State-Space Models with Endogenous Markov Regime Switching Parameters

Kyu Ho Kang*
(Hanyang University)

This version: December 2011

Abstract

This study extends the endogenous Markov-switching model of Kim, Piger and Startz' (2008) to a general state-space model. It also complements Kim's (1991) regime switching dynamic linear model by allowing the discrete regime to be jointly determined with observed or unobserved continuous state variables. The estimation framework involves an efficient Bayesian Markov chain Monte Carlo (MCMC) scheme to simulate the latent state variable that controls the regime shifts. A simulation exercise shows that neglecting endogeneity leads to biased inference, and it also demonstrates the effectiveness of this procedure in correctly estimating the regime shifts. This method is then applied to the dynamic Nelson-Siegel yield curve model where the unobserved time-varying level, slope and curvature factors are contemporaneously correlated with the Markov switching volatility regimes. (JEL classification: C11; C51; C58)

Keywords: Bayesian MCMC estimation, Marginal likelihood, Particle filter, Dynamic Nelson-Siegel model

1 Introduction

This study extends the endogenous Markov-switching model of Kim, Piger, and Startz (2008) to a general state-space model. This extension is motivated by the fact that state-space models with Markov switching parameters have been widely used for modeling the heterogenous dynamics of time series data (e.g., Kim (1994); Kim and Piger (2002) and Kuan, Huang, and Tsay (2005)). The popularity of this approach arises from its

*Department of Economics and Finance, Hanyang University, Seoul, South Korea 133-791. E-mail: kyu@kyukang.net.

flexibility in modeling. In standard state-space models with exogenous regime switching parameters one can model two different types of dynamic state variables: those with a discrete state space and those with a continuous state space. The evolution of the discrete state variables is usually modeled by a first-order Markov process and determines the set of the state-contingent model parameters at each data point. Given these parameters, the continuous state variables, termed as factors, for convenience, are generated from continuous probability distributions (e.g., ARMA process). They are responsible for the dynamics of the observations within the discrete state of the economy. In other words, the discrete state variables (or regimes) are generated independent of the continuous state variables and the dependent variables. Subsequently, the continuous state variables and the observations are sequentially generated conditioned on the model parameters.

Similar to the study by Kim et al. (2008), we assume that the regimes follow a discrete-time Markov process. It is then straightforward to endogenize the realization of the regimes by allowing for non-zero correlation between the regimes and the observed or unobserved factors. As a result, the present paper also complements the regime switching dynamic linear models by allowing the discrete regime to be jointly (rather than sequentially) determined with observed or unobserved continuous state variables[1].

It is important to note that the instantaneous interaction between the regimes and factors can arise when there are missing variables that are supposed to influence the realizations of both regimes and factors. Once the correlation is allowed, one confronts the endogeneity problem caused by the missing variables. The intuition behind this is that the regime indicator can be regarded as a dummy variable although it is hidden. The shocks to the factors form a fraction of the prediction error. Therefore, the error term and the explanatory variable are contemporaneously correlated at each time point,

[1]As in some dynamical systems the evolution of the outcomes is jointly driven by both types of state variables, it is essential to consider contemporaneous feedback among them when modeling such systems. In such cases, the instantaneous interaction between the continuous state variables can be modeled by allowing for non-zero correlation of the innovations (see, for instance, Morley, Nelson, and Zivot (2003) and Diebold, Rudebusch, and Aruoba (2006)). Also, the co-movement of more than two discrete state variables can be permitted by estimating the unrestricted transition matrix of the aggregate Markov chain process as presented in the study by Ang, Bekaert, and Wei (2008) and Kim (1993). While these modeling strategies are applicable to a variety of models, they are restrictive because they only permit the instantaneous interaction among the same type of state variables.

and thus the regime switching process becomes endogenous. This tends to result in inaccuracy of the estimates if the problem is simply ignored.

This generalization is particularly important in many of the macroeconomics and finance models. For example, in the three-factor Nelson-Siegel yield curve model, the slope and curvature of the term structure, both of which are modeled as unobserved factors, affect the future risk premia (Cochrane and Piazzesi (2008)). Moreover, they are highly likely to be contemporaneously correlated with the Markov switching volatility regimes because the size of shock volatilities determines the magnitude of the risk. Therefore, it is more theoretically and statistically natural to model the joint realization of the two different types of the unobserved underlying stochastic processes.

The goal of this study is to develop and estimate linear Gaussian state space models in which the discrete and continuous state variables are determined jointly, not sequentially. The Bayesian MCMC (Markov chain Monte Carlo) estimation scheme proposed here builds on the work by Albert and Chib (1993), Filardo and Gordon (1998) and Kim et al. (2008). Kim et al. (2008) developed an endogenous Markov-switching framework for linear models that do not involve the dynamics of continuous state variables. Thus, this study extends their work to a general state-space model, and the central idea here is to introduce and simulate the latent state variables that control the regime shifts at each time point. This idea of data augmentation is based on the work of Albert and Chib (1993) who introduced the exact Bayesian methods for modeling categorical response data.[2] Filardo and Gordon (1998) applied this method to estimate the parameters entering the probit specification of time-varying transition probabilities. Our estimation technique described in the present study is rounded off by proposing a particle filtering algorithm for likelihood computation and addressing the Bayesian model choice problem between endogenous and exogenous regime switching models based on the marginal likelihoods.

The rest of the paper is organized as follows. Section 2 presents the generalized class of state space models with endogenous Markov switching parameters. Section

[2]In a related paper Chib and Dueker (2004) developed a Bayesian non-Markovian regime switching model. In their setup, the regime states depended on the sign of an autoregressive latent variable, which was allowed to be endogenous in the sense that regimes were determined jointly with the observed data.

3 discusses the Bayesian MCMC estimation method. Section 4 provides a simulation exercise followed by an empirical finance application. Finally, concluding notes are discussed in Section 5.

2 The Class of Models

Consider a class of linear Gaussian state-space models where at each time point the model parameters are chosen by a Markov process with finite states indexed by $s_t \in \{1, 2, .., K\}$. Suppose that some of the unobserved continuous and discrete variables are contemporaneously correlated. This class of models can be represented as follows

$$\mathbf{y}_t = \mathbf{a}_{s_t} + \mathbf{b}_{s_t}\mathbf{f}_t + H_{s_t}\mathbf{x}_t + D_{s_t}e_t, \quad e_t \sim \text{i.i.d.}\mathcal{N}_q(0, \mathbf{I}_q) \tag{2.1}$$

$$\mathbf{f}_t = \mu_{s_t} + G_{s_t}\mathbf{f}_{t-1} + L_{s_t}\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.}\mathcal{N}_k(0, \mathbf{I}_k) \tag{2.2}$$

where $\mathbf{y}_t$ is a q-dimensional vector of dependent variables and $\mathbf{f}_t$ is a k-dimensional vector of observed or unobserved continuous state variables. $\mathbf{x}_t$ is a h-dimensional vector of the observed strictly exogenous or predetermined variables that are assumed to be covariance-stationary. $\mathbf{a}_{s_t}$, $\mathbf{b}_{s_t}$ and H_{s_t} are $q \times 1$, $q \times k$ and $q \times h$ matrices, respectively. $\mu_{s_t} : k \times 1$ and $G_{s_t} : k \times k$ determine the state-dependent intercept and persistence of the continuous state variables. $D_{s_t} : q \times q$ and $L_{s_t} : k \times k$ capture the volatilities of the measurement errors and the continuous state shocks, respectively. Furthermore, as is standard in the literature, it is assumed that (i) e_t and ε_t are independent, (ii) $\mathbb{E}[e_t|\mathbf{S}_n] = 0$ and (iii) $\mathbb{E}[e_t e_t'|\mathbf{S}_n] = \mathbf{I}_q$.

In what follows, as in Hamilton (1989), we assume without loss of generality that the number of regimes is two (i.e. K=2). The two-regime case is not only convenient to explain and understand, but also popular in many empirical applications. Additionally, it is supposed that the unobserved discrete state variable s_t is governed by a first-order Markov chain with transition probabilities given by

$$p(s_t = k|s_{t-1} = j, \mathbf{z}_t) = p_{jk}(\mathbf{z}_t) \tag{2.3}$$

where $\mathbf{z}_t$ is a vector of covariance-stationary exogenous or predetermined variables. $\mathbf{z}_t$ may include some elements of $\mathbf{x}_t$ and the lagged dependent variables. The Markov chain is assumed to be stationary and independent of all observations of those elements of $\mathbf{x}_t$ not included in $\mathbf{z}_t$. To formulate this Markov process, it is convenient to introduce another latent variable γ_t, such that the influence of $\mathbf{z}_t$ on the transition probabilities is modeled through a probit specification as reported by Kim et al. (2008).

$$s_t = \begin{cases} 1 & \text{if } \gamma_t < \alpha_{s_{t-1}} + \beta'_{s_{t-1}}\mathbf{z}_t \\ 2 & \text{if } \gamma_t > \alpha_{s_{t-1}} + \beta'_{s_{t-1}}\mathbf{z}_t \end{cases} \quad \text{where } \gamma_t \sim \text{i.i.d.}\mathcal{N}(0,1) \tag{2.4}$$

In other words, the transition probabilities have the form

$$\begin{aligned} p_{j1}(\mathbf{z}_t) &= \Pr\left[\gamma_t < \alpha_j + \beta_j'\mathbf{z}_t\right] = \Phi\left(\alpha_j + \beta_j'\mathbf{z}_t\right) \\ p_{j2}(\mathbf{z}_t) &= 1 - \Phi\left(\alpha_j + \beta_j'\mathbf{z}_t\right) \end{aligned} \tag{2.5}$$

where Φ denotes the standard normal distribution function. It should be noted that by allowing for non-zero correlation between the regime shock γ_t and the innovations ε_t one is able to capture a bi-directional contemporaneous feedback between the unobserved discrete state variable s_t and the unobserved continuous state variables $\mathbf{f}_t$

$$\begin{bmatrix} \varepsilon_t \\ \gamma_t \end{bmatrix} \sim \text{i.i.d.}\mathcal{N}_{k+1}\left(0_{(k+1)\times 1}, \begin{pmatrix} \mathbf{I}_k & \rho \\ \rho' & 1 \end{pmatrix}\right) \tag{2.6}$$

where $\rho = \begin{pmatrix} \rho_1 & \rho_2 & \cdots & \rho_k \end{pmatrix}'$. Hence, the presence of some none-zero $\rho_i's$ $(i = 1, 2, .., k)$ implies endogenous regime changes, the realization of which at the next period is determined jointly with the vector of continuous unobserved (or observable) variables conditioned on their lagged values. This is the distinguishing feature of this paper. Notice that it is possible to make the measurement error e_t correlated with γ_t by moving e_t into the vector $\mathbf{f}_t$. When all ρ_i's are zero, one gets back to the familiar Markov switching model with time-varying transition probabilities, similar to that presented in the study by Gray (1996), Diebold, Lee, and Weinbach (1994), Filardo and Gordon (1994) and Filardo and Gordon (1998). When $k = q = \mathbf{b}_{s_t} = 1$ and $G_{s_t} = \mu_{s_t} = D_{s_t} = 0$, this class of models immediately reduces to that of Kim et al. (2008).

