Journal of Economic Theory and Econometrics, Vol. 34, No. 4, Dec. 2023, 115–148

Non-Markovian Regime-Switching Models*

Chang-Jin Kim† Jaeho Kim‡

Abstract To date, almost all extensions and applications of Hamilton's (1989) regime switching model have been based on the assumption that the latent regime-indicator variable follows a Markovian switching process. This paper doubts the universal validity of this assumption and develops an MCMC algorithm for estimation of the non-Markovian regime switching model which employs an autoregressive continuous latent variable in specifying the dynamics of the discrete latent regime-indicator variable within the Probit specification. We show that, in spite of the non-Markovian nature of the discrete regime indicator variable, the Markovian property of this continuous latent variable allows us to successfully estimate the model. Our empirical results suggest that, for modeling volatility of the stock return, the non-Markovian switching model is strongly preferred to the Markovian switching model. However, for modeling the regime-switching nature of the business cycle based on real GDP, the convention of assuming Markovian switching seems to be valid.

Keywords Non-markovian regime switching, Markovian regime switching, exogenous switching, endogenous switching

JEL Classification C11, C22

*We are deeply grateful to the two anonymous reviewers for their thoughtful comments and suggestions. This work was supported by the Sogang Research Support Grant (Grant # 202310032.01).

†Department. of Economics, University of Washington; changjin@uw.edu

‡Corresponding author: School of Economics, Sogang University; jaehoecon@sogang.ac.kr

Submitted November 21, 2023, Accepted December 21, 2023

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 Markovian switching process.[1] 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 modelling 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). Dueker (2005), 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. Our study is closely related with Dueker (2005), Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) in that our econometric approach enables estimation of unobserved non-Markovian binary processes.

To date, little attention has been paid to regime switching models with non-Markovian latent regime indicator variables, except in Chib and Dueker (2004)[2]. We revisit the following version of the non-Markovian regime switching model

[1] Albert and Chib (1993a) is the first paper that presents Bayesian estimation of a Markov-switching model in which the pdf of the dependent variable (y_t) depends on not just the current state S_t but previous values of S_t as well. For an overview of econometric analysis of time series that are subject to changes in regime, readers are referred to Hamilton (2016). For the economic applications of the Markov regime switching model, readers are referred to Yang *et al.* (2019), Choi (2019), and Jeong *et al.* (2022).

[2] To estimate the model, Chib and Dueker (2004) cast equations (1), (2) and (3) into a state-space model and employ linear approximation to the non-linear measurement equation with the extended Kalman filter. 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 only valid under the assumption of normality. We have confirmed that their estimation method does not work for our simulation and empirical applications.

considered by them:

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

where S_t is a latent regime-indicator variable. The transitional dynamics of S_t is given by the following Probit specification:

$$
\begin{aligned}
& S_t = 1[S_t^* \geq 0], \\
& S_t^* = \alpha(1-\psi) + \psi S_{t-1}^* + \omega_t, \ \ \omega_t \sim i.i.d.N(0,1),
\end{aligned} \tag{2}
$$

where $1[.]$ is the indicator function; $|\psi| < 1$; and S_t^* is a continuous latent variable with $E(S_t^*) = \alpha$.[3] The joint distribution of ε_t and ω_t is specified as:[4]

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

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}^*, carries 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}^*.

The crucial feature of the Non-Markovian regime-switching model lies in its ability to capture varying durations of regimes, unlike the conventional Markov-switching model. The latter assumes a constant expected duration for each

[3] In this study, we have not considered higher order Auto-Regressive (AR) processes due to the challenges in identifying AR coefficients. Accurate identification typically requires a substantial number of regime switches within the data, a scenario that is relatively rare in the fields of economics and finance.

[4] When equation (2) is replaced by

$$
\begin{aligned}
& S_t = 1[S_t^* \geq 0], \\
& S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t, \ \ \omega_t \sim i.i.d.N(0,1),
\end{aligned}
$$

we have a first-order Markov switching model with endogenous switching (Kim *et al.*, 2008).

regime, solely reliant on the immediate past state, due to the time-invariant nature of its transition probabilities.[5] In contrast, the Non-Markovian model introduces a dynamic layer through S^*_{t-1}, a cumulative sum of past shocks that encodes the entire regime switching history. This accumulated shock value then dynamically modulates the expected duration of the current regime. For instance, a low positive value of S^*_{t-1} suggests a brief stay in regime 1, while a large positive value implies that the expected duration would be much longer. This flexibility makes the Non-Markovian model a natural choice for capturing temporal variations in regime durations.

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.[6] 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.

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. Derivation of the full conditional distribution for S^*_t extends the work of Albert and Chib (1993b), who present MCMC methods for static Probit models. 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 modelling of postwar real GDP and the volatility modelling of weekly stock returns.

Our proposed estimation algorithm for regime dynamics relies on a single-move sampling approach, where latent variables are drawn one at a time. Given that this single-move sampler conditions the latent variables on other time periods, it is typically characterized by a slower convergence speed. To address this concern, we recommend using convergence diagnostics such as those pro-

[5]For example, if $S_{t-1} = 1$ and the corresponding regime transition probability $\Pr[S_t = 1|S_{t-1} = 1, S_{t-2}, ..., S_1] = \Pr[S_t = 1|S_{t-1} = 1] = 0.99$, the expected duration of regime 1 is 100, calculated as $\frac{1}{1-0.99}$.

[6]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.

posed by Geweke (1992), Gelman and Rubin (1992), and Raftery and Banfield (1991) to ensure satisfactory convergence. In our empirical applications, we specifically utilize the Potential Scale Reduction Factor (PSRF) from Gelman and Rubin (1992) as our convergence diagnostic tool.

While the proposed model successfully captures the dynamics of two regimes, extending it to a general N-regime case requires a more sophisticated framework. The multinomial Probit specification of Hwu *et al.* (2021) offers a promising approach. However, incorporating this framework would introduce additional complexity in both parameter estimation and computational demands due to the presence of multiple continuous latent variables influencing the discrete regime-indicator variable. Therefore, exploring this extension remains an open question for future studies.

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 works 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 the second equation in (2) as

$$S_t^* = \alpha + \omega_t + \psi\omega_{t-1} + \psi^2\omega_{t-2} + \ldots, \tag{4}$$

where, due to equation (2), $Pr[S_{t-j} = 1|S_{t-j-1}^*]$ is positively related to ω_{t-j} for $j = 0, 1, 2, \ldots$. 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

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

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 $\rho = 0$:

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

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

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

As the joint distribution of S_t and S_{t-1}^* conditional on past information depends on all the history of past regimes (S_{t-1}, S_{t-2},..., S_0), evaluation of equation (7) is intractable within the classical framework.[7]

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 $\tilde{S}_T = \begin{pmatrix} S_0 & S_1 & S_2 & \dots & S_T \end{pmatrix}'$ and $\tilde{S}_T^* = \begin{pmatrix} S_0^* & S_1^* & S_2^* & \dots & S_T^* \end{pmatrix}'$.

