switching: } S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t$$

$$\text{Markovian Switching: } S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t,$$

$$\begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} \sim i.i.d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{M_t} \\ \rho_{M_t} & 1 \end{bmatrix}\right)$$

		<i>Non-Markov Switching</i>					<i>Markov Switching</i>				
Prior		Posterior					Posterior				
		Mean	Mode	(SD)	90 % HPDI	Mean	Mode	(SD)	90 % HPDI		
$\beta_{0,0}$	$N(1,2^2)$	1.39	1.43	(0.24)	0.99	1.76	1.40	1.38	(0.18)	1.09	1.69
$\beta_{1,0}$	$N(-1,2^2)$	-0.48	-0.32	(0.51)	-0.93	0.10	-0.61	-0.40	(0.48)	-1.26	0.13
$\beta_{0,1}$	$N(0.5,2^2)$	0.95	0.95	(0.09)	0.81	1.11	0.96	0.96	(0.08)	0.82	1.09
$\beta_{1,1}$	$N(-0.5,2^2)$	0.12	0.12	(0.26)	-0.28	0.55	0.08	0.20	(0.30)	-0.42	0.55
$\sigma_0^2$	$IG(1,1)$	0.94	0.91	(0.17)	0.67	1.24	0.97	0.94	(0.18)	0.69	1.26
$\sigma_1^2$	$IG(2,1)$	0.21	0.21	(0.04)	0.15	0.28	0.22	0.21	(0.04)	0.15	0.28
$\rho_0$	$TN(0,1^2)$	0.26	0.63	(0.44)	-0.82	0.88	0.41	0.57	(0.29)	0.00	0.87
$\rho_1$	$TN(0,1^2)$	0.28	0.53	(0.39)	-0.41	0.94	0.31	0.58	(0.34)	-0.17	0.88
$\alpha$	$N(-1, 2^2)$	-1.16	-0.87	(0.74)	-2.17	-0.07	-	-	-	-	-
$\psi$	$N(0.9,1^2)$	0.78	0.82	(0.09)	0.64	0.94	-	-	-	-	-
$\alpha_0$	$N(-1.5, 2^2)$	-	-	-	-	-	-1.36	-1.34	(0.26)	-1.77	-0.92
$\alpha_1$	$N(1, 2^2)$	-	-	-	-	-	0.59	0.53	(0.30)	0.14	1.09
WAIC		269.75					275.20				

- Note:
1. Total iterations / Burn-in = 20,000 / 70,000
  2. TN(.) represents a truncated normal distribution defined between -1 and 1.
  3. WAIC =  $-\sum_{t=1}^T \log(\bar{p}d_t) + \sum_{t=1}^T V_R(\log(p d_t))$  where  $\bar{p}d_t$  is the MCMC sample average of predictive density  $p(y_t | \theta^{(r)}, S_t^{(r)})$  and  $V_R(\log(p d_t))$  is the MCMC sample variance of  $\log(p(y_t | \theta^{(r)}, S_t^{(r)}))$ , where  $\theta^{(r)}$  is the vector of the parameters drawn at the r-th iteration. A model with a lower WAIC is preferred in a model comparison.

**Table 3.A. Posterior Moments from Non-Markovian and Markovian Exogenous Switching Models of Volatility [Weekly Excessive Return: 1990:M1:W1~2017:M5:W4]**

$$y_t = \beta_0 + \beta_1 y_{t-1} + \sigma_{S_t} \varepsilon_t,$$

$$\text{Non-Markovian Switching: } S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t,$$

$$\text{Markovian Switching: } S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t,$$

$$\begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} \sim i.i.d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