The instantaneous interaction between the shocks to the factor and regimes arises when there is a missing variable, denoted by h_t, in both the transition equation (2.2) and the transition probabilities (2.4). Suppose that h_t affects the realizations of factors and regimes simultaneously, however it is omitted because it is not observed from the data.

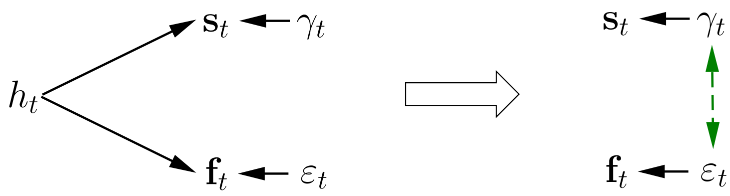


Figure 1: **Directed graph of correlation between the state shocks**. This figure describes how the shocks to the regimes and factors can be correlated when there is a missing variable, denoted by h_t, in the transition equation and the transition probabilities

In that case, ε_t and γ_t will contain common information about the missing variable, as described in Figure (1). Note that s_t can be regarded as a dummy variable although it is hidden, and ε_t is a fraction of prediction error. Hence, the explanatory dummy variable and the error term are instantaneously correlated. This is why this endogenous regime switching state space model is conceptually similar to the linear regression model with the endogeneity problem caused by missing variables.

Several other interesting models can be constructed as a special cases in this framework. Noteworthy examples include time-varying coefficient models, dynamic common factor models and unobserved component models. As an example of how this situation may arise, let y_t be a vector of yields with various maturities and $\mathbf{f}_t$ be the slope factor among them in a dynamic Nelson-Siegel model where the volatility is regime-specific. Additionally, let h_t be the market-wide or systematic risk of the economy. Subsequently the slope factor captures the risk premium and the volatility measures the time-varying size of the risk, thus both of them are supposed to be influenced by the missing variable h_t. Therefore, the Markov switching volatility and the factor process are highly likely to be positively correlated.

Figure 2 summarizes the data generating process through a directed acyclic graph. In the beginning of period t, a regime and a vector of continuous state variables occur simultaneously. In particular, this realization of the regime at time t is governed by the regime in the previous period and the current innovations to the continuous state variables s_t as indicated by the direction of the two arrows connecting s_{t-1} to s_t and $\mathbf{f_t}$ to s_t. Subsequently, given the regime at time t, the corresponding model parameters $\mathbf{\Theta}_{s_t}$ are taken from the full collection of model parameters. These include $\mathbf{a}_{\mathbf{s_t}}$ and $\mathbf{b}_{s_t}$,

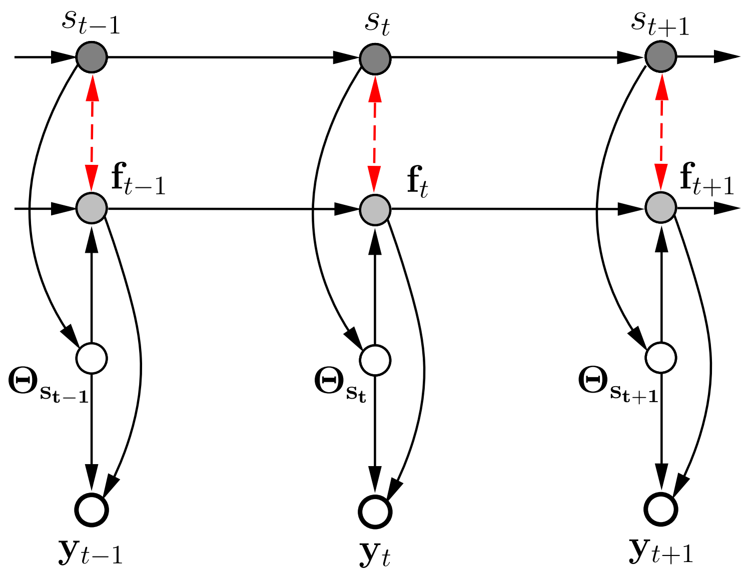


Figure 2: **Directed graph of model linkages**. This is a summary of the data generating process. In the beginning of period t, a regime and a vector of the continuous state variables occur simultaneously conditioned on $\mathbf{f}_{t-1}$ and s_{t-1}. Subsequently, given the regime s_t, the corresponding model parameters $\mathbf{\Theta}_{s_t}$ are taken from the full collection of the model parameters. Finally, after simulating the measurement error e_t, $\mathbf{y}_t$ is generated from equation (2.1)

for example. Conditioned on the parameters, s_t and $\mathbf{f}_{t-1}$, $\mathbf{f}_t$ is computed by the regime-specific autoregressive process in (2.2). Finally, from (2.1), $\mathbf{a}_{\mathbf{s_t}}$, $\mathbf{b}_{s_t}$, $\mathbf{f}_t$ and a simulated measurement error e_t at each time point one can construct the observations $\mathbf{y}_t$. Note that in standard state space models with exogenous regime switching parameters the dashed line in Figure 2 is absent because s_t is assumed to be drawn independent of the innovations to $\mathbf{f}_t$.

3 Prior-Posterior Analysis

3.1 Markov Chain Monte Carlo Scheme

Let $\mathbf{Y}_t = \{\mathbf{y}_i\}_{i=1,2,..,t}$, $\mathbf{X}_t = \{\mathbf{x}_i\}_{i=1,2,..,t}$ and $\mathbf{Z}_t = \{\mathbf{z}_i\}_{i=1,2,..,t}$ denote the observations through time t. Similarly, $\mathbf{F}_t = \{\mathbf{f}_i\}_{i=1,2,..,t}$, $\mathbf{S}_t = \{s_i\}_{i=0,1,2,..,t}$ and $\boldsymbol{\gamma}_t = \{\gamma_i\}_{i=1,2,..,t}$ denote the collection of the state variables through time t. Let $\mathbf{\Theta}$ be the parameters in the evolution and transition equations, and let $\mathbf{P}$ denote those in the transition probabilities. That is, $\mathbf{\Theta}$ is the collection of the parameters in $\{\mathbf{a}_{s_t},\ \mathbf{b}_{s_t},\ H_{s_t},\ D_{s_t},\ \mu_{s_t},\ G_{s_t},\ L_{s_t},\ \mathbf{f}_0\}$ and $\mathbf{P}$ $=\{\alpha_1,\ \alpha_2,\ \beta_1',\ \beta_2',\ \rho_1,\ \rho_2,\ ..,\ \rho_k\}$. Let us specify the prior density $\pi(\mathbf{\Theta}, \mathbf{P})$ on the parameters and suppose that data on $\mathbf{Y}_n$, $\mathbf{Z}_n$ and $\mathbf{X}_n$ are available. In the context of Bayesian state space models with regime switching parameters, the joint posterior distribution of the regime states, continuous latent variables and model parameters (conditioned on the data, including any exogenous variables) are obtained as

$$\begin{aligned} & \pi\left(\mathbf{\Theta}, \mathbf{P}, \mathbf{S}_n, \boldsymbol{\gamma}_n, \mathbf{F}_n | \mathbf{Y}_n, \mathbf{\Omega}_n\right) \\ \propto \quad & f\left(\mathbf{Y}_n | \mathbf{\Omega}_n, \mathbf{S}_n, \boldsymbol{\gamma}_n, \mathbf{F}_n, \mathbf{\Theta}, \mathbf{P}\right) \times p\left(\mathbf{S}_n, \boldsymbol{\gamma}_n, \mathbf{F}_n | \mathbf{\Omega}_n, \mathbf{\Theta}, \mathbf{P}\right) \times \pi(\mathbf{\Theta}, \mathbf{P}) \end{aligned} \tag{3.1}$$

Here $\mathbf{\Omega}_t = \mathbf{X}_t \cup \mathbf{Z}_t$ is the collection of exogenous variables at time t whose dynamics are not analyzed through the transition equation (2.2).

As interest centers mainly on the model parameters $\mathbf{\Theta}$ and $\mathbf{P}$, the MCMC sampling scheme developed here is targeted at obtaining the (marginal) posterior density of $\pi\left(\mathbf{\Theta}, \mathbf{P} | \mathbf{Y}_n, \mathbf{\Omega}_n\right)$ by integrating out $(\mathbf{S}_n, \boldsymbol{\gamma}_n, \mathbf{F}_n)$ from $\pi(\mathbf{\Theta},\ \mathbf{P},\ \mathbf{S}_n,\ \boldsymbol{\gamma}_n,\ \mathbf{F}_n | \mathbf{Y}_n,\ \mathbf{\Omega}_n)$. This is achieved by sampling the parameters and the states recursively. In the first step, the parameters in $\mathbf{\Theta}$ are simulated on $(\mathbf{S}_n,\ \mathbf{\Omega}_n,\ \mathbf{W}_n,\ \mathbf{P},\ \boldsymbol{\gamma}_n)$ and then $\mathbf{P}$ is sampled in turn. Next, the discrete states $\mathbf{S}_n$ are drawn conditioned on $(\mathbf{W}_n,\ \mathbf{\Omega}_n)$ and the other parameters where $\mathbf{W}_t = \mathbf{Y}_t \cup \mathbf{F}_t$. Finally, one sequentially simulates $\boldsymbol{\gamma}_n$ and $\mathbf{F}_n$ conditioned on the most recent values of the conditioning variables. The following is the step-by-step summary of the MCMC algorithm and the details of each of the steps.

Algorithm 1: MCMC sampling

Step 1 Initialize $(\mathbf{S}_n, \mathbf{F}_n, \mathbf{P}, \boldsymbol{\gamma}_n)$ and fix n_0 (the burn-in) and n_1 (the MCMC sample size)

Step 2 Sample $\mathbf{\Theta} | \mathbf{W}_n, \mathbf{\Omega}_n, \mathbf{S}_n, \boldsymbol{\gamma}_n, \mathbf{P}$

Step 3 Sample $\mathbf{P} | \mathbf{W}_n, \mathbf{\Omega}_n, \mathbf{S}_n, \mathbf{\Theta}$

Step 4 Sample $\mathbf{S}_n | \mathbf{W}_n, \mathbf{\Omega}_n, \mathbf{\Theta}, \mathbf{P}$

Step 5 Sample $\boldsymbol{\gamma}_n|\mathbf{W}_n, \boldsymbol{\Omega}_n, \mathbf{S}_n, \boldsymbol{\Theta}, \mathbf{P}$

Step 6 Sample $\mathbf{F}_n|\mathbf{Y}_n, \boldsymbol{\Omega}_n, \mathbf{S}_n, \boldsymbol{\gamma}_n, \boldsymbol{\Theta}, \mathbf{P}$

Step 7 Repeat Steps 2-6, discard the draws from the first n_0 iterations and save the subsequent n_1 draws.