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:

$$f(\tilde{\beta}, \tilde{\sigma}^2, \tilde{A}, \tilde{S}_T^*, \tilde{S}_T|\tilde{Y}_T) = f(\tilde{\beta}, \tilde{\sigma}^2|\tilde{A}, \tilde{S}_T^*, \tilde{S}_T, \tilde{Y}_T) f(\tilde{A}|\tilde{S}_T^*, \tilde{S}_T, \tilde{Y}_T) f(\tilde{S}_T^*, \tilde{S}_T|\tilde{Y}_T) \tag{8}$$
$$= f(\tilde{\beta}, \tilde{\sigma}^2|\tilde{S}_T, \tilde{Y}_T) f(\tilde{A}|\tilde{S}_T^*) f(\tilde{S}_T^*, \tilde{S}_T|\tilde{Y}_T),$$

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

---

[7] 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

$$f(S_t|I_{t-1}) = \sum_{S_{t-1}} f(S_t, S_{t-1}|I_{t-1})$$
$$= \sum_{S_{t-1}} f(S_t|S_{t-1}) f(S_{t-1}|I_{t-1}),$$

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).

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^*, $t = 1,2,...,T$, from the following full conditional distribution:

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

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^*.

Equation (9) allows us to take advantage of the Markovian property of S_t^*. Furthermore, the decompositions in equations (8) and (9) lead to the following steps for the MCMC procedure, which can be repeated until convergence is achieved:[8]

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 $t = 1,2,...,T$. 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.

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,

[8] Hereafter, we suppress the model parameters in the full conditional distributions associated with S_t and/or S_t^* for notational simplification.

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 $\phi(.)$ 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 $1[.]$ denoting the indicator function;

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

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}:

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

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

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

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

$$
\begin{aligned}
& f(S_t = 1|\tilde{S}^*_{\neq t}, \tilde{S}_{\neq t}, \tilde{Y}_T) \\
& \propto \int_{-\infty}^{\infty} \left[\frac{1}{\sigma_1}\phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right)\right]\left[\frac{1}{\sqrt{V}}\phi\left(\frac{S^*_t - \mu_t}{\sqrt{V}}\right)\right] f(S_t = 1|S^*_t)dS^*_t \qquad (14)\\
& = \int_{0}^{\infty} \left[\frac{1}{\sigma_1}\phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right)\right]\left[\frac{1}{\sqrt{V}}\phi\left(\frac{S^*_t - \mu_t}{\sqrt{V}}\right)\right] dS^*_t \\
& = \frac{1}{\sigma_1}\phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right)\Phi\left(\frac{\mu_t}{\sqrt{V}}\right),
\end{aligned}
$$

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

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

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:

$$
\begin{aligned}
f(S^*_t|\tilde{S}^*_{\neq t}, \tilde{S}_{\neq t}, S_t, \tilde{Y}_T) & \propto f(S^*_t, S_t|\tilde{S}^*_{\neq t}, \tilde{S}_{\neq t}, \tilde{Y}_T) \qquad (16)\\
& \propto \left[\frac{1}{\sigma_{S_t}}\phi\left(\frac{y_t - x_t'\beta_{S_t}}{\sigma_{S_t}}\right)\right]\left[\frac{1}{\sqrt{V}}\phi\left(\frac{S^*_t - \mu_t}{\sqrt{V}}\right)\right] f(S_t|S^*_t) \\
& \propto \frac{1}{\sqrt{V}}\phi\left(\frac{S^*_t - \mu_t}{\sqrt{V}}\right) f(S_t|S^*_t),
\end{aligned}
$$

which suggests that we can generate S^*_t from the following truncated normal distribution:

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

where μ_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 ω_t as a function ε_t (i.e., $\omega_t = \rho\varepsilon_t + \sqrt{(1-\rho^2)}\omega_t^*$), so that we can rewrite equation (2) as:

$$S_{t+j}^* = \alpha(1-\psi) + \psi S_{t+j-1}^* + \rho\varepsilon_{t+j} + \sqrt{1-\rho^2}\omega_{t+j}^*, \tag{18}$$

where $\omega_{t+j}^* \sim i.i.d.N(0,1)$, $E(\omega_{t+j}^*\varepsilon_{t+j}) = 0$, and $\varepsilon_{t+j} = \frac{y_{t+j}-x_{t+j}'\beta_{S_{t+j}}}{\sigma_{S_{t+j}}}$.

We note that, unlike in the case of exogenous regime switching, derivation of the full conditionals for $\tilde{A} = \begin{pmatrix}\alpha & \psi & \rho\end{pmatrix}'$, $\tilde{\beta} = \begin{pmatrix}\beta_0 & \beta_1\end{pmatrix}'$, and $\tilde{\sigma}^2 = \begin{pmatrix}\sigma_0^2 & \sigma_1^2\end{pmatrix}'$ 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.[9] 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 ε_{t+1}, which is a function of y_{t+1} and S_{t+1}, as well 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:

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

[9]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.

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

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

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

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

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

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

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:

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

where

$$V = \frac{1-\rho^2}{1+\psi^2};\ \textit{and}\ \ \mu_t = \alpha+\frac{1}{1+\psi^2}(\psi(S^*_{t+1}-\alpha)-\rho\psi\varepsilon_{t+1}+\psi(S^*_{t-1}-\alpha)+\rho\varepsilon_t), \quad (24)$$

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

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

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

$$f(S_t = 1|\tilde{S}^*_{\neq t}, \tilde{S}_{\neq t}, \tilde{Y}_T) \tag{28}$$
$$\propto \int_{-\infty}^{\infty} \left[\frac{1}{\sigma_1}\phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right)\right]\left[g(\varepsilon_t(S_t = 1))\frac{1}{\sqrt{V}}\phi\left(\frac{S_t^* - \mu_t}{\sqrt{V}}\right)\right] f(S_t = 1|S_t^*)dS_t^*$$
$$= \int_{0}^{\infty} \left[\frac{1}{\sigma_1}\phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right)\right]\left[g(\varepsilon_t(S_t = 1))\frac{1}{\sqrt{V}}\phi\left(\frac{S_t^* - \mu_t}{\sqrt{V}}\right)\right] dS_t^*$$
$$= \left[\frac{1}{\sigma_1}\phi\left(\frac{y_t - x_t'\beta_1}{\sigma_1}\right)\right]\left[g(\varepsilon_t(S_t = 1))\Phi\left(\frac{\mu_t}{\sqrt{V}}\right)\right]$$

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

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

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). Conditional on S_t, the first term on the right-hand-side of equation (26) and the $g(\varepsilon_t(S_t))$ term are a part of the normalizing constant. We have the following result:

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

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

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

where μ_t and V are given in equation (24).

3.3. GENERATING $\tilde{A} = (\alpha, \psi, \rho)'$ CONDITIONAL ON $\tilde{\beta}$, $\tilde{\sigma}^2$, $\tilde{S}_T^*$, $\tilde{S}_T$, AND $\tilde{Y}_T$

Generating ρ conditional on α, ψ, $\tilde{\beta}$, $\tilde{\sigma}^2$, $\tilde{S}_T^*$, $\tilde{S}_T$, and $\tilde{Y}_T$

The full conditional density from which ρ is drawn can be derived as:[10]

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

where $f(\rho)$ is the prior density of ρ; 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 ρ parameter suppressed. By substituting equations (20) and (21) into equation (32), we obtain the following target density of ρ:

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

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

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

[10]For the sake of notational brevity, we suppress α, ψ, $\tilde{\beta}$, and $\tilde{\sigma}^2$ in the conditional density of ρ.