		<i>Non-Markov Switching</i>					<i>Markov Switching</i>				
Prior		Posterior					Posterior				
		<u>Mean</u>	<u>Mode</u>	<u>(SD)</u>	<u>90 % HPDI</u>		<u>Mean</u>	<u>Mode</u>	<u>(SD)</u>	<u>90 % HPDI</u>	
$\beta_0$	$N(0, 2^2)$	0.28	0.30	(0.05)	0.20	0.36	0.26	0.26	(0.05)	0.18	0.35
$\beta_1$	$N(0, 2^2)$	-0.08	-0.07	(0.03)	-0.13	-0.04	-0.09	-0.09	(0.03)	-0.13	-0.04
$\sigma_0^2$	$IG(1, 1)$	2.20	2.19	(0.14)	1.98	2.46	2.61	2.67	(0.20)	2.28	2.93
$\sigma_1^2$	$IG(1, 2)$	12.92	12.58	(1.28)	10.86	14.99	16.28	16.19	(2.25)	12.18	19.82
$\alpha$	$N(-1, 2^2)$	-2.25	-2.25	(1.22)	-4.20	-0.24	-	-	-	-	-
$\psi$	$N(0.9, 1^2)$	0.98	0.98	(0.01)	0.96	0.99	-	-	-	-	-
$\alpha_0$	$N(-1.5, 2^2)$	-	-	-	-	-	-2.16	-2.15	(0.14)	-2.41	-1.93
$\alpha_1$	$N(1, 2^2)$	-	-	-	-	-	1.51	1.44	(0.20)	1.20	1.85
WAIC		3004.76					3016.62				

- Note:
1. Total iterations / Burn-in = 20,000 / 70,000
  2. TN(.) represents a truncated normal distribution defined between -1 and 1.
  3. WAIC =  $-\sum_{t=1}^T \log(\overline{pd}_t) + \sum_{t=1}^T V_R(\log(pd_t))$  where  $\overline{pd}_t$  is the MCMC sample average of predictive density  $f(y_t | \theta^{(r)}, S_t^{(r)})$  and  $V_R(\log(pd_t))$  is the MCMC sample variance of  $\log(f(y_t | \theta^{(r)}, S_t^{(r)}))$ , where  $\theta^{(r)}$  is the vector of the parameters drawn at the r-th iteration. A model with a lower WAIC is preferred in a model comparison.

**Table 3.B. Posterior Moments from Non-Markovian and Markovian Endogenous Switching Models of Volatility [Weekly Excessive Return: 1990:M1:W1~2017:M5:W4]**

$$y_t = \beta_0 + \beta_1 y_{t-1} + \sigma_{S_t} \varepsilon_t,$$

$$\text{Non-Markovian Switching: } S_t^* = \alpha(1 - \psi) + \psi S_{t-1}^* + \omega_t,$$

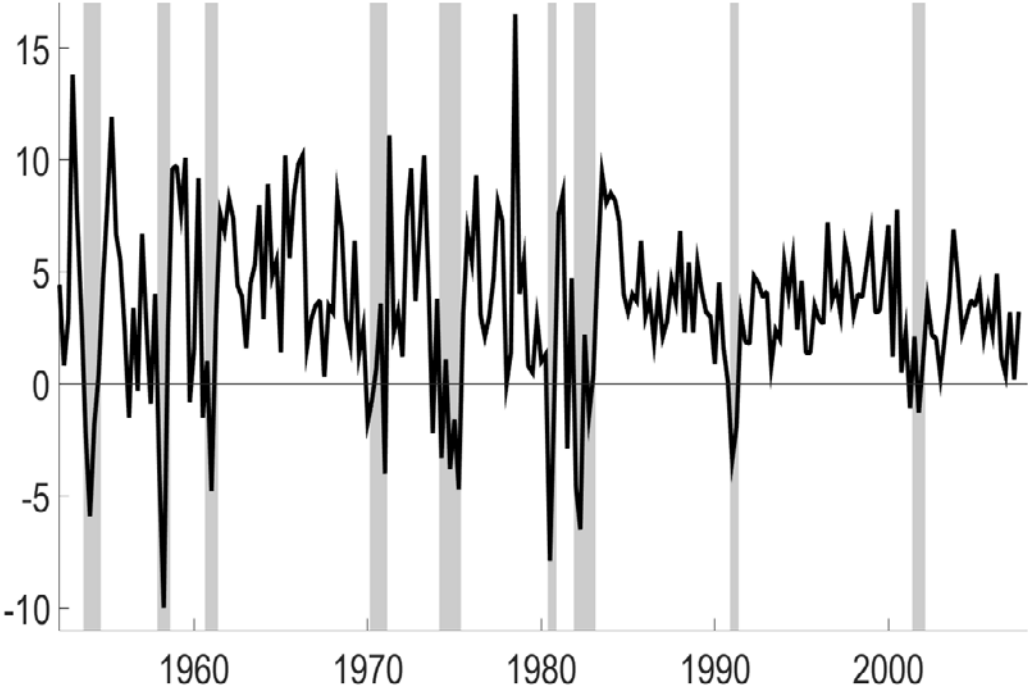
$$\text{Markovian Switching: } S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t,$$

$$\begin{bmatrix} \varepsilon_t \\ \omega_t \end{bmatrix} \sim i.i.d N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