3.1.1 Simulation of $\boldsymbol{\Theta}$

In general, given the high dimensionality of the parameter space in most realistic applications, the simulation of $\boldsymbol{\Theta}$ conditioned on ($\boldsymbol{\Omega}_n$,$\mathbf{W}_n$,$\mathbf{S}_n$,$\boldsymbol{\gamma}_n$,$\mathbf{P}$) is better dealt with the tailored multiple block Metropolis-Hastings (MH) algorithm (Chib and Greenberg (1995)). In this method the parameters in $\boldsymbol{\Theta}$ are first blocked into various sub-blocks. Then each of these sub-blocks is sampled in sequence by drawing a value from a tailored proposal density constructed for that particular block. This proposal is then accepted or rejected by the usual MH probability move. For instance, suppose that in the jth iteration, there are g sub-blocks of $\boldsymbol{\Theta}$

$$\boldsymbol{\Theta}_1,\ \boldsymbol{\Theta}_2,\ .\ .,\ \boldsymbol{\Theta}_g$$

Then the proposal density $q\left(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i}, \mathbf{W}_n, \boldsymbol{\gamma}_n, \mathbf{S}_n, \mathbf{P}\right)$ for the ith block, conditioned on the most current value of the remaining blocks $\boldsymbol{\Theta}_{-i}$, is constructed by a quadratic approximation at the mode of the current target density $\pi\left(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i}, \mathbf{W}_n, \boldsymbol{\gamma}_n, \mathbf{S}_n, \mathbf{P}\right)$. In practice, the student t distribution with relatively low degrees of freedom (e.g., 15) is a good choice for the proposal density, that is

$$q\left(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_{-i}, \mathbf{P}, \mathbf{W}_n, \boldsymbol{\gamma}_n, \boldsymbol{\Omega}_\mathbf{n}, \mathbf{S}_n\right) = St\left(\boldsymbol{\Theta}_i|\boldsymbol{\Theta}_i, \mathbf{P}, \mathbf{V}_{\boldsymbol{\Theta}_i}, \mathbf{15}\right) \tag{3.2}$$

where

$$\hat{\boldsymbol{\Theta}}_i = \arg\max_{\boldsymbol{\Theta}_\mathbf{i}} \ln\{f(\mathbf{W}_\mathbf{n}|\boldsymbol{\gamma}_\mathbf{n}, \boldsymbol{\Omega}_\mathbf{n}, \mathbf{S}_\mathbf{n}, \boldsymbol{\Theta}_i, \boldsymbol{\Theta}_{-i}, \mathbf{P})\pi(\boldsymbol{\Theta}_i)\} \tag{3.3}$$

$$\text{and } \mathbf{V}_{\boldsymbol{\Theta}_i} = \left(-\frac{\partial^2 \ln\{f(\mathbf{W}_\mathbf{n}|\boldsymbol{\gamma}_\mathbf{n}, \boldsymbol{\Omega}_\mathbf{n}, \mathbf{S}_\mathbf{n}, \boldsymbol{\Theta}_i, \boldsymbol{\Theta}_{-i}, \mathbf{P})\pi(\boldsymbol{\Theta}_i)}{\partial\boldsymbol{\Theta}_i\partial\boldsymbol{\Theta}_i'}\right)^{-1}_{|\boldsymbol{\Theta}_i=\boldsymbol{\Theta}_i}.$$

Let $\mathcal{R}$ be the collection of restrictions on the parameter space for identifying the factors and regimes. The generated proposal value $\boldsymbol{\Theta}_i^\dagger$ is immediately rejected if it violates any of the constraints in $\mathcal{R}$, otherwise, it is accepted as the next value in the chain with probability

$$\alpha\left(\boldsymbol{\Theta}_i^{(j-1)}, \boldsymbol{\Theta}_i^{\dagger}|\boldsymbol{\Theta}_{-i}\right) \tag{3.4}$$
$$= \min\left\{\frac{f(\mathbf{W_n}|\gamma_\mathbf{n}, \boldsymbol{\Omega}_\mathbf{n}, \mathbf{S_n}, \boldsymbol{\Theta}_i^{\dagger}, \boldsymbol{\Theta}_{-i}, \mathbf{P})\pi\left(\boldsymbol{\Theta}_i^{\dagger}\right)}{f\left(\mathbf{W_n}|\gamma_\mathbf{n}, \boldsymbol{\Omega}_\mathbf{n}, \mathbf{S_n}, \boldsymbol{\Theta}_i^{(j-1)}, \boldsymbol{\Theta}_{-i}, \mathbf{P}\right)\pi\left(\boldsymbol{\Theta}_i^{(j-1)}\right)} \frac{St\left(\boldsymbol{\Theta}_i^{(j-1)}|\boldsymbol{\Theta}_i, \mathbf{P}, \mathbf{V}_{\boldsymbol{\Theta}_i}, \mathbf{15}\right)}{St\left(\boldsymbol{\Theta}_i^{\dagger}|\boldsymbol{\Theta}_i, \mathbf{P}, \mathbf{V}_{\boldsymbol{\Theta}_i}, \mathbf{15}\right)}, 1\right\}.$$

The simulation of $\boldsymbol{\Theta}$ is complete when all the sub-blocks $\{\boldsymbol{\Theta}_i\}_{i=1,2,..,g}$ are sequentially updated as mentioned earlier.

On letting $\mathcal{N}_q(x|a,b)$ denote the q-dimensional multivariate normal density of x with a mean of a and a variance of b, the required joint density of $\mathbf{W}_n$ conditioned on $(\boldsymbol{\Omega}_n, \boldsymbol{\gamma}_n, \mathbf{S}_n, \boldsymbol{\Theta}, \mathbf{P})$ is then a product of conditional predictive densities:

$$\begin{aligned} & f\left(\mathbf{W}_n|\gamma_n, \boldsymbol{\Omega}_n, \mathbf{S}_n, \boldsymbol{\Theta}, \mathbf{P}\right) \\ = & \prod_{t=1}^{n} \mathcal{N}_q\left(\mathbf{y}_t|\mathbf{y}_{t|t-1}, D_{s_t}D'_{s_t}\right) \times \mathcal{N}_k\left(\mathbf{f}_t|\mathbf{f}_{t|t-1}, L_{s_t}\left(\mathbf{I}_k - \rho\rho'\right)L'_{s_t}\right) \end{aligned} \tag{3.5}$$

where

$$\mathbf{y}_{t|t-1} = \mathbb{E}\left[\mathbf{y}_t|\mathbf{f}_t, \boldsymbol{\Omega}_n, \mathbf{W}_{t-1}, \mathbf{S}_n, \gamma_n\right] = \mathbf{a}_{s_t} + \mathbf{b}_{s_t}\mathbf{f}_t + H_{s_t}\mathbf{x}_t \tag{3.6}$$

$$\mathbf{f}_{t|t-1} = \mathbb{E}\left[\mathbf{f}_t|\boldsymbol{\Omega}_n, \mathbf{W}_{t-1}, \mathbf{S}_n, \gamma_n\right] = \mu_{s_t} + G_{s_t}\mathbf{f}_{t-1} + L_{s_t}\rho\gamma_t \tag{3.7}$$

As ε_t and γ_t are jointly normally distributed, it follows that $\mathbb{E}\left[\varepsilon_t|\gamma_t\right] = \rho\gamma_t$ and $\mathbb{V}\left[\varepsilon_t|\gamma_t\right] = \mathbf{I}_k - \rho\rho'$. It is important to note that the mean of ε_t conditioned on the state at time t is non-zero, which implies that ignoring the last term in equation (3.6) causes the ordinary omitted variable problem. $L_{s_t}\rho\gamma_t$ plays an important role in correcting the error in μ_{s_t} or $\mathbf{a}_{s_t}$ depending on the identifying restrictions for the unconditional mean of $\mathbf{y}_t$. It should be noted that the conditional mean of γ_t given s_t and s_{t-1} is non-zero because γ_t is a draw from a truncated normal distribution as described in equation (2.4). In addition, without $\boldsymbol{\gamma}_n$ the density of $\mathbf{y}_t$ conditioned on s_t and s_{t-1} is no longer Gaussian as pointed out by Kim et al. (2008).

When the likelihood function tends to be ill-behaved in these problems, either because of inherent nonlinearities or complicated constraints (or both), one may calculate $\boldsymbol{\Theta}_i$

using a suitably designed version of the simulated annealing algorithm like in Chib and Ergashev (2009). This stochastic optimization method is more reliable than the standard Newton-Raphson class of deterministic optimizers especially when the multi-modality of the likelihood surface is severe. Moreover, if there is no convenient way of grouping the parameters based on their correlation structure, one can resort to the randomized blocking scheme introduced by Chib and Ramamurthy (2010) to achieve better efficiency and faster convergence to the stationary distribution.

3.1.2 Simulation of P

We now discuss sampling $\mathbf{P}$. In our formulation its full conditional $p\left(\mathbf{P}|\mathbf{S}_n, \mathbf{\Omega}_n, \mathbf{F}_n, \mathbf{\Theta}\right)$ is independent of $\mathbf{Y}_n$ as in standard regime switching models. However, because of the non-zero contemporaneous correlation between the regime process and the continuous state variables, it depends on $\mathbf{F}_n$. Also as the full conditional is not tractable in general, we rely on a MH method as above. Given the prior density of $\mathbf{P}$ that is assumed to be independent of $\mathbf{\Theta}$, we need to discuss the joint density for $(\mathbf{S}_n, \mathbf{F}_n)$ in order to apply the MH method as in the previous step because the full conditional is proportional to the product of the joint density and its prior

$$p\left(\mathbf{P}|\mathbf{S}_n, \mathbf{W}_n, \mathbf{\Omega}_n, \mathbf{\Theta}\right) \propto p\left(\mathbf{S}_n, \mathbf{F}_n|\mathbf{\Omega}_n, \mathbf{\Theta}, \mathbf{P}\right) \pi\left(\mathbf{P}\right) \tag{3.8}$$

Next, we should note that the joint density has the form

$$\begin{aligned} p\left(\mathbf{S}_n, \mathbf{F}_n|\mathbf{\Omega}_n, \mathbf{\Theta}, \mathbf{P}\right) &= p(s_0) f(\mathbf{f}_0) \prod\nolimits_{t=1}^{n} p[s_t, \mathbf{f}_t|s_{t-1}, \mathbf{f}_{t-1}, \mathbf{z}_t, \mathbf{\Theta}, \mathbf{P}] \\ &\propto \prod\nolimits_{t=1}^{n} p[s_t, \mathbf{f}_t|s_{t-1}, \mathbf{f}_{t-1}, \mathbf{z}_t, \mathbf{\Theta}, \mathbf{P}] \end{aligned} \tag{3.9}$$

where $p(\cdot)$ denotes a joint density of discrete and continuous random variables or discrete random variables only. Each density in the product can be computed as follows.

$$p(s_t, \mathbf{f}_t|s_{t-1} = j, \mathbf{f}_{t-1}, \mathbf{z}_t, \mathbf{\Theta}, \mathbf{P}) \tag{3.10}$$

$$= \quad f\left(\mathbf{f}_t|s_t, s_{t-1} = j, \mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_t|s_{t-1} = j, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) \tag{3.11}$$

The first term in the product of the equation (3.11) is the density of $\mathbf{f}_t$ conditioned on s_t and $s_{t-1} = j$, and the event $\{s_t = 1 \text{ and } s_{t-1} = j\}$ is equivalent to $\{\gamma_t < \alpha_j +$