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

For the Metropolis-Hastings algorithm, let ρ_o denote the accepted ρ at the previous MCMC iteration and ρ_n is a newly generated candidate from the following random walk candidate generating distribution:

$$\rho_n = \rho_o + \varepsilon, \;\; \varepsilon \sim N(0,c)_{1[-\rho_0-1<\varepsilon<-\rho_0+1]}, \tag{35}$$

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

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

Generating α conditional on ρ, ψ, $\tilde{\beta}$, $\tilde{\sigma}^2$, $\tilde{S}_T$, $\tilde{S}_T^*$, and $\tilde{Y}_T$

Rearranging equation (18), we have

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

and drawing α from the full conditional distribution derived based on this equation is standard.

Generating ψ conditional on ρ, α, $\tilde{\beta}$, $\tilde{\sigma}^2$, $\tilde{S}_T$, $\tilde{S}_T^*$, and $\tilde{Y}_T$

Rearranging equation (18), we have

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

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 ψ 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 ε_t and ω_t in equation (3), we can write ε_t as a function of ω_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:

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

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

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

$y_t^* = y_t - \sigma_{S_t}\rho\omega_t$. Considering the shock ω_t in this MCMC step is crucial as emphasized by Kim et al. (2008) because the regime switching β_{S_t} will be correlated with the error term ε_t, which causes an endogeneity issue without controlling for ω_t.

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

As ρ is irrelevant in making inference on $\tilde{\sigma}^2$, this step is based on equation (1). To further explain the validity of this MCMC step, we rewrite the joint distribution of the two variables as

$$\begin{pmatrix} y_t \\ S_t^* \end{pmatrix} \mid S_t, \tilde{S}_{t-1}^* \sim N\left(\begin{pmatrix} x_t\beta_{S_t} \\ S_{t-1}^* \end{pmatrix}, \begin{pmatrix} \sigma_{S_t}^2 & \rho\sigma_{S_t} \\ \sigma_{S_t} & 1 \end{pmatrix}\right). \tag{41}$$

Because the error terms can be obtained, their joint distribution is easily specified as

$$\begin{pmatrix} e_t \\ \omega_t \end{pmatrix} \mid S_t, \tilde{S}_{t-1}^* \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{S_t}^2 & \rho\sigma_{S_t} \\ \sigma_{S_t} & 1 \end{pmatrix}\right) \tag{42}$$

whose marginal distribution of $e_t = \sigma_{S_t}\varepsilon_t$ is simply $N(0,\sigma_{S_t}^2)$. Based on the idea of collapsed Gibbs sampling that integrates out S_t^* for all time periods, we use the marginal distribution of e_t instead of the joint distribution of e_t and ω_t when sampling $\sigma_{S_t}^2$. Thus, conditional on $\tilde{\beta}$, it is straightforward to generate $\tilde{\sigma}^2$ from an appropriate posterior distribution. Alternatively, we can employ the conditional density of $\tilde{S}_t^*$ conditioning on ε_t in equation (39) for the posterior sampling step. In this case, the posterior sampling should be performed by an Metropolis-Hastings algorithm. If we adopt a random walk candidate generating distribution as for ρ, the acceptance probability is given by

$$\alpha(\sigma_{i,n}^2,\sigma_{i,o}^2) = min\{1, \frac{\prod_{t=1}^T [f(y_t|y_{t-1},\omega_t,\sigma_{i,n}^2,\rho)]f(\sigma_{i,n}^2)}{\prod_{t=1}^T [f(y_t|y_{t-1},\omega_t,\sigma_{i,o}^2,\rho)]f(\sigma_{i,o}^2)}\}. \tag{43}$$

for $i = 0,1$ where $f(y_t|y_{t-1}, \omega_t, \sigma_i^2, \rho)$ is the observation density based on equation (39) and $f(\sigma_{i,n}^2)$ is the prior density. The support of the prior distribution is $\sigma_{i,n}^2 > 0$. We have confirmed that all simulation and empirical results from the two described algorithms are identical. In the following sections, we use the the Metropolis-Hastings algorithm to draw the posterior samples of $\sigma_{i,n}^2$.

4. SIMULATION STUDY

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:

$$\begin{aligned} & y_t = \beta_0(1-S_t) + \beta_1 S_t + (\gamma_0(1-S_t) + \gamma_1 S_t)x_t + (\sigma_0(1-S_t) + \sigma_1 S_t)\varepsilon_t, \quad (44) \\ & S_t = 1[S_t^* \geq 0], \;\; S_t^* = \alpha(1-\psi) + \psi S_{t-1}^* + \omega_t, \\ & \begin{pmatrix} \varepsilon_t \\ \omega_t \end{pmatrix} \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \end{aligned}$$

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

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

Case #1: [Exogenous Switching, No Exogenous Variable]

$$\rho = 0;\; \beta_0 = 1;\; \beta_1 = -1;\; \gamma_0 = 0;\; \gamma_1 = 0;\; \sigma_0^2 = 0.5;\; \sigma_1^2 = 1;\; \psi = 0.9;\; \alpha = 0$$

Case #2: [Endogenous Switching, No Exogenous Variable]

$$\rho = 0.9;\; \beta_0 = 1;\; \beta_1 = -1;\; \gamma_0 = 0;\; \gamma_1 = 0;\; \sigma_0^2 = 0.5;\; \sigma_1^2 = 1;\; \psi = 0.9;\; \alpha = 0$$

Case #3: [Endogenous Switching, No Exogenous Variable]

$$\rho = 0.9;\; \beta_0 = 1;\; \beta_1 = -1;\; \gamma_0 = 0;\; \gamma_1 = 0;\; \sigma_0^2 = 0.5;\; \sigma_1^2 = 1;\; \psi = 0.98;\; \alpha = 0$$

Case #4: [Endogenous Switching, Exogenous Variable]

$$\rho = 0.9;\ \beta_0 = 1;\ \beta_1 = -1;\ \gamma_0 = -1;\ \gamma_1 = 1;\ \sigma_0^2 = 0.5;\ \sigma_1^2 = 1;\ \psi = 0.9;\ \alpha = 0$$

All cases except Case #4 do not consider any exogenous variable. By comparing Case #2 and Case #3, we can test for how the persistence of the latent AR(1) variable influences the performance of the proposed algorithm. By comparing Case #2 and Case #4, we can test for how the presence of the exogenous variable x_t influences the performance of the proposed algorithm.

For each case, we generate 500 samples of 500 or 1,000 observations. For estimation, we apply both the non-Markovian and Markovian switching models. For each model and for each case, we obtain the sampling distribution of the posterior mean for each parameter. Moreover, we compute a Bayesian model selection criterion, the Watanabe-Akaike Information Criterion (WAIC) developed by Watanabe (2010) for each generated sample to see if WAIC selects a right model.

In this simulation exercise, we adopt weakly informative priors for all parameters. For the parameters β_{S_t} and γ_{S_t}, we assume $N(0,5^2)$. The prior for $\sigma^2_{S_t}$ is $IG(3, 3\times 0.2)$ where 3 is the shape parameter and 3×0.2 is the scale parameter. For the Markov regime-switching model, we set the priors for the parameters α_0 and α_1 to follow normal distributions, $N(-1.5,1)$ and $N(1.5,1)$, respectively. In the non-Markovian regime-switching model, the parameters α and Ψ are assumed to follow a normal distribution, $N(0,1)$, and a truncated normal distribution, $TN(0.9, 0.1^2)$ over the interval $[-1,1]$, respectively. It is important to note that the prior for ψ is not informative because any dynamics of S_t observed in real data are likely to be captured within the range $[0.8,1]$. Lastly, the prior for the correlation parameter ρ is assumed to be a truncated normal distribution defined between -1 and 1.