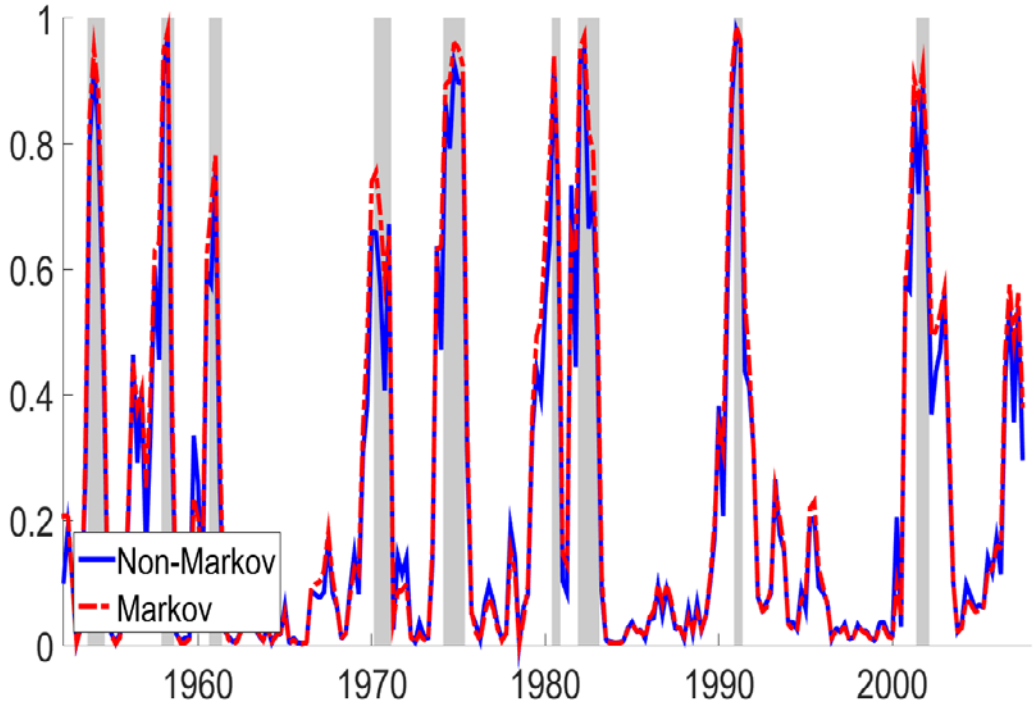
		<i>Non-Markov Switching</i>				<i>Markov Switching</i>					
Prior		Posterior				Posterior					
		Mean	Mode	(SD)	90 % HPDI		Mean	Mode	(SD)	90 % HPDI	
$\beta_0$	$N(0, 2^2)$	0.26	0.26	(0.04)	0.19	0.34	0.25	0.27	(0.05)	0.17	0.34
$\beta_1$	$N(0, 2^2)$	-0.09	-0.09	(0.02)	-0.13	-0.04	-0.09	-0.09	(0.02)	-0.13	-0.04
$\sigma_0^2$	$IG(1, 1)$	2.14	2.12	(0.13)	1.92	2.37	2.69	2.55	(0.18)	2.39	3.01
$\sigma_1^2$	$IG(1, 2)$	12.80	12.64	(1.21)	10.82	14.81	16.68	15.90	(1.98)	13.43	19.88
$\rho$	$TN(0, 1^2)$	-0.68	-0.68	(0.06)	-0.79	-0.57	-0.45	-0.46	(0.09)	-0.60	-0.29
$\alpha$	$N(-1, 2^2)$	-2.06	-1.71	(1.02)	-3.62	-0.40	-	-	-	-	-
$\psi$	$N(0.9, 1^2)$	0.97	0.97	(0.01)	0.95	0.98	-	-	-	-	-
$\alpha_0$	$N(-1.5, 2^2)$	-	-	-	-	-	-2.13	-2.09	(0.13)	-2.36	-1.91
$\alpha_1$	$N(1, 2^2)$	-	-	-	-	-	1.43	1.42	(0.15)	1.17	1.68
WAIC		2981.47				3014.45					