$\beta_j'\mathbf{z}_t\}$. Due to the correlation between γ_t and ε_t one may assume that $\gamma_t = \rho'\varepsilon_t - \sqrt{1-\rho'\rho}\omega_t$ where ε_t and ω_t are i.i.d. $\mathcal{N}(0,1)$ variables. Consequently, $\{\gamma_t < \alpha_j + \beta_j'\mathbf{z}_t\}$ is equivalent to $\{\rho'\varepsilon_t/\sqrt{1-\rho'\rho} < (\alpha_j + \beta_j'\mathbf{z}_t)/\sqrt{1-\rho'\rho} + \omega_t\}$. It follows that $f(\mathbf{f}_t|s_t, s_{t-1} = j, \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})$ is a conditional normal density with hidden truncation and, thus, belongs to the class of skew-normal density with the skewness parameter of $\rho'/\sqrt{1-\rho'\rho}$, and in the case of $s_t = 1$ it has the form

$$\begin{aligned} & f(\mathbf{f}_t|s_t = 1, s_{t-1} = j, \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \qquad (3.12) \\ = & \frac{\phi_k\left(L_1^{-1}(\mathbf{f}_t - \mu_1 - G_1\mathbf{f}_{t-1})\right)}{|L_1|\, p(s_t = 1|s_{t-1} = j, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})}\Phi\left(\frac{\alpha_j + \beta_j'\mathbf{z}_t - \rho'\left(L_1^{-1}(\mathbf{f}_t - \mu_1 - G_1\mathbf{f}_{t-1})\right)}{\sqrt{1-\rho'\rho}}\right) \end{aligned}$$

where $\phi_k(\cdot)$ and $\Phi(\cdot)$ denote the $k-$dimensional multivariate standard normal density and the c.d.f of the standard normal distribution, respectively. The higher correlation ρ indicates more severe endogeneity and causes larger skewness[3]. Then it immediately follows that the equation (3.10) can be rewritten as

$$\frac{\phi_k\left(L_1^{-1}(\mathbf{f}_t - \mu_1 - G_1\mathbf{f}_{t-1})\right)}{|L_1|}\Phi\left(\frac{\alpha_j + \beta_j'\mathbf{z}_t - \rho'\left(L_1^{-1}(\mathbf{f}_t - \mu_1 - G_1\mathbf{f}_{t-1})\right)}{\sqrt{1-\rho'\rho}}\right) \qquad (3.13)$$

As the density of $\varepsilon_t = L_1^{-1}(\mathbf{f}_t - \mu_1 - G_1\mathbf{f}_{t-1})$ does not involve the model parameters in $\mathbf{P}$, $p[s_t = 1, \mathbf{f}_t|s_{t-1}, \mathbf{f}_{t-1}, \mathbf{z}_t, \boldsymbol{\Theta}, \mathbf{P}]$ becomes proportional to

$$\Phi\left(\frac{\alpha_j + \beta_j'\mathbf{z}_t - \rho'\left(L_1^{-1}(\mathbf{f}_t - \mu_1 - G_1\mathbf{f}_{t-1})\right)}{\sqrt{1-\rho'\rho}}\right) \qquad (3.14)$$

In exactly the same way, for the case of $s_t = 2$,

$$\begin{aligned} & p[s_t = 2, \mathbf{f}_t|s_{t-1} = j, \mathbf{f}_{t-1}, \mathbf{z}_t, \boldsymbol{\Theta}, \mathbf{P}] \qquad (3.15) \\ = & \frac{\phi_k\left(L_2^{-1}(\mathbf{f}_t - \mu_2 - G_2\mathbf{f}_{t-1})\right)}{|L_2|}\left[1 - \Phi\left(\frac{\alpha_j + \beta_j'\mathbf{z}_t - \rho'\left(L_2^{-1}(\mathbf{f}_t - \mu_2 - G_2\mathbf{f}_{t-1})\right)}{\sqrt{1-\rho'\rho}}\right)\right] \\ \propto & \;\Phi\left(\frac{-\alpha_j - \beta_j'\mathbf{z}_t + \rho'\left(L_2^{-1}(\mathbf{f}_t - \mu_2 - G_2\mathbf{f}_{t-1})\right)}{\sqrt{1-\rho'\rho}}\right) \qquad (3.16) \end{aligned}$$

Now, by using the equations (3.14) and (3.15) one can compute the likelihood density for $\mathbf{P}$ conditioned on $(\mathbf{S}_n, \boldsymbol{\Omega}_n, \mathbf{F}_n, \boldsymbol{\Theta})$, $p(\mathbf{S}_n, \mathbf{F}_n|\boldsymbol{\Omega}_n, \boldsymbol{\Theta}, \mathbf{P})$ without the normalizing constant.

[3]For more details about skew-normal distributions, see Arnold and Beaver (2002)

This enables us to sample $\mathbf{P}$ from its full conditional distribution by the MH method described in the previous subsection.

One can see that when all elements in ρ and β_j are zeros for $j = 1, 2$ (i.e. $p_{ij}(\mathbf{z}_t) = p_{ij}$ and regime shifts are exogenous), multiplying a beta prior by the likelihood function of $\mathbf{P}$ conditioned on $\mathbf{S}_n$ immediately gives the result that the updated distribution is also beta (Chib (1996)). In that case, the full conditional distribution of the transition matrix can be derived without considering the sampling model. Therefore, an exogenous Markov regime switching model with constant transition probabilities is a special case of our formulation presented in this study when $\rho = 0_{k\times 1}$ and $\beta_j = 0$ for all $j \in \{1, 2\}$. It also implies that neglecting endogeneity can lead to biased inference in α_{s_t} as well as μ_{s_t}.

It should be noted that some parameters can be sampled through a Gibbs-sampling method like in Filardo and Gordon (1998). For example, the posterior of the correlation parameter ρ can be conditioned on γ_t and "realized" factor shocks and measurement errors given the other sampled parameters, continuous and discrete states. Given a normal conjugate prior on $(\alpha_{s_r}, \beta_{s_t})$, they can be updated from the normal posterior distribution. Despite that, we rely on the M-H step because this approach is more generic and, hence, can be applied to the case even when the parameters in $\mathbf{P}$ appear in the measurement equation (e.g., affine term structure models with regime shifts (Bansal and Zhou (2002) and Chib and Kang (2010)))[4]. In addition, by letting all the parameters be updated through the M-H steps, one can easily adopt the method of Chib and Jeliazkov (2001) for the marginal likelihood calculation.

3.1.3 Simulation of $\{s_t\}$

To sample $\mathbf{S}_n$ we provide a generalized version of the multi-move method of Chib (1996) where exogenous Markov mixture models are analyzed. The objective in this subsection is to draw a sequence of values of $\{s_t\}_{t=0,1,2,..,n}$ jointly from $p(\mathbf{S}_n|\mathbf{W}_n, \boldsymbol{\Omega}_n, \boldsymbol{\Theta}, \mathbf{P})$ where the state process is endogenous. It can be easily seen that this sampling is simulating s_t from $p\left(s_t|\mathbf{W}_n, \boldsymbol{\Omega}_n, s_{t+1}, \boldsymbol{\Theta}, \mathbf{P}\right)$, and by Bayes theorem

$$p\left(s_t|\mathbf{W}_n, \boldsymbol{\Omega}_n, s_{t+1}, \boldsymbol{\Theta}, \mathbf{P}\right) \propto p\left(s_t|\mathbf{W}_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}\right) p\left(s_{t+1}|s_t, \boldsymbol{\Omega}_{t+1}, \mathbf{f}_{t+1}, \boldsymbol{\Theta}, \mathbf{P}\right) \tag{3.17}$$

[4] In this case $\boldsymbol{\Theta}$ and $\mathbf{P}$ should be sampled in one block.

In this approach, the sampling of $\mathbf{S}_n$ is achieved by one forward and backward pass through the data.

In the forward pass, one recursively obtains the sequence of filtered probabilities $p\left(s_t|\mathbf{W}_n, \mathbf{\Omega}_n, \mathbf{\Theta}, \mathbf{P}\right)$ by calculating

$$\begin{aligned} & p\left(s_t|\mathbf{W}_t, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) \\ = & \frac{\sum_{j=1}^{2} f\left(\mathbf{y}_t, \mathbf{f}_t|\mathbf{W}_{t-1}, \mathbf{\Omega}_t, s_{t-1}=j, s_t, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_{t-1}=j, s_t|\mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right)}{f\left(\mathbf{y}_t, \mathbf{f}_t|\mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right)} \end{aligned} \tag{3.18}$$

Then the term in the numerator can be written as

$$\begin{aligned} & f\left(\mathbf{y}_t, \mathbf{f}_t|\mathbf{W}_{t-1}, \mathbf{\Omega}_t, s_{t-1}, s_t, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_{t-1}, s_t|\mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) && (3.19) \\ = & f\left(\mathbf{y}_t|\mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{f}_t, s_t, s_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) f\left(\mathbf{f}_t|\mathbf{W}_{t-1}, \mathbf{\Omega}_t, s_t, s_{t-1}=j, \mathbf{\Theta}, \mathbf{P}\right) \\ & \times p\left(s_t|s_{t-1}=j, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_{t-1}|\mathbf{W}_{t-1}, \mathbf{\Omega}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \\ = & f\left(\mathbf{y}_t|\mathbf{f}_t, s_t, s_{t-1}, \mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) f\left(\mathbf{f}_t, s_t|s_{t-1}=j, \mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) && (3.20) \\ & \times p\left(s_{t-1}|\mathbf{W}_{t-1}, \mathbf{\Omega}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \end{aligned}$$

since

$$\begin{aligned} & f\left(\mathbf{f}_t, s_t|s_{t-1}=j, \mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) \\ = & f\left(\mathbf{f}_t|s_t, s_{t-1}=j, \mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_t|s_{t-1}=j, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) \end{aligned} \tag{3.21}$$

The first term of the equation (3.20) is easily obtained by

$$f\left(\mathbf{y}_t|\mathbf{f}_t, \mathbf{W}_{t-1}, s_t, s_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) = \mathcal{N}_q\left(\mathbf{y}_t|\mathbf{y}_{t|t-1}, D_{s_t} D'_{s_t}\right) \tag{3.22}$$

The second term is the joint density of $\mathbf{f}_t$ and s_t conditioned on $s_{t-1} = j$, and it is already given by the equations (3.13) and (3.15). The remaining term $p(s_{t-1}|\mathbf{W}_{t-1}, \mathbf{\Omega}_{t-1}, \mathbf{\Theta}, \mathbf{P})$ can be obtained from the calculation at the previous time point, which completes the inference about the numerator for non-zero values of ρ. When all the elements of ρ are zeros, this step reduces to an exogenous switching case.