Table 1 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 noticeable 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 $T = 1000$. As shown in the last row of Table 1, WAIC selects a right model with high probabilities.

Table 2 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

Table 1: Sampling Distributions of Posterior Mean based on Markovian and Non-Markovian Models [DGP: Exogenous Switching, No Exogenous Variable]

$$y_t = \beta_{S_t} + \sigma_{S_t}\varepsilon_t, \quad E[\varepsilon_t] = 0,\ E[\varepsilon_t \cdot \omega_t] = 0$$

$$\text{Non-Markovian Switching (M1):} \quad S_t^* = \alpha(1-\psi) + \psi S_{t-1}^* + \omega_t,$$

$$\text{Markovian Switching (M2):} \quad S_t^* = \alpha_{s-1} + \omega_t,$$

Parameter	True Value	Estimated Model			
		Non-Markovian Switching		Markovian Switching	
		$T = 500$	$T = 1,000$	$T = 500$	$T = 1,000$
β_0	1	1.00 (0.06)	1.00 (0.04)	0.99 (0.07)	1.00 (0.04)
β_1	-1	-0.96 (0.11)	-1.00 (0.07)	-0.88 (0.12)	-0.93 (0.07)
σ_0^2	0.5	0.51 (0.06)	0.50 (0.04)	0.53 (0.07)	0.52 (0.05)
σ_1^2	1	1.06 (0.13)	1.03 (0.10)	1.20 (0.15)	1.15 (0.11)
α	0	0.07 (0.35)	0.05 (0.37)	-	-
ψ	0.9	0.91 (0.03)	0.91 (0.03)	-	-
α_0	-	-	-	-1.28 (0.16)	-1.27 (0.13)
α_1	-	-	-	1.37 (0.17)	1.33 (0.15)
WAIC	Pr(M1 is preferred) $= 0.88(T = 500), 0.94(T = 1,000)$				

Table 2: Sampling Distributions of Posterior Mean based on Markovian and Non-Markovian Models [DGP: Endogenous Switching, No Exogenous Variable]

$$y_t = \beta_{S_t} + \sigma_{S_t}\varepsilon_t, \quad E[\varepsilon_t] = 0,\ E[\varepsilon_t \cdot \omega_t] = \rho$$

$$\text{Non-Markovian Switching (M1):} \quad S_t^* = \alpha(1-\psi) + \psi S_{t-1}^* + \omega_t,$$

$$\text{Markovian Switching (M2):} \quad S_t^* = \alpha_{s-1} + \omega_t,$$

Parameter	True Value	Estimated Model			
		Non-Markovian Switching		Markovian Switching	
		$T = 500$	$T = 1,000$	$T = 500$	$T = 1,000$
β_0	1	0.97 (0.04)	0.98 (0.02)	0.85 (0.10)	0.87 (0.02)
β_1	-1	-0.98 (0.09)	-0.98 (0.03)	-0.80 (0.10)	-0.78 (0.03)
σ_0^2	0.5	0.51 (0.06)	0.52 (0.02)	0.50 (0.10)	0.47 (0.02)
σ_1^2	1	1.02 (0.10)	1.01 (0.04)	1.00 (0.10)	0.99 (0.04)
α	0	-0.09 (0.43)	0.00 (0.20)	-	-
ψ	0.9	0.91 (0.02)	0.90 (0.01)	-	-
ρ	0.9	0.90 (0.04)	0.90 (0.01)	0.57 (0.20)	0.52 (0.05)
α_0	-	-	-	-1.39 (0.17)	-1.38 (0.07)
α_1	-	-	-	1.38 (0.12)	1.41 (0.08)
WAIC	Pr(M1 is preferred) $= 0.99(T = 500), 1(T = 1,000)$				

Table 3: Sampling Distributions of Posterior Mean based on Markovian and Non-Markovian Models [DGP: Endogenous Switching, No Exogenous Variable]

$$y_t = \beta_{S_t} + \sigma_{S_t}\varepsilon_t, \quad E[\varepsilon_t] = 0, \; E[\varepsilon_t \cdot \omega_t] = \rho$$

$$\text{Non-Markovian Switching (M1):} \quad S_t^* = \alpha(1-\psi) + \psi S_{t-1}^* + \omega_t,$$

$$\text{Markovian Switching (M2):} \quad S_t^* = \alpha_{s-1} + \omega_t,$$

Parameter	True Value	Estimated Model: Non-Markovian Switching $T=500$	Non-Markovian Switching $T=1{,}000$	Markovian Switching $T=500$	Markovian Switching $T=1{,}000$
β_0	1	0.98 (0.03)	0.99 (0.02)	0.91 (0.06)	0.67 (0.43)
β_1	-1	-1.01 (0.05)	-0.98 (0.04)	-0.83 (0.17)	-0.68 (0.33)
σ_0^2	0.5	0.50 (0.05)	0.50 (0.05)	0.57 (0.27)	0.74 (0.39)
σ_1^2	1	0.99 (0.11)	0.99 (0.08)	1.13 (0.25)	1.26 (0.29)
α	0	-0.14 (0.50)	0.06 (0.55)	-	-
ψ	0.98	0.98 (0.01)	0.98 (0.01)	-	-
ρ	0.9	0.85 (0.05)	0.84 (0.04)	0.43 (0.37)	0.51 (0.32)
α_0	-	-	-	-2.01 (0.45)	-1.69 (0.78)
α_1	-	-	-	1.93 (0.70)	2.08 (0.85)
WAIC	$\Pr(\text{M1 is preferred}) = 0.99(T=500), 1(T=1{,}000)$				

for a misspecified model show considerable bias in some parameter estimates regardless of the sample size. For example, the estimates of the ρ parameter considerably underestimate the true value when a Markovian switching model is employed. While the true value of ρ 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, ρ 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. As for Case #1, WAIC works well for model comparison.