- Note:
1. Total iterations / Burn-in = 20,000 / 70,000
  2. TN(.) represents a truncated normal distribution defined between -1 and 1.
  3. WAIC =  $-\sum_{t=1}^T \log(\bar{p}d_t) + \sum_{t=1}^T V_R(\log(pd_t))$  where  $\bar{p}d_t$  is the MCMC sample average of predictive density  $f(y_t | \theta^{(r)}, S_t^{(r)})$  and  $V_R(\log(pd_t))$  is the MCMC sample variance of  $\log(f(y_t | \theta^{(r)}, S_t^{(r)}))$ , where  $\theta^{(r)}$  is the vector of the parameters drawn at the r-th iteration. A model with a lower WAIC is preferred in a model comparison.

Figure 1. Quarterly Real GDP Growth Rate [1952Q1-2007Q2]



**Figure 2.A. Posterior Probability of Recession from Markovian and non-Markovian Exogenous Switching Models of Business Cycle [Real GDP Growth: 1952Q1-2007Q2]**





**Figure 2.B. Posterior Mean of  $S_t^*$  from Markovian and non-Markovian Exogenous Switching Models of Business Cycle [Real GDP Growth: 1952Q1-2007Q2]**

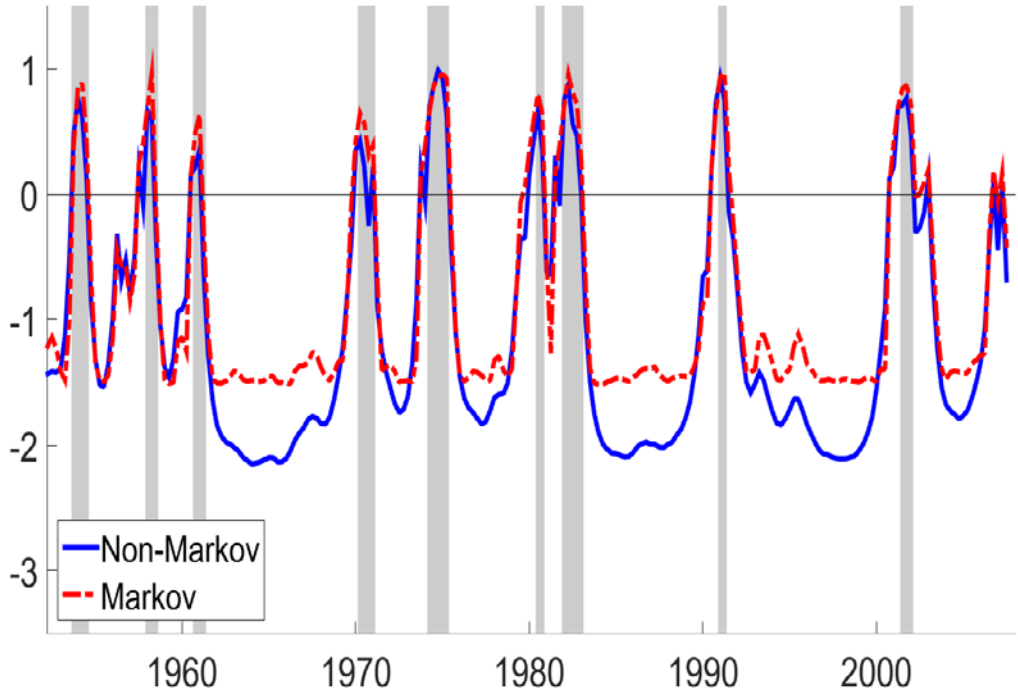
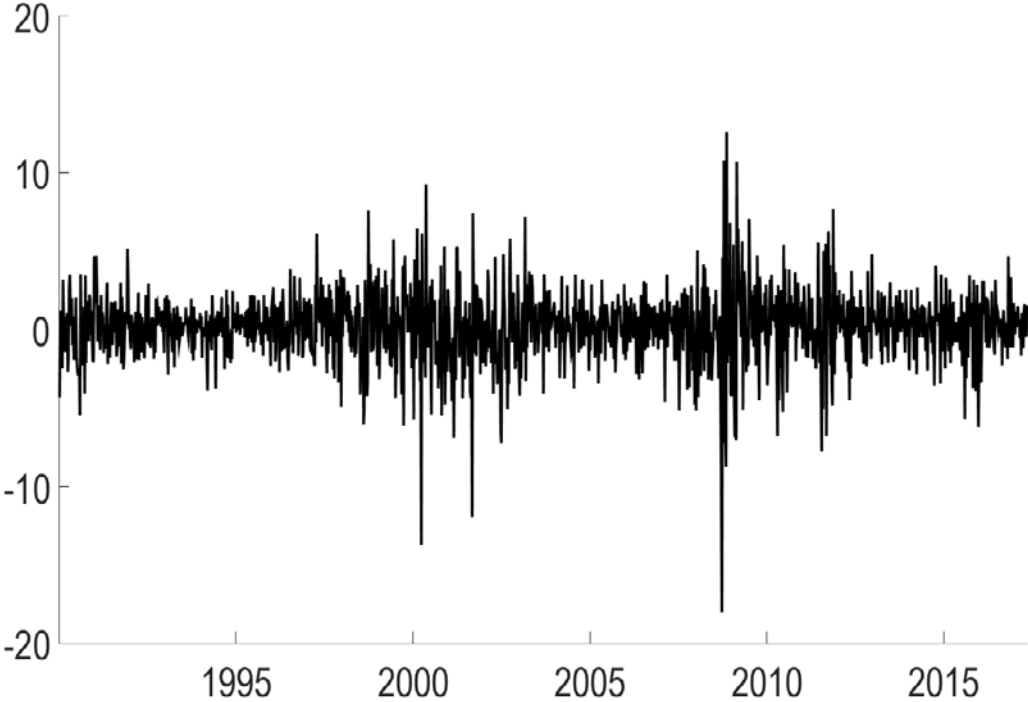
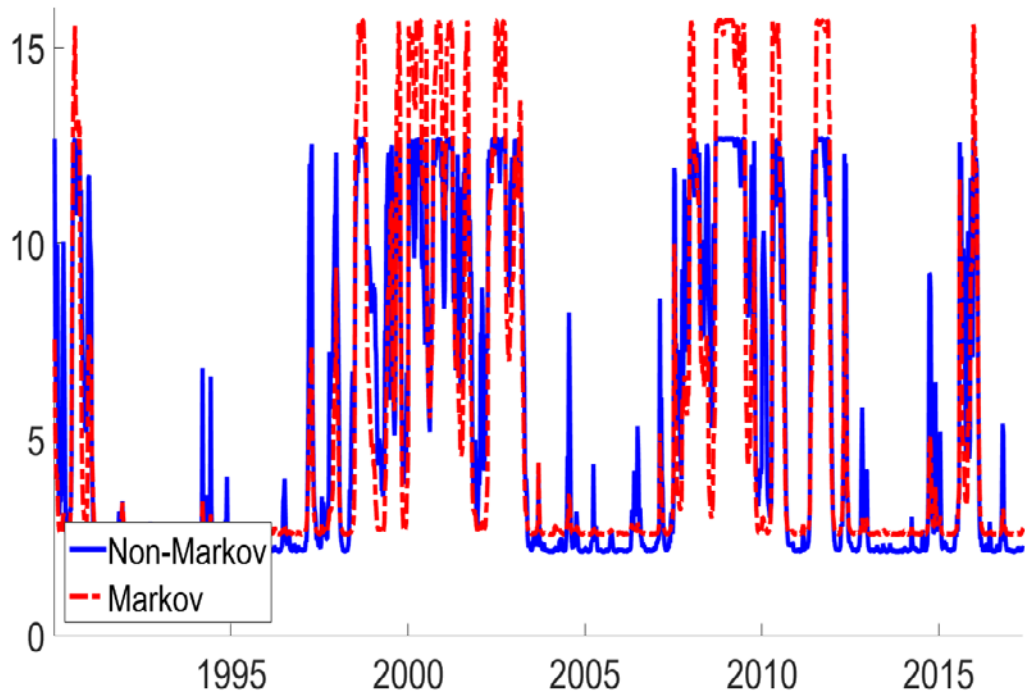


Figure 3. Weekly Excess Stock Returns [Jan, 1990 - May, 2017]



**Figure 4.A. Posterior Mean of the Volatility from Markovian and non-Markovian Endogenous Switching Models [Excess Returns: Jan, 1990 - May, 2017]**



**Figure 4.B. Posterior Mean of  $S_t^*$  from Markovian and non-Markovian Endogenous Switching Models of Volatility [Excess Returns: Jan, 1990 - May, 2017]**