With the numerator at hand, the conditional joint density of $\mathbf{y}_t$ and $\mathbf{f}_t$ which is the denominator of equation (3.18) is given by the law of total probability

$$f\left(\mathbf{y}_t, \mathbf{f}_t|\mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) \tag{3.23}$$

$$= \sum_{s_t=1}^{2} \sum_{s_{t-1}=1}^{2} f\left(\mathbf{y}_t, \mathbf{f}_t | s_{t-1}, s_t, \mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_{t-1}, s_t | \mathbf{W}_{t-1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right)$$

These calculations are initialized at $t = 1$ by treating the initial probability $p(s_0|\mathbf{W}_0,$ $\mathbf{\Omega}_0$, $\mathbf{\Theta}$, $\mathbf{P})$ as an additional parameter to be estimated or approximating it by the unconditional probability $p(s_0|\mathbf{\Theta}, \mathbf{P})$. This completes the computation of the forward pass for $t = 1, 2, .., n$.

Now the backward recursion can be carried out by the method of composition according to the scheme described in the equation (3.17). First, one can simulate s_n from $s_n|\mathbf{W}_n, \mathbf{\Omega}_n, \mathbf{\Theta}, \mathbf{P}$ given in the equation (3.18). Then, given s_{t+1}, s_t $(t = 1, 2, .., t-1)$ can be sampled from its posterior mass function

$$\begin{aligned} & p\left(s_t = j | s_{t+1}, \mathbf{W}_{t+1}, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) \\ = & \frac{p\left(s_{t+1} | s_t = j, \mathbf{W}_{t+1}, \mathbf{\Omega}_{t+1}, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_t = j | \mathbf{W}_t, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right)}{\sum_{j=1}^{2} p\left(s_{t+1} | s_t = j, \mathbf{W}_{t+1}, \mathbf{\Omega}_{t+1}, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_t = j | \mathbf{W}_t, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right)} \end{aligned} \tag{3.24}$$

where

$$\begin{aligned} & p\left(s_{t+1} | s_t = j, \mathbf{W}_{t+1}, \mathbf{\Omega}_{t+1}, \mathbf{\Theta}, \mathbf{P}\right) \\ = & \Phi\left(\frac{\alpha_j + \beta_j' \mathbf{z}_{t+1} - \rho' \left(L_{s_{t+1}}^{-1} \left(\mathbf{f}_{t+1} - \mu_{s_{t+1}} - G_{s_{t+1}} \mathbf{f}_t\right)\right)}{\sqrt{1 - \rho' \rho}}\right) \end{aligned} \tag{3.25}$$

, which is not equal to $\Phi(\alpha_j + \beta_j' \mathbf{z}_{t+1})$ for non-zero ρ. This is equivalent to updating s_t by combining $p\left(s_t | \mathbf{W}_t, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right)$ and information contained in the previously generated s_{t+1} and the given $\mathbf{f}_{t+1}$. The Monte Carlo estimate of $p\left(s_t | \mathbf{W}_n, \mathbf{\Omega}_n\right)$ can be obtained by taking the average of s_t over the MCMC iterations. It is interesting to note that when ρ is a vector of zeros, this sampling procedure exactly reduces to the method of Chib (1996).

3.1.4 Simulation of $\{\gamma_t\}$

One of the most important features in the algorithm presented in this paper is introduction an auxiliary variable γ_t into the sampling following Albert and Chib (1993). Suppose that s_t is equal to 1. Then given $s_{t-1} = j$, $\mathbf{f}_t$ and $\mathbf{f}_{t-1}$, ε_t has the value of $L_{s_t}^{-1} \left(\mathbf{f}_t - \mu_{s_t} - G_{s_t} \mathbf{f}_{t-1}\right)$. From the Bayes theorem, the conditional density of γ_t is the following.

$$f\left(\gamma_t | s_t = 1, s_{t-1} = j, \mathbf{W}_t, \mathbf{\Omega}_t, \mathbf{\Theta}, \mathbf{P}\right) \propto \mathcal{N}\left(\gamma_t | \rho' \varepsilon_t, 1 - \rho'\rho\right) \times I\left[\gamma_t < \gamma_t^*\right] \tag{3.26}$$

where $\gamma_t^* = \alpha_j + \beta_j' \mathbf{z}_t$ and $I[\cdot]$ is an indicator function. Given ρ, ε_t determines its posterior mode and the information $s_t = 1$ serves to truncate the support of γ_t. Thus once ε_t, $s_t = 1$ and $s_{t-1} = j$ are given, the posterior draw of γ_t is sampled from a normal distribution $\mathcal{N}\left(\rho'\varepsilon_t, 1 - \rho'\rho\right)$ whose value is bounded below γ_t^*, denoted by $\mathcal{TN}_{(-\infty, \gamma_t^*]}(\rho'\varepsilon_t, 1 - \rho'\rho)$. By a similar argument, the support of γ_t is (γ_t^*, ∞) when $s_t = 2$ and $s_{t-1} = j$, and thus it is sampled from $\mathcal{TN}_{(\gamma_t^*, \infty)}(\rho'\varepsilon_t, 1 - \rho'\rho)$. In summary, each of these truncated normal distributions is drawn as follows.

$$\begin{aligned} \gamma_t | \varepsilon_t, s_t &= 1, s_{t-1} = j \sim \mathcal{TN}_{(-\infty, \gamma_t^*]}(\rho'\varepsilon_t, 1 - \rho'\rho) \\ \gamma_t | \varepsilon_t, s_t &= 2, s_{t-1} = j \sim \mathcal{TN}_{(\gamma_t^*, \infty)}(\rho'\varepsilon_t, 1 - \rho'\rho) \end{aligned} \tag{3.27}$$

3.1.5 Simulation of $\{\mathbf{f}_t\}$

The MCMC sampling scheme is completed by sampling $\mathbf{F}_n$ conditioned on ($\mathbf{Y}_n$, $\mathbf{S}_n$, $\mathbf{\Omega}_n$, $\boldsymbol{\gamma}_n$, $\mathbf{\Theta}$, $\mathbf{P}$). For this, we modify and use the multi-move Gibbs-sampling suggested by Carter and Kohn (1994) that generates the whole time series of he continuous state variables $\mathbf{f}_t$ in one block. This approach consists of two steps: Kalman filter and backward updating. The objective of the Kalman filter step is to obtain the inference of $\mathbf{f}_t$ based on the information up to time t as follows.

$$\mathbf{f}_{t|t} = \mathbb{E}\left[\mathbf{f}_t | \mathbf{Y}_t, \mathbf{S}_n, \mathbf{\Omega}_n, \boldsymbol{\gamma}_n, \mathbf{\Theta}, \mathbf{P}\right] = \mathbf{f}_{t|t-1} + \mathbf{K}_t \eta_{t|t-1} \tag{3.28}$$

$$\mathbf{P}_{t|t} = \mathbb{V}\left[\mathbf{f}_t | \mathbf{Y}_t, \mathbf{S}_n, \mathbf{\Omega}_n, \boldsymbol{\gamma}_n, \mathbf{\Theta}, \mathbf{P}\right] = \left(\mathbf{I}_k - \mathbf{K}_t \bar{\mathbf{b}}_{s_t}\right) \mathbf{P}_{t|t-1} \tag{3.29}$$

where for given $\mathbf{f}_{t-1|t-1}$ and $\mathbf{P}_{t-1|t-1}$,

$$\mathbf{f}_{t|t-1} = \mu_{s_t} + G_{s_t} \mathbf{f}_{t-1|t-1} + L_{s_t} \rho \gamma_t \tag{3.30}$$

$$\mathbf{P}_{t|t-1} = G_{s_t} \mathbf{P}_{t-1|t-1} G_{s_t}' + L_{s_t} \left(\mathbf{I}_k - \rho\rho'\right) L_{s_t}' \tag{3.31}$$

$$\eta_{t|t-1} = \mathbf{y}_t - \mathbf{a}_{s_t} - \mathbf{b}_{s_t} \mathbf{f}_{t|t-1} - \mathbf{H}_{s_t} \mathbf{x}_t \tag{3.32}$$

$$f_{t|t-1} = \mathbf{b}_{s_t} \mathbf{P}_{t|t-1} \mathbf{b}_{s_t}' + D_{s_t} D_{s_t}' \tag{3.33}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{b}'_{s_t} f^{-1}_{t|t-1} \tag{3.34}$$

One distinguishing feature from the basic filter is that the last term of equation (3.30) and (3.31) exists due to the non-zero correlation between ε_t and γ_t. These recursions are initialized by setting $\mathbf{f}_{0|0}$ to be a vector of additional parameters to be estimated and $\mathbf{P}_{0|0}$ to be zero.

Given the filtered values of $\mathbf{f}_t$ and their variance-covariance (i.e. $\mathbf{f}_{t|t}$ and $\mathbf{P}_{t|t}$) from equations (3.28) and (3.29), the backward updating can be done by the standard sampling procedure

$$\mathbf{f}_t|\mathbf{Y}_t, \mathbf{S}_n, \mathbf{f}_{t+1}, \mathbf{\Omega}_n, \boldsymbol{\gamma}_n, \mathbf{\Theta}, \mathbf{P} \sim \mathcal{N}_k\left(\mathbf{f}_{t|t,\mathbf{f}_{t+1}}, \mathbf{P}_{t|t,\mathbf{f}_{t+1}}\right) \tag{3.35}$$

where

$$\mathbf{f}_{t|t,\mathbf{f}_{t+1}} = \mathbf{f}_{t|t} + \mathbf{P}_{t|t}G'_{s_{t+1}}\left(\mathbf{P}^*_{t+1|t}\right)^{-1}\left(\mathbf{f}_{t+1} - \mathbf{f}^*_{t+1|t}\right) \tag{3.36}$$

$$\mathbf{P}_{t|t,\mathbf{f}_{t+1}} = \mathbf{P}_{t|t} - \mathbf{P}_{t|t}G'_{s_{t+1}}\left(\mathbf{P}^*_{t+1|t}\right)^{-1} G_{s_{t+1}}\mathbf{P}_{t|t} \tag{3.37}$$

$$\mathbf{f}^*_{t+1|t} = \mu_{s_{t+1}} + G_{s_{t+1}}\mathbf{f}_{t|t} + L_{s_{t+1}}\rho\gamma_{t+1} \tag{3.38}$$

$$\mathbf{P}^*_{t+1|t} = G_{s_{t+1}}\mathbf{P}_{t|t}G'_{s_{t+1}} + L_{s_{t+1}}\left(\mathbf{I}_k - \rho\rho'\right)L'_{s_{t+1}} \tag{3.39}$$

Note that this sampling stage is not necessary when the continuous state variables are observable.