In Table 3 of Case #3, our analysis shifts focus to a scenario with a more persistent S_t^*, resulting in regimes of greater persistence and the reduced number of regime switches for a given sample size. Within this DGP, the superior efficacy of the non-Markovian regime-switching model is consistently observed. It's important to note that assuming the regimes follow a first-order Markov process leads to significantly biased estimates of the model parameters. Interestingly, even with a large sample size, this bias issue remains unresolved in the Markov

Table 4: Sampling Distributions of Posterior Mean based on Markovian and Non-Markovian Models [DGP: Endogenous Switching, Exogenous Variable]

$$y_t = \beta_{S_t} + x_t \gamma_{S_t} + \sigma_{S_t} \varepsilon_t, \quad E[\varepsilon_t] = 0, \; E[\varepsilon_t \cdot \omega_t] = \rho$$

$$\text{Non-Markovian Switching (M1):} \quad S_t^* = \alpha(1-\psi) + \psi S_{t-1}^* + \omega_t,$$

$$\text{Markovian Switching (M2):} \quad S_t^* = \alpha_{s-1} + \omega_t,$$

Parameter	True Value	Estimated Model			
		Non-Markovian Switching		Markovian Switching	
		$T = 500$	$T = 1,000$	$T = 500$	$T = 1,000$
β_0	1	0.98 (0.04)	0.99 (0.04)	0.96 (0.06)	0.96 (0.05)
β_1	-1	-0.98 (0.07)	-0.98 (0.05)	-0.95 (0.09)	-0.94 (0.06)
γ_0	1	0.99 (0.03)	1.00 (0.02)	1.00 (0.03)	1.01 (0.02)
γ_1	-1	-0.99 (0.03)	-0.99 (0.03)	-0.98 (0.04)	-1.00 (0.03)
σ_0^2	0.5	0.51 (0.06)	0.53 (0.05)	0.48 (0.05)	0.50 (0.04)
σ_1^2	1	1.01 (0.10)	1.03 (0.07)	1.00 (0.10)	1.00 (0.07)
α	0	-0.08 (0.38)	-0.04 (0.31)	-	-
ψ	0.9	0.91 (0.02)	0.90 (0.02)	-	-
ρ	0.9	0.90 (0.04)	0.91 (0.02)	0.66 (0.09)	0.65 (0.07)
α_0	-1.5	-	-	-1.18 (0.19)	-1.15 (0.14)
α_1	1.2	-	-	1.15 (0.19)	1.14 (0.14)
WAIC	$\Pr(\text{M1 is preferred}) = 0.99(T = 500), 1(T = 1,000)$				

regime-switching model.

In Case #4 where an additional exogenous variable x_t is incorporated, we obtain the same conclusion. The results of Table 4 reveal that the Non-Markovian regime-switching model exhibits marginally superior performance compared to its Markovian counterpart in a more generalized scenario. We think that the presence of additional regime-dependent parameters in the model, enriching the dataset with more information on regime changes, tends to diminish the impact of adopting different regime dynamics assumptions.

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

We evaluate the performance of regime-switching models using two real datasets: market stock returns and real GDP growth rates. The regime-switching models are estimated utilizing the proposed estimation algorithm. For the purpose of model comparison, we employ WAIC, as in the simulation section. A model is considered preferable if it exhibits a lower WAIC value.

Additionally, to verify satisfactory convergence of the MCMC chain, we report the Potential Scale Reduction Factor (PSRF) proposed by Gelman and Rubin (1992). A PSRF value close to 1 indicates stronger evidence of the MCMC chain successfully converging to the stationary posterior distribution, which is crucial for the reliability of our model estimates.

In our estimation process, we have deliberately chosen to use weakly informative priors for all model parameters. This approach is aimed at ensuring that the priors exert minimal influence on the Bayesian posterior estimates, thereby maintaining the objectivity of our analysis. The prior distributions are reported in the second column of each result table.

5.1. NON-MARKOVIAN VERSUS MARKOVIAN SWITCHING MODELS OF WEEKLY STOCK RETURN VOLATILITY

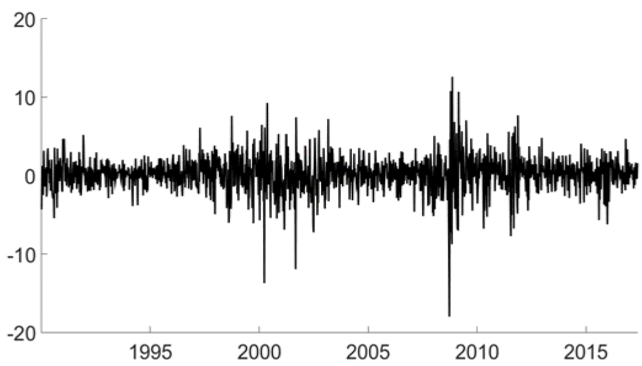

Figure 1: Weekly Excess Stock Returns [Jan, 1990 - May, 2017]

Recognizing abrupt changes in stock return volatility is crucial because financial markets rapidly integrate new information. Traditional stochastic volatility models, which assume gradual changes in volatility, might fail to capture dramatic shifts. Additionally, the duration of volatility regimes can vary over time, a characteristic not accounted for in conventional Markov switching models. For instance, economic events like wars, supply chain disruptions, and unforeseen monetary policies can impact financial markets for different durations. In this context, our proposed Non-Markovian regime switching model is more appropriately suited than the Markov regime switching model for capturing these dynamics.

In this section, we also consider the leverage effect within a regime-switching model of the stock return volatility. In the literature on stochastic volatility, asymmetry in the stock return is typically modeled by introducing the leverage effect. By denoting ζ_t as the innovation to the stock return volatility and ε_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 modelling the leverage effect.

By noting that ω_t in equation (2) is equivalent to ζ_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:

$$r_t = \beta_0 + \beta_1 r_{t-1} + (\sigma_0(1-S_t) + \sigma_1 S_t)\varepsilon_t, \qquad \sigma_0^2 < \sigma_1^2,$$

$$\text{Non-Markovian Switching: } S_t = 1[S_t^* \geq 0], \;\; S_t^* = \alpha(1-\psi) + \psi S_{t-1}^* + \omega_t,$$

$$\text{Markovian Switching: } S_t = 1[S_t^* \geq 0], \;\; S_t^* = \alpha_0 + \alpha_1 S_{t-1} + \omega_t,$$

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

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.[11] The sample for our study deliberately omits the period of the Covid pandemic and subsequent years. Incorporating this turbulent period necessitates specialized modeling tools beyond the scope of our current framework. Thus, we earmark the extension of our model to include this period as an avenue for future research because our study's primary objective is to develop a valid Bayesian estimation method for non-Markovian regime-switching models. The data are plotted in Figure 1.

[11] The excess return data are freely available at the data library of Kenneth R. French's home page.

Table 5: Bayesian Posterior Estimates of Exogenous Regime Switching Models [Weekly Excessive Return: 1990:M1-W1-2017:M5:W4]

Parameter	Prior Distribution	Non-Markov Switching Mean (SD)	Non-Markov Switching 90% HPDI	Non-Markov Switching PSRF	Markov Switching Mean (SD)	Markov Switching 90% HPDI	Markov Switching PSRF
β_0	$\mathcal{N}(0,2^2)$	0.28 (0.05)	[0.21, 0.36]	1.000	0.26 (0.05)	[0.18, 0.35]	1.000
β_1	$\mathcal{N}(0,2^2)$	-0.08 (0.03)	[-0.13, -0.04]	1.000	-0.08 (0.03)	[-0.13, -0.04]	1.000
σ_0^2	$IG(1,1)$	2.24 (0.15)	[2.00, 2.47]	1.005	2.60 (0.20)	[2.28, 2.93]	1.007
σ_1^2	$IG(1,2)$	13.29 (1.39)	[11.19, 15.54]	1.005	15.95 (2.23)	[12.60, 19.41]	1.009
α	$\mathcal{N}(-1,2^2)$	-2.39 (1.18)	[-4.27, -0.53]	1.011	-	-	-
ψ	$\mathcal{N}(0.9,1^2)$	0.98 (0.01)	[0.96, 0.99]	1.032	-	-	-
α_0	$\mathcal{N}(-1.5,2^2)$	-	-	-	-2.17 (0.15)	[-2.41, -1.92]	1.003
α_1	$\mathcal{N}(1,2^2)$	-	-	-	1.54 (0.18)	[1.25, 1.84]	1.008
WAIC		3001.457			3018.112		

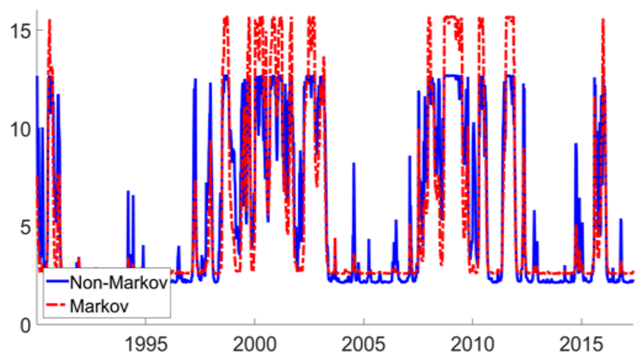

Figure 2: Posterior Mean of the Volatility from Markovian and non-Markovian Endogenous Switching Models [Excess Returns: Jan, 1990 - May, 2017]

Table 6: Bayesian Posterior Estimates of Endogenous Regime Switching Models [Weekly Excessive Return: 1990:M1-W1-2017:M5:W4]

	Prior	Non-Markov Switching			Markov Switching		
Parameter	Distribution	Mean (SD)	90% HPDI	PSRF	Mean (SD)	90% HPDI	PSRF
β_0	$\mathcal{N}(0,2^2)$	0.26 (0.04)	[0.19, 0.33]	1.001	0.26 (0.05)	[0.18, 0.34]	1.000
β_1	$\mathcal{N}(0,2^2)$	-0.09 (0.03)	[-0.13, -0.04]	1.001	-0.09 (0.03)	[-0.14, -0.05]	1.000
σ_0^2	$IG(1,1)$	2.12 (0.13)	[1.90, 2.34]	1.010	2.65 (0.20)	[2.33, 2.99]	1.007
σ_1^2	$IG(1,2)$	12.96 (1.21)	[11.06, 14.92]	1.008	17.36 (2.46)	[13.54, 21.23]	1.009
α	$\mathcal{N}(-1,2^2)$	-1.84 (0.77)	[-3.06, -0.60]	1.016	-	-	-
ψ	$\mathcal{N}(0.9,1^2)$	0.97 (0.01)	[0.96, 0.99]	1.005	-	-	-
ρ	$TN(0,1)$	-0.68 (0.07)	[-0.80, -0.58]	1.004	-0.43 (0.10)	[-0.60, -0.26]	-
α_0	$\mathcal{N}(-1.5,2^2)$	-	-	-	-2.10 (0.13)	[-2.31, -1.88]	1.003
α_1	$\mathcal{N}(1,2^2)$	-	-	-	1.40 (0.15)	[1.15, 1.66]	1.008
WAIC			2981.444			3017.501	

Table 5 and Table 6 report Bayesian estimates for four competing models, i.e., non-Markovian and Markovian switching models with endogenous switching ($\rho \neq 0$) and those with exogenous switching ($\rho = 0$). The reported WAICs in the tables indicate that endogenous switching or the leverage effect is an important feature of the stock return volatility. That is, the Bayesian model selection criterion favors models with endogenous switching or the leverage effect to those with exogenous switching. We thus focus our attention on comparing these two models with endogenous switching.

Table 6 reports that the posterior mean of ρ is -0.43 with a small posterior standard deviation (0.10) for the Markovian switching model, while it is -0.68 with a smaller posterior standard deviation (0.07) 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) is strongly preferred to the Markovian switching model (WAIC=3017).

Figure 2 plots and compares the posterior means of time-varying volatility obtained from these models. Figure 3 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 very different from those based on a non-Markovian switching model, which is strongly preferred by the Bayesian model selection criterion. From Figure 2 and the reported WAIC, we conclude that the durations of the high and low volatility regimes are not constant over time.

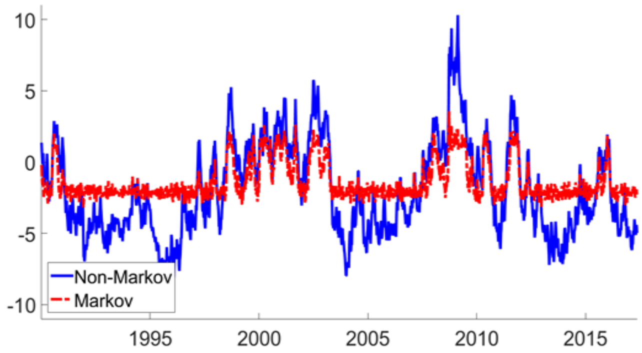

Figure 3: Posterior Mean of S_t^* from Markovian and non-Markovian Endogenous Switching Models of Volatility [Excess Returns: Jan, 1990 - May, 2017]

5.2. MARKOVIAN VERSUS NON-MARKOVIAN SWITCHING MODELS OF REAL GDP

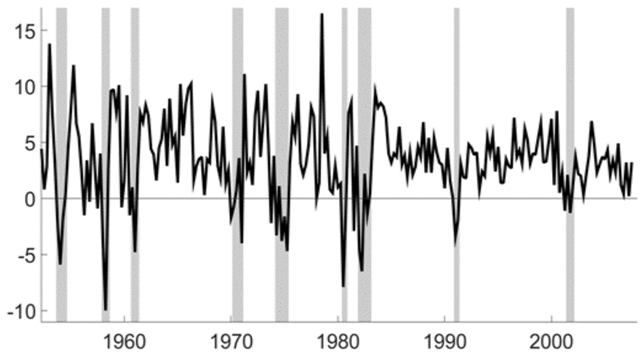

Figure 4: Quarterly Real GDP Growth Rate [1952Q1-2007Q2]

It is well known that the durations of economic expansion and recession are asymmetric in the maceroconomics literature. Regime switching models have been frequently used to capture this important nature of the real GDP data. However, it is not certain that the durations of regimes differ across the different economic recessions and expansions. In this section, comparing the Markovin and

Table 7: Bayesian Posterior Estimates of Exogenous Regime Switching Models [Quarterly Real GDP Growth Rate: 1952:Q1-2007:Q2]

	Prior	Non-Markov Switching			Markov Switching		
Parameter	Distribution	Mean (SD)	90% HPDI	PSRF	Mean (SD)	90% HPDI	PSRF
$\beta_{0,0}$	$\mathcal{N}(1,2^2)$	1.29 (0.16)	[1.03, 1.54]	1.000	1.31 (0.15)	[1.06, 1.56]	1.001
$\beta_{1,0}$	$\mathcal{N}(-1,2^2)$	-0.36 (0.37)	[-0.75, 0.16]	1.001	-0.27 (0.30)	[-0.66, 0.20]	1.001
$\beta_{0,1}$	$\mathcal{N}(0.5,2^2)$	0.92 (0.07)	[0.81, 1.04]	1.000	0.93 (0.07)	[0.81, 1.04]	1.000
$\beta_{1,1}$	$\mathcal{N}(-0.5,2^2)$	0.20 (0.24)	[-0.16, 0.59]	1.001	0.24 (0.22)	[-0.10, 0.59]	1.001
σ_0^2	$IG(1,1)$	0.87 (0.16)	[0.63, 1.13]	1.000	0.88 (0.15)	[0.65, 1.12]	1.000
σ_1^2	$IG(1,1)$	0.21 (0.04)	[0.15, 0.27]	1.000	0.21 (0.04)	[0.15, 0.27]	1.000
α	$\mathcal{N}(-1,2^2)$	-1.18 (0.61)	[-2.05, -0.18]	1.001	-	-	-
ψ	$\mathcal{N}(0.9,1^2)$	0.75 (0.11)	[0.60, 0.92]	1.003	-	-	-
α_0	$\mathcal{N}(-1.5,2^2)$	-	-	-	-1.42 (0.26)	[-1.83, -1.00]	1.001
α_1	$\mathcal{N}(1,2^2)$	-	-	-	0.69 (0.31)	[0.22, 1.19]	1.000
WAIC			253.23			252.63	