3.2 Marginal Likelihood Calculation

One of the goals of this study is to evaluate the extent to which the endogenous regime switching model is an improvement over the exogenous regime-switching model or non-switching model. For this, we do the comparison in terms of marginal likelihoods and their ratios, Bayes factors. As in the previous sections, we suppress the dependence on the model indicator $\mathcal{M}$ in our notation since all the MCMC computations in this study must be repeated for all competing models. The marginal likelihood of any given model $m(\mathbf{Y}_n|\mathbf{\Omega}_n)$ is obtained as

$$m(\mathbf{Y}_n|\mathbf{\Omega}_n) = \int p(\mathbf{Y}_n|\mathbf{\Omega_n}, \mathbf{\Theta}, \mathbf{P})\pi(\mathbf{\Theta}, \mathbf{P})d(\mathbf{\Theta}, \mathbf{P}) \tag{3.40}$$

As the likelihood density of $\mathbf{y}_t$ conditioned on the model specification is not standard, the direct computation of the marginal likelihood is not feasible. Thus, a simulation-based approach is employed. As is well known, provided one has an estimate of posterior ordinate $\pi(\boldsymbol{\Theta}^*, \mathbf{P}^*|\mathbf{Y}_n)$ the marginal likelihood can be estimated on the log scale as

$$\ln \hat{m}(\mathbf{Y}_n|\boldsymbol{\Omega}_n) = \ln f\left(\mathbf{Y}_n|\boldsymbol{\Omega}_n, \boldsymbol{\Theta}^*, \mathbf{P}^*\right) + \ln \pi(\boldsymbol{\Theta}^*, \mathbf{P}^*) - \ln \hat{\pi}(\boldsymbol{\Theta}^*, \mathbf{P}^*|\mathbf{Y}_n) \tag{3.41}$$

where $(\boldsymbol{\Theta}^*, \mathbf{P}^*)$ is some specified (e.g., high-density) point of $(\boldsymbol{\Theta}, \mathbf{P})$.

Note that the first term of the right-hand side of this expression is the likelihood ordinate evaluated at the single point. It can be obtained by integrating out the state variables.

$$\begin{aligned} & f\left(\mathbf{Y}_n|\boldsymbol{\Omega}_n, \boldsymbol{\Theta}^*, \mathbf{P}^*\right) \\ = & \int f\left(\mathbf{Y}_n|\mathbf{F}_n, \mathbf{S}_n, \boldsymbol{\Omega}_n, \boldsymbol{\Theta}^*, \mathbf{P}^*\right) p\left(\mathbf{F}_n, \mathbf{S}_n|\boldsymbol{\Omega}_n, \boldsymbol{\Theta}^*, \mathbf{P}^*\right) d\left(\mathbf{F}_n, \mathbf{S}_n\right) \end{aligned} \tag{3.42}$$

where $f\left(\mathbf{Y}_n|\mathbf{F}_n, \mathbf{S}_n, \boldsymbol{\Omega}_n, \boldsymbol{\Theta}^*, \mathbf{P}^*\right)$ is the joint density of observations conditioned on the state variables and the model parameters, and $p\left(\mathbf{F}_n, \mathbf{S}_n|\boldsymbol{\Omega}_n, \boldsymbol{\Theta}^*, \mathbf{P}^*\right)$ is the prior distribution of the state variables given the model parameters. However, because of the non-zero correlation between the joint dynamics of the state variables, this integration is obviously infeasible by direct means. The next subsection shows how the likelihood estimation can be done by simulation.

The calculation of the prior density, the second term of the right-hand side of the equation (3.41), is straightforward. Finally, the third one, which is the log posterior ordinate, is obtained from a marginal-conditional decomposition following Chib and Jeliazkov (2001). All are evaluated at $(\boldsymbol{\Theta}^*, \mathbf{P}^*)$.

3.2.1 Particle Filtering and Likelihood Estimation

As mentioned earlier, we complete the calculation of the marginal likelihood by estimating the likelihood function. This is calculated based on the earlier work of Chib, Nardari, and Shephard (2002) and Kaufmann (2000), and a particle filtering method is employed. Chib et al. (2002) discuss a particle filter for the stochastic volatility model that is expressed as a state space model with exogenous and independent switching. Kaufmann

(2000) estimates the likelihood function of a factor model with Markov switching in the transition equation by particle filtering. However, their particle filter method has to be modified before it can be applied to our model with endogenous first-order Markov regime switching parameters. The main issue in the likelihood inference by simulation is updating the distribution of the state variables $(s_t, \mathbf{f}_t)$ conditioned on the current information at time t. As a part of the particle filtering procedure the auxiliary particle filter, introduced by Pitt and Shephard (1999), is implemented to obtain draws of $(s_t, \mathbf{f}_t)$ conditioned on $(\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})$ where $\mathcal{F}_t = \{\mathbf{Y}_t, \mathbf{\Omega}_n\}$ denotes the history of the observations up to time t. This auxiliary particle filtering procedure recursively generates the sequence of draws of $\{s_t, \mathbf{f}_t\}_{t=1,2,..,n}$ from the filtered distributions $p(s_t, \mathbf{f}_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})$.

From the Bayes theorem we have

$$\begin{aligned} & p\left(s_t, \mathbf{f}_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}\right) \\ \propto \; & \mathcal{N}_q\left(\mathbf{y}_t|\mathbf{a}_{s_t} + \mathbf{b}_{s_t}\mathbf{f}_t + H_{s_t}\mathbf{x}_t, D_{s_t}D'_{s_t}\right) p\left(s_t, \mathbf{f}_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \end{aligned} \tag{3.43}$$

where

$$\begin{aligned} & p\left(s_t, \mathbf{f}_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \\ = & \int p\left(s_t, \mathbf{f}_t|s_{t-1}, \mathbf{f}_{t-1}, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_{t-1}, \mathbf{f}_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) d\left(s_{t-1}, \mathbf{f}_{t-1}\right). \end{aligned} \tag{3.44}$$

Suppose that we have $\{s^j_{t-1}, \mathbf{f}^j_{t-1}\}_{j=1,2,..,M} \sim p(s_{t-1}, \mathbf{f}_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})$. Then it follows from the equations (3.13) and (3.15) that the integral of $p\left(s_t, \mathbf{f}_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right)$ can be approximated as

$$\begin{aligned} & p\left(s_t = 1, \mathbf{f}_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \\ \simeq \; & \frac{1}{M}\sum_{j=1}^{M} \frac{\phi_k\left(L_1^{-1}\left(\mathbf{f}_t - \mu_1 - G_1\mathbf{f}^j_{t-1}\right)\right)}{|L_1|} \Phi\left(\frac{\alpha_{s^j_{t-1}} + \beta'_{s^j_{t-1}}\mathbf{z}_t - \rho'\left(L_1^{-1}\left(\mathbf{f}_t - \mu_1 - G_1\mathbf{f}^j_{t-1}\right)\right)}{\sqrt{1-\rho'\rho}}\right) \end{aligned} \tag{3.45}$$

and

$$\begin{aligned} & p\left(s_t = 2, \mathbf{f}_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \\ \simeq \; & \frac{1}{M}\sum_{j=1}^{M} \frac{\phi_k\left(L_2^{-1}\left(\mathbf{f}_t - \mu_2 - G_2\mathbf{f}^j_{t-1}\right)\right)}{|L_2|} \Phi\left(\frac{-\alpha_{s^j_{t-1}} - \beta'_{s^j_{t-1}}\mathbf{z}_t + \rho'\left(L_2^{-1}\left(\mathbf{f}_t - \mu_2 - G_2\mathbf{f}^j_{t-1}\right)\right)}{\sqrt{1-\rho'\rho}}\right) \end{aligned} \tag{3.46}$$

To sample $(s_t, \mathbf{f}_t)$ from the approximated filtered distribution $p\left(s_t, \mathbf{f}_t | \mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}\right)$ implied by the equations (3.43), (3.45) and (3.46) we employ the auxiliary particle filtering method. This filtering procedure uses the importance sampling method, in which, first, the values of the state variables are proposed and subsequently, they are reweighted to produce draws from the target distribution. For each time point t the steps of our auxiliary particle filter algorithm can be summarized as follows.

Algorithm 2: Auxiliary particle filter

Step 1.1 $h_{t-1}^j = (s_{t-1}^j, \mathbf{f}_{t-1}^j)$, $h_{t-1} = (s_{t-1}, \mathbf{f}_{t-1})$, $M = 10,000$ and $R = 50,000$ are set. Given $\{h_{t-1}^1, h_{t-1}^2, ..., h_{t-1}^M\}$ from $(h_{t-1} | \mathcal{F}_{t-1}, \mathbf{\Theta}^*, \mathbf{P}^*)$, sample $\{\hat{\gamma}_t^{*1}, \hat{\gamma}_t^{*2}, ..., \hat{\gamma}_t^{*M}\}$ from $\mathcal{N}_1(0,1)$.

Step 1.2 Given $\{\hat{\gamma}_t^{*1}, \hat{\gamma}_t^{*2}, ..., \hat{\gamma}_t^{*M}\}$ and $\{h_{t-1}^1, h_{t-1}^2, ..., h_{t-1}^M\}$, sample $\{\hat{s}_t^{*1}, \hat{s}_t^{*2}, ..., \hat{s}_t^{*M}\}$ from

$$\hat{s}_t^{*j} = \begin{cases} 1 & \text{if } \hat{\gamma}_t^{*j} < \alpha_{s_{t-1}^j} + \beta'_{s_{t-1}^j}\mathbf{z}_t \\ 2 & \text{if } \hat{\gamma}_t^{*j} > \alpha_{s_{t-1}^j} + \beta'_{s_{t-1}^j}\mathbf{z}_t \end{cases} \quad \text{for each } j = 1, .., M \tag{3.47}$$

Step 1.3 Given $\{\hat{s}_t^{*j}, \hat{\gamma}_t^{*j}, \mathbf{f}_{t-1}^j\}_{j=1,2,..,M}$, calculate

$$\begin{aligned} \hat{\mathbf{f}}_t^{*j} &= \mu_{\hat{s}_t^{*j}} + G_{\hat{s}_t^{*j}}\mathbf{f}_{t-1}^j + L_{\hat{s}_t^{*j}}\rho\hat{\gamma}_t^{*j} && (3.48) \\ w_j &= \mathcal{N}_q\left(\mathbf{y}_t | \mathbf{a}_{\hat{s}_t^{*j}} + \mathbf{b}_{\hat{s}_t^{*j}}\hat{\mathbf{f}}_t^{*j} + H_{\hat{s}_t^{*j}}\mathbf{x}_t, D_{\hat{s}_t^{*j}}D'_{\hat{s}_t^{*j}}\right) && (3.49) \end{aligned}$$

Step 1.4 Sample R times the integers $1, 2, .., M$ with probability proportional to $\{w_j^* = w_j / \sum_{j=1}^M w_j\}$, and let the sampled indices be $k_1, k_2, .., k_R$ and associate these with $\{s_{t-1}^{k_i}, \mathbf{f}_{t-1}^{k_i}\}_{i=1,2,..,R}$ and $\{\hat{s}_t^{*k_i}, \hat{\mathbf{f}}_t^{*k_i}\}_{i=1,2,..,R}$.