Table 8: Bayesian Posterior Estimates of Endogenous Regime Switching Models [Quarterly Real GDP Growth Rate: 1952:Q1-2007:Q2]

	Prior	Non-Markov Switching			Markov Switching		
Parameter	Distribution	Mean (SD)	90% HPDI	PSRF	Mean (SD)	90% HPDI	PSRF
$\beta_{0,0}$	$\mathcal{N}(1,2^2)$	1.40 (0.23)	[1.02, 1.76]	1.002	1.41 (0.19)	[1.10, 1.70]	1.000
$\beta_{1,0}$	$\mathcal{N}(-1,2^2)$	-0.45 (0.35)	[-0.91, 0.08]	1.000	-0.62 (0.48)	[-1.27, 0.11]	1.001
$\beta_{0,1}$	$\mathcal{N}(0.5,2^2)$	0.96 (0.09)	[0.82, 1.11]	1.001	0.96 (0.09)	[0.82, 1.10]	1.000
$\beta_{1,1}$	$\mathcal{N}(-0.5,2^2)$	0.13 (0.26)	[-0.29, 0.57]	1.000	0.09 (0.31)	[-0.42, 0.57]	1.001
σ_0^2	$IG(1,1)$	0.94 (0.18)	[0.67, 1.22]	1.000	0.98 (0.19)	[0.69, 1.28]	1.000
σ_1^2	$IG(1,1)$	0.22 (0.04)	[0.15, 0.28]	1.000	0.22 (0.04)	[0.16, 0.30]	1.000
ρ_0	$TN(0,1)$	0.27 (0.43)	[-0.36, 0.89]	1.003	0.43 (0.30)	[0.00, 0.89]	1.000
ρ_1	$TN(0,1)$	0.27 (0.38)	[-0.42, 0.91]	1.003	0.31 (0.34)	[-0.25, 0.90]	1.000
α	$\mathcal{N}(-1,2^2)$	-1.11 (0.63)	[-2.05, -0.11]	1.000	-	-	-
ψ	$\mathcal{N}(0.9,1^2)$	0.78 (0.09)	[0.65, 0.92]	1.003	-	-	-
α_0	$\mathcal{N}(-1.5,2^2)$	-	-	-	-1.36 (0.26)	[-1.77, -0.93]	1.000
α_1	$\mathcal{N}(1,2^2)$	-	-	-	0.59 (0.30)	[0.12, 1.09]	1.000
WAIC			269.31			275.92	

non-Markovian models, we test if the durations of economic regimes are statistically different.

Consider the following model specification for the log of real GDP covering the period 1952Q1 - 2007Q2:

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

where $M_t = 0$ if $t < 1984Q4$, and $M_t = 1$ otherwise; y_t is the log difference of real GDP plotted in Figure 4. β_{0,M_t} and β_{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 (σ_{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 (1999)'s model, in which the regime-indicator variable follows a first-order Markovian exogenous switching process, to the case of non-Markovian endogenous switching. In our research, similar to what is done in the stock market context, we intentionally exclude the Global Financial Crisis period and the years following it from our sample. This decision is due to the complex nature of this period, which would require advanced modeling techniques that are not part of our existing framework. Therefore, we identify the inclusion of this period in our model as a potential area for future investigation.

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 8, a non-Markovian switching model seems to be preferred to a Markovian switching model by WAIC. However, the results in Table 7 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 7.

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 5, and they are almost identical. Plots of the latent variable S_t^* from the two models presented in Figure 6 are also very close. Besides, the WAIC's for these two models are very close, suggesting that the two models are equally preferred. This result suggests

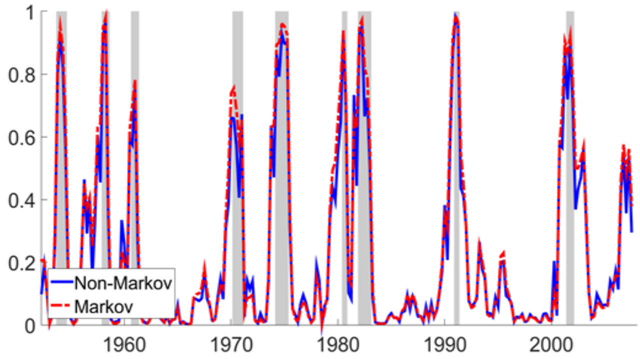

Figure 5: Posterior Probability of Recession from Markovian and non-Markovian Exogenous Switching Models of Business Cycle [Real GDP Growth: 1952Q1-2007Q2]

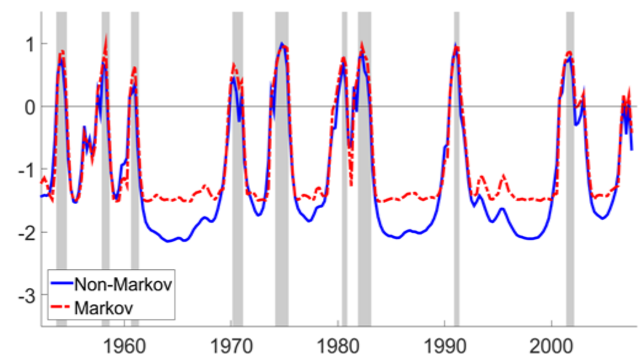

Figure 6: Posterior Mean of S_t^* from Markovian and non-Markovian Exogenous Switching Models of Business Cycle [Real GDP Growth: 1952Q1-2007Q2]

that there is no compelling evidence to support the existence of heterogeneous durations in economic recessions and expansions. While this empirical finding is intriguing, it is challenging to determine whether it stems from a paucity of data or if it truly reflects the inherent characteristics of the economic regimes.

6. CONCLUSION

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. Our empirical results suggests that, for modeling the regime-switching nature of the business cycle based on real GDP, the convention of assuming Markovian switching for the regime-indicator variable seems to be valid. For modeling volatility of the stock return, however, the non-Markovian switching model is strongly preferred to the Markovian switching model.

REFERENCES

Choi, Seungmoon (2022). "Return Predictability using an Endogenous Regime Switching Model," *Journal of Economic Theory and Econometrics,* 33(1), 1–27.

Choi, Seungmoon (2019). "Estimation of Regime-Switching Continuous-Time Stochastic Volatility Models Using KOSPI 200," *Journal of Economic Theory and Econometrics,* 30(1), 59–95.

Yang, Hyunjin and Han, H and Kim, CS (2019). "Additive endogenous regime switching garch model," *Journal of Economic Theory and Econometrics,* 30(2), 20–54.