Step 2.1 For each k_i simulate

$$\begin{aligned} s_t^{*i} &= \begin{cases} 1 & \text{if } \gamma_t^{*k_i} < \alpha_{s_{t-1}^{k_i}} + \beta'_{s_{t-1}^{k_i}}\mathbf{z}_t \\ 2 & \text{if } \gamma_t^{*k_i} > \alpha_{s_{t-1}^{k_i}} + \beta'_{s_{t-1}^{k_i}}\mathbf{z}_t \end{cases} \quad \text{for each } i = 1, 2, .., R && (3.50) \\ \text{and } \mathbf{f}_t^{*i} &= \mu_{s_t^{*i}} + G_{s_t^{*i}}\mathbf{f}_{t-1}^{k_i} + L_{s_t^{*i}}\varepsilon_t^i && (3.51) \end{aligned}$$

where

$$\begin{bmatrix} \varepsilon_t^i \\ \gamma_t^i \end{bmatrix} \sim \text{i.i.d.}\mathcal{N}_{k+1}\left(0_{(k+1)\times 1}, \begin{pmatrix} \mathbf{I}_k & \rho \\ \rho' & 1 \end{pmatrix}\right)$$

Step 2.2 Resample $\{s_t^{*i}, \mathbf{f}_t^{*i}\}_{i=1,2,..,R}$ M times with probabilities proportional to the importance weights $\{W_i^* = W_i / \sum_{i=1}^{R} W_i\}$ where

$$W_i = \frac{\mathcal{N}_q\left(\mathbf{y}_t|\mathbf{a}_{s_t^{*i}} + \mathbf{b}_{s_t^{*i}}\mathbf{f}_t^{*i} + H_{s_t^{*i}}\mathbf{x}_t, D_{s_t^{*i}}D'_{s_t^{*i}}\right)}{\mathcal{N}_q\left(\mathbf{y}_t|\mathbf{a}_{\hat{s}_t^{*k_i}} + \mathbf{b}_{\hat{s}_t^{*k_i}}\hat{\mathbf{f}}_t^{*k_i} + H_{\hat{s}_t^{*k_i}}\mathbf{x}_t, D_{\hat{s}_t^{*k_i}}D'_{\hat{s}_t^{*k_i}}\right)} \tag{3.52}$$

, which produces the filtered sample $\{s_t^j, \mathbf{f}_t^j\}_{j=1,2,..,M}$ from $(s_t, \mathbf{f}_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})$.

Once one has a sample from $(s_t, \mathbf{f}_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})$, one can estimate the log likelihood ordinate by using the particle filtering steps. Specifically, the draws $(s_{t-1}, \mathbf{f}_{t-1})$ from $(s_{t-1}, \mathbf{f}_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})$ based on Algorithm 2 allow us to sample $(s_t, \mathbf{f}_t)$ from as $(s_t, \mathbf{f}_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})$. Next, one can approximate the one-step ahead predictive density of $\mathbf{y}_t$

$$\begin{aligned} & f\left(\mathbf{y}_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \\ = & \int f\left(\mathbf{y}_t|s_t, \mathbf{f}_t, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_t, \mathbf{f}_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) d\left(s_t, \mathbf{f}_t\right) \end{aligned} \tag{3.53}$$

by averaging $\mathcal{N}_q(\mathbf{y}_t|\mathbf{a}_{s_t} + \mathbf{b}_{s_t}\mathbf{f}_t + H_{s_t}\mathbf{x}_t, D_{s_t}D'_{s_t})$ over those draws $(s_t, \mathbf{f}_t)$. These filtering steps can be summarized as follows.

Algorithm 3: Likelihood Function

Step 1 Set $t = 1$, initialize $(\mathbf{\Theta}, \mathbf{P})$ and obtain a sample of $\{s_{t-1}^{(g)}, \mathbf{f}_{t-1}^{(g)}\}_{g=1,..,M}$ from $(s_{t-1}, \mathbf{f}_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})$.

Step 2 Sample $\{\gamma_t^{(g)}, \varepsilon_t^{(g)}\}_{g=1,..,M}$ from

$$\mathcal{N}_{k+1}\left(0_{(k+1)\times 1}, \begin{pmatrix} \mathbf{I}_k & \rho \\ \rho' & 1 \end{pmatrix}\right) \tag{3.54}$$

Step 3 Given $\{\gamma_t^{(g)}, \mathbf{f}_{t-1}^{(g)}\}_{g=1,..,M}$, sample $\{s_t^{(g)}\}_{g=1,..,M}$ from

$$s_t^{(g)} = \begin{cases} 1 & \text{if } \gamma_t^{(g)} < \alpha_{s_{t-1}^{(g)}} + \beta'_{s_{t-1}^{(g)}}\mathbf{z}_t \\ 2 & \text{if } \gamma_t^{(g)} > \alpha_{s_{t-1}^{(g)}} + \beta'_{s_{t-1}^{(g)}}\mathbf{z}_t \end{cases} \quad \text{for each } g = 1, 2, .., M \tag{3.55}$$

Step 4 Given $\{\mathbf{f}_{t-1}^{(g)}, s_t^{(g)}, \varepsilon_t^{(g)}\}_{g=1,..,M}$, calculate

$$\mathbf{f}_t^{(g)} = \mu_{s_t^{(g)}} + G_{s_t^{(g)}}\mathbf{f}_{t-1}^{(g)} + L_{s_t^{(g)}}\varepsilon_t^{(g)} \tag{3.56}$$

Step 5 Estimate the one-step ahead predictive density as

$$\hat{f}\left(\mathbf{y}_t|\mathcal{F}_{t-1},\boldsymbol{\Theta},\mathbf{P}\right) \simeq \frac{1}{M}\sum_{g=1}^{M}\mathcal{N}_q\left(\mathbf{y}_t|\mathbf{a}_{s_t^{(g)}}+\mathbf{b}_{s_t^{(g)}}\mathbf{f}_t^{(g)}+H_{s_t^{(g)}}\mathbf{x}_t, D_{s_t^{(g)}}D'_{s_t^{(g)}}\right) \tag{3.57}$$

Step 6 Apply the auxiliary particle filter in Algorithm 2 to draw $\{s_t^{(g)},\mathbf{f}_t^{(g)}\}_{g=1,..,M}$ from $(s_t,\mathbf{f}_t|\mathcal{F}_t,\boldsymbol{\Theta},\mathbf{P})$, set $t=t+1$ and goto Step 2.

Step 7 Return the log likelihood ordinate

$$\ln f\left(\mathbf{Y}_n|\boldsymbol{\Omega},\boldsymbol{\Theta},\mathbf{P}\right) = \sum_{t=1}^{n}\ln\hat{f}\left(\mathbf{y}_t|\mathcal{F}_{t-1},\boldsymbol{\Theta},\mathbf{P}\right) \tag{3.58}$$

, which is a consistent estimate of the log-likelihood as demonstrated by Omori, Chib, Shephard, and Nakajima (2007).

Of course, a straightforward modification of the algorithm is needed for the exogenous regime switching model because the endogenous switching model nests it.

4 Examples

4.1 Simulation Study

This subsection provides an evidence of efficient performance of the MCMC algorithm, and illustrates the importance of the assumption of endogenous regime switchings based on a simulation study. Consider a bivariate dynamic common factor as an example of state space models with $n = 1,000$ observations. A sequence of 2×1 vector of observations $\{y_t\}$ is assumed to be generated from the following process.

$$\mathbf{y}_t = \mathbf{f}_t + D_{s_t}e_t \tag{4.1}$$

$$\mathbf{f}_t = \mu_{s_t} + L_{s_t}\varepsilon_t \tag{4.2}$$

where $e_t \sim$i.i.d.$\mathcal{N}_2(0,\mathbf{I}_2)$ and $\varepsilon_t \sim$i.i.d.$\mathcal{N}_1(0,\mathbf{I}_1)$. In simple terms, the unobserved common scalar $\mathbf{f}_t$ is a regime switching process in mean and variance, and the regime $\{s_t\}$ is governed by a two-state Markov-switching process with fixed transition probabilities. The variance of the measure errors are also subject to regime shifts over time. Under the notations in the previous sections, $q = 2$, $k = 1$, $h = 0$, $\mathbf{a}_{s_t} = H_{s_t} = G_{s_t} = 0$, $\beta_{s_t} = 0$ and $\mathbf{b}_{s_t} = 1$.

Table 1 shows the results of the simulation experiment examining the proposed MCMC method of the endogenous and exogenous switching models where (i, i) element of D_{s_t} denoted by $d_{s_t}^{(i)}$ under diffuse prior. For all the parameters, each table shows the mean of posterior distributions along with 95% credibility intervals. In particular, Table 1(b) clearly demonstrates the inaccuracy of estimation that occurs when the endogenous regime process is treated as exogenous whereas the estimates in Table 1(a) are close to the corresponding true values, and all true values are contained in the 95% credible intervals. The estimates for μ_{s_t} in Table 1(b) are far from the true values, and this becomes more severe for higher values of ρ. The sign of the difference between the true values and estimates depends on the direction of the regime switchings. In this example, μ_1 is underestimated whereas μ_2 is overestimated. This is because the regime shifts from regime 1 to regime 2 are always negative in terms of the μ_{s_t}'s and the correlation coefficient ρ is positive. Finally, it should be noted that the MCMC algorithm presented in this study is highly efficient in the sense that the sampled draws for all model parameters display very low inefficiency factors around 7.5[5].

4.2 Application: A three-factor Nelson-Siegel yield curve model

As an empirical application, let us now consider a three-factor Nelson-Siegel model that originally motivates the proposed modeling approach. It is well-known that the dynamic Nelson-Siegel model is commonly used in practice for both fitting and forecasting the term structure of interest rates due to its convenient and parsimonious functional form. Following Diebold and Li (2006), the vector of yields with τ period maturity $y_t(\tau)$

$$\mathbf{y}_t = \begin{pmatrix} y_t(\tau_1) & y_t(\tau_2) & \cdots & y_t(\tau_N) \end{pmatrix}$$

is statistically modeled by

$$\mathbf{y}_t = \mathbf{\Lambda} \times \mathbf{f}_t + D_{s_t} e_t \tag{4.3}$$

[5] For any sampled sequence of posterior draws the inefficiency factor is computed as $1+2\sum_{k=1}^{500} AC(k)$ where $AC(k)$ is the k-order autocorrelation computed from the sampled variates.

parameters	$s_t = 1$				$s_t = 2$			
	true	mean	2.5%	97.5%.	true	mean	2.5%	97.5%.
μ_{s_t}	4.000	4.057	3.699	4.420	-4.000	-3.954	-4.680	-3.219
$d_{s_t}^{(1)}$	1.000	1.000	0.8943	0.131	0.500	0.581	0.338	0.722
$d_{s_t}^{(2)}$	9.000	9.306	8.660	9.984	9.000	8.917	8.274	9.602
α_{s_t}	0.000	0.142	0.009	0.277	0.000	-0.046	-0.186	0.093
β_{s_t}	1.000	0.989	0.833	1.162	-1.000	-1.039	-1.175	-0.907
L	1.000	1.000	0.904	1.103	6.000	5.860	5.066	6.634
ρ	0.600	0.626	0.563	0.716				

(a) unrestricted model ($\rho \neq 0$)

parameters	$s_t = 1$				$s_t = 2$			
	true	mean	2.5%	97.5%.	true	mean	2.5%	97.5%.
μ_{s_t}	4.000	3.245	2.946	3.547	-4.000	-2.838	-3.480	-2.206
$d_{s_t}^{(1)}$	0.500	0.097	0.012	0.378	0.500	0.690	0.597	0.744
$d_{s_t}^{(2)}$	9.000	9.375	8.726	10.067	9.000	9.087	8.409	9.756
α_{s_t}	0.000	0.033	-0.107	0.175	0.000	-0.063	-0.211	0.085
β_{s_t}	1.000	0.890	0.758	1.031	-1.000	-1.012	-1.158	-0.874
L	4.000	3.868	3.668	4.081	6.000	5.505	4.725	6.322