Albert, J.H. and S. Chib (1993a). "Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance," *Journal of Business & Economic Statistics,* 11(1), 1–15.

Albert, J. H., and S. Chib (1993b). "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association,* 88, 669–679.

Chauvet, M. and S. Potter (2005). "Forecasting Recessions Using the Yield Curve," *Journal of Forecasting,* 24(2), 77–103.

Chib, S. and M. Dueker (2004). "Non-Markovian Regime Switching with Endogenous States and Time-Varying State Strengths," unpublished manuscript, Washington University.

Dueker, M. (2005). "Dynamic Forecasts of Qualitative Variables: A Qual VAR Model of U.S. Recessions," *Journal of Business & Economic Statistics,* 23(1), 96–104.

Gelman, A., and D. B. Rubin (1992). "Inference from iterative simulation using multiple sequences," *Statistical Science,* 7(4), 457–472.

Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments," *Bayesian Statistics,* 4, 641–649.

Hamilton, James D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica,* 57(2), 357–384.

Hamilton, James D. (2016). “Macroeconomic Regimes and Regime Shifts,” *Handbook of Macroeconomics,* 2A, 163–201, edited by Harald Uhlig and John Taylor, Amsterdam: Elsevier.

Harvey, A. and N. Shephard (1996). “Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns,” *Journal of Business and Economic Statistics,* 14(4), 429–434.

Hering, A.S., K. Kazor, and W. Kleiber (2015). “A Markov-Switching Vector Autoregressive Stochastic Wind Generator for Multiple Spatial and Temporal Scales,” *Resources,* 4, 70–92.

Hwu, S. T., Kim, C. J., and Piger, J. (2021). “An N-state endogenous Markov-switching model with applications in Macroeconomics and Finance,” *Macroeconomic Dynamics,* 25(8), 1937–1965.

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.

Kauppi, H. and P. Saikkonen (2008). “Predicting US recessions with dynamic binary response models,” *The Review of Economics and Statistics,* 90(4), 777–791.

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.

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.

Li, W.K. (1994). “Time Series Models Based on Generalized Linear Models: Some Further Results,” *Biometrics,* 50, 506–511.

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,” *Structural Equation Modeling,* 23(2), 221–233.

Raftery, A. E. (1985). “A Model for High-Order Markov Chains,” *Journal of Royal Statistical Society,* Series B, 47, 528–539.

Raftery, A. E., and J. D. Banfield (1991). "Stopping the Gibbs sampler, the use of morphology, and other issues in spatial statistics," *Annals of the Institute of Statistical Mathematics,* 43(1), 32–43.

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.

Startz, R. (2008). "Binomial Autoregressive Moving Average Models with an Application to US Recessions," *Journal of Business and Economic Statistics,* 26 (1), 1–8.

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.

Yu, J. (2005). "On Leverage in a Stochastic Volatility Model," *Journal of Econometrics,* 127, 165–178.

Zeger, S.L. and B. Qaqish (1988). "Markov Regression Models for time series: A quasi-likelihood approach," *Biometrics,* 44, 1019–1031.

A. APPENDIX: 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 $f(y_t, S_t|S_{t-1}^*)$:

$$\begin{aligned}\int_{-\infty}^{y_t^*} f(y_t, S_t|S_{t-1}^*)dy_t &= \int_{-\infty}^{y_t^*} f(y_t|S_{t-1}^*)f(S_t|S_{t-1}^*, y_t)dy_t \\ &= \int_{-\infty}^{\frac{y_t^*-x_t'\beta_{S_t}}{\sigma_{S_t}}} f(\varepsilon_t|S_{t-1}^*)f(S_t|S_{t-1}^*, \varepsilon_t)d\varepsilon_t \\ &= \int_{-\infty}^{\frac{y_t^*-x_t'\beta_{S_t}}{\sigma_{S_t}}} \phi(\varepsilon_t)f(S_t|S_{t-1}^*, \varepsilon_t)d\varepsilon_t,\end{aligned} \tag{45}$$

where $\phi(.)$ 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 (42) with respect to y_t*, we obtain

$$(y_t^*, S_t|S_{t-1}^*) = \frac{1}{\sigma_{S_t}}\phi\left(\frac{y_t^* - x_t'\beta_{S_t}}{\sigma_{S_t}}\right)f(S_t|S_{t-1}^*, \varepsilon_t), \tag{46}$$

which implies that equation (21) holds.

B. APPENDIX: DERIVATION OF EQUATION (23)

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

$$\eta_t = \psi\eta_{t-1} + \rho\varepsilon_t + \sqrt{1-\rho^2}\omega_t^*, \quad \omega_t \sim i.i.d.N(0,1), \tag{47}$$

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

$$f(S_{t+j}^*|S_{t+j-1}^*, \varepsilon_{t+j}) = \frac{1}{\sqrt{1-\rho^2}}\phi\left(\frac{S_{t+j}^* - \alpha(1-\psi) - \psi S_{t+j-1}^* - \rho\varepsilon_{t+j}}{\sqrt{(1-\rho^2)}}\right), \quad j = 0, 1, \tag{48}$$

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

$$
\begin{aligned}
& f(S^*_{t+1}|S^*_t, \varepsilon_{t+1}) f(S^*_t|S^*_{t-1}, \varepsilon_t) \qquad (49)\\
& \propto \exp\left[-\frac{1}{2}\left(\frac{(\eta_{t+1} - \psi\eta_t - \rho\varepsilon_{t+1})^2}{1-\rho^2} + \frac{(\eta_t - \psi\eta_{t-1} - \rho\varepsilon_t)^2}{1-\rho^2}\right)\right]\\
& \propto exp\left[-\frac{\rho^2\varepsilon_t^2 + 2\rho\psi\eta_{t-1}\varepsilon_t - (1+\psi^2)\left(\frac{1}{1+\psi^2}(\psi\eta_{t+1} - \rho\psi\varepsilon_{t+1} + \psi\eta_{t-1} + \rho\varepsilon_t)\right)^2}{2(1-\rho^2)}\right]\\
& \qquad \times exp\left[-\frac{1}{2\frac{1-\rho^2}{1+\psi^2}}\left(\eta_t - \frac{1}{1+\psi^2}(\psi\eta_{t+1} - \rho\psi\varepsilon_{t+1} + \psi\eta_{t-1} + \rho\varepsilon_t)\right)^2\right]\\
& \propto g(\varepsilon_t(S_t))\frac{1}{\sqrt{V}}\phi\left(\frac{S^*_t - \mu_t}{V}\right),
\end{aligned}
$$

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

$$\mu_t = \alpha + \frac{1}{1+\psi^2}(\psi(S^*_{t+1} - \alpha) - \rho\psi\varepsilon_{t+1} + \psi(S^*_{t-1} - \alpha) + \rho\varepsilon_t); \quad and\ V = \frac{1-\rho^2}{1+\psi^2}, \qquad (50)$$

and

$$g(\varepsilon_t(S_t)) = \exp\{-\frac{\rho^2\varepsilon_t^2 + 2\rho\psi(S^*_{t-1} - \alpha)\varepsilon_t - (1+\psi^2)(\mu_t - \alpha)^2}{2(1-\rho^2)}\}, \qquad (51)$$

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