(b) restricted model ($\rho = 0$)

Table 1: **Estimates of Model Parameters** *This table presents the true values, the posterior mean and the 95% credibility intervals of the model parameters based on 50,000 MCMC draws beyond a burn-in of 5,000.*

where

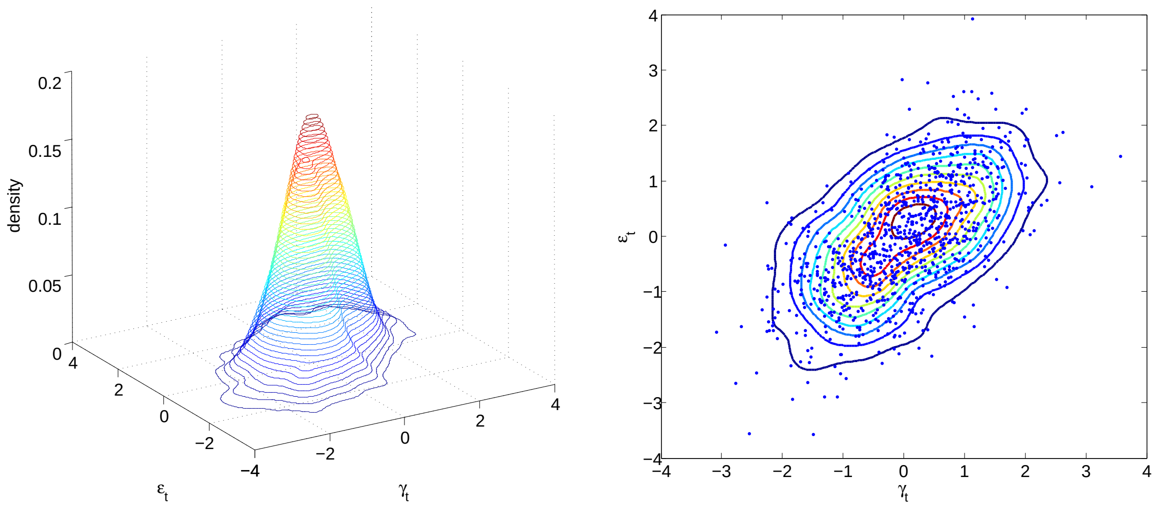
$$\mathbf{\Lambda} = \begin{pmatrix} 1 & \frac{1-e^{\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{\tau_1\lambda}}{\tau_1\lambda} - e^{\tau_1\lambda} \\ 1 & \frac{1-e^{\tau_2\lambda}}{\tau_1\lambda} & \frac{1-e^{\tau_2\lambda}}{\tau_1\lambda} - e^{\tau_2\lambda} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{\tau_N\lambda}}{\tau_1\lambda} & \frac{1-e^{\tau_N\lambda}}{\tau_1\lambda} - e^{\tau_N\lambda} \end{pmatrix} \tag{4.4}$$

$$\mathbf{f}_t = \begin{pmatrix} \mathbf{f}_t^L & \mathbf{f}_t^S & \mathbf{f}_t^C \end{pmatrix}' \tag{4.5}$$

$$e_t = \begin{pmatrix} e_t(\tau_1) & e_t(\tau_2) & \cdots & e_t(\tau_N) \end{pmatrix}' \tag{4.6}$$

The latent dynamic factors, $\mathbf{f}_t^L$, $\mathbf{f}_t^S$ and $\mathbf{f}_t^C$ are usually interpreted as level, slope and curvature, respectively. The vector of the dynamic factors $\mathbf{f}_t$ is assumed to follow the first-order stationary vector autoregressive process.

$$\mathbf{f}_t = \mu_{s_t} + G\mathbf{f}_{t-1} + L_{s_t}\varepsilon_t \tag{4.7}$$



(a) 3 dimensional kernel smoothed plot (b) 2 dimensional contour plot

Figure 3: **Kernel smoothed bivariate plots for estimated γ_t and ε_t** *These graphs are based on 50,000 simulated draws of the posterior simulation.*

The coefficient λ, referring to the shape parameter, determines the exponential decay rate of the factor loadings, $\mathbf{\Lambda}$. Following Diebold and Li (2006) the value is fixed at 0.125, so that it maximizes the loadings on the curvature factor at 12 months. In order to deal with the potential heteroscedasticity it is assumed that the volatility of the measurement errors and the factor shocks, denoted by diagonal matrices D_{s_t} and L_{s_t} respectively, varies over time depending on a regime indicator s_t. The intercept term in the transition equation μ_{s_t} is also assumed to be regime-specific. The identification restriction that the (1,1) element of $D_{s_t=1}$ is less than that of $D_{s_t=2}$ is imposed through the prior used in this study. Hence, regime 1 (i.e. $s_t = 1$) corresponds to the high volatility state of the economy. For simplicity, we maintain the assumption that it follows a two-state Markov process with constant transition probabilities.

$$s_t = \begin{cases} 1 & \text{if } \gamma_t < \alpha_{s_{t-1}} \\ 2 & \text{if } \gamma_t > \alpha_{s_{t-1}} \end{cases} \tag{4.8}$$

Theoretically, the risk of long-term bonds, which is usually measured by the slope or curvature in this framework, depends on the size of the conditional volatility. Therefore, it is natural to consider the potentially active feedback between the the dynamic slope or curvature and the conditional volatility although this possibility has not been examined by the literature in the context of Nelson-Siegel models. To do that, the regime switching is made endogenous by allowing for contemporaneous correlation between the factor shocks ε_t and the latent variable γ_t (or s_t) whereas ε_t and e_t are independently normally distributed. These assumptions can be summarized by

$$\begin{bmatrix} e_t \\ \varepsilon_t \\ \gamma_t \end{bmatrix} \sim \text{i.i.d.}\mathcal{N}\left(0_{(k+1)\times 1}, \begin{pmatrix} I_N & 0_{N\times 3} & 0_{N\times 1} \\ 0_{3\times N} & I_3 & \rho \\ 0_{1\times N} & \rho' & 1 \end{pmatrix}\right) \tag{4.9}$$

where $\rho = \begin{pmatrix} \rho_1 & \rho_2 & \rho_3 \end{pmatrix}'$.

Note that the equations (4.3) through (4.9) immediately form a state-space system conditioned on s_t, and it belongs to the class of models discussed in this paper. The empirical results are based on the diffuse prior, as shown earlier, as well as the collection of monthly historical yields of treasury bills with maturities 1, 2, 3, 4, 6, 8, 12, 16 and 20 quarters for the sample period 1984:M1 to 2008:M12. These data are available online from the Board of Governors of the Federal Reserve System (Gurkaynak, Sack, and Wright (2007))

parameters	density	range	mean	S.D.
$\mu_{s_t}^{(1)}$	normal	$(-\infty,+\infty)$	0.10	1.00
$\mu_{s_t}^{(2)}$	normal	$(-\infty,+\infty)$	-0.10	1.00
$\mu_{s_t}^{(3)}$	normal	$(-\infty,+\infty)$	-0.05	1.00
α_1	normal	$(-\infty,+\infty)$	0.50	1.00
α_2	normal	$(-\infty,+\infty)$	-0.50	1.00
$G^{(1)}$	beta	(0,1)	0.95	0.05
$G^{(2)}$	beta	(0,1)	0.95	0.05
$G^{(3)}$	beta	(0,1)	0.95	0.05
ρ_1	uniform	(-1,1)	0.00	0.58
ρ_2	uniform	(-1,1)	0.00	0.58
ρ_3	uniform	(-1,1)	0.00	0.58
$d_{s_t}^*$	inverse gamma	$(0,+\infty)$	1.00	0.50
$10\times L_{s_t}$	inverse gamma	$(0,+\infty)$	5.00	1.00

Table 2: **Prior distribution for the endogenous regime switching model parameters** $\mu_{s_t}^{(i)}$ denotes the ith element of μ_{s_t} and $G^{(i)}$ is the (i,i) element of G

In this paper, the prior on the parameters is formulated based on the simulation-based method suggested by (Chib and Ergashev (2009)). Our prior is that the yield curve is gently upward sloping on average allowing for considerable apriori variation of the yield curve between -5% and 30%. Furthermore, the prior distribution of the regime-specific parameters is identical across regimes, so that the regimes are identified by the information from the data rather than our prior. All parameters are assumed to be a priori independent. The vector of the initial latent factors $\mathbf{f}_0$ follows the steady-state distribution in the regime at time 0 conditioned on s_0. Finally, the standard deviation of the unknown pricing errors, $d_{s_t}^{(i)\prime}s$ is collected in reparameterized form as

$$d_{s_t}^* = \{d_{s_t}^{*(i)} = C_{s_t}^{(i)} d_{s_t}^{(i)},\ i = 1, 2, .., 9\}$$

where $C_1^{(1)} = C_2^{(3)} = C_2^{(4)} = C_2^{(6)} = C_2^{(7)} = 100$, $C_1^{(2)} = C_1^{(3)} = C_1^{(4)} = 2000$, $C_1^{(5)} = 500$, $C_1^{(6)} = 400$, $C_1^{(7)} = 50$, $C_1^{(8)} = 20$, $C_1^{(9)} = C_2^{(8)} = C_2^{(9)} = 10$, $C_2^{(1)} = 5$, $C_2^{(2)} = 25$. These positive multipliers are useful to increase the magnitude of the measurement error variances and mitigate the numerical difficulty. Table 2 summarizes our prior.

parameters	Unrestricted model ($\rho \neq 0$)				Restricted model ($\rho = 0$)			
	mean	95% C.I.		ineff.	mean	95% C.I.		ineff.
$G^{(1)}$	0.975	0.961	0.986	46.56	0.978	0.963	0.992	14.54
$G^{(2)}$	0.952	0.930	0.972	7.32	0.944	0.913	0.974	4.04
$G^{(3)}$	0.866	0.820	0.911	18.35	0.872	0.830	0.914	7.43
$f_0^{(1)}$	11.684	11.083	12.262	4.53	11.708	11.099	12.309	1.39
$f_0^{(2)}$	-2.650	-3.305	-2.012	2.86	-2.620	-3.293	-1.937	1.39
$f_0^{(3)}$	-2.444	-3.323	-1.570	1.36	-2.441	-3.299	-1.586	1.12
ρ_1	-0.655	-0.736	-0.565	14.92	0.000			
ρ_2	-0.609	-0.688	-0.526	7.25	0.000			
ρ_3	0.064	-0.028	0.156	13.86	0.000			

Table 3: **Estimates of regime-independent parameters** *This table presents the posterior mean, 95% credibility interval and inefficiency factor (ineff.) of the regime-independent parameters of both restricted and unrestricted models based on 50,000 MCMC draws beyond a burn-in of 5,000.* $d_{s_t}^{(i)}$ *denotes the ith diagonal element of* D_{s_t}

We now discuss the posterior estimates of the parameters. Tables 3-5 summarize the posterior distribution of the model parameters from the restricted and unrestricted models based on 50,000 of the MCMC algorithm beyond a burn-in of 5,000. The efficiency of the MCMC sampling is measured in terms of the acceptance rate in the M-H step and the inefficiency factors. These values are found to be 62.3% and 13.3 on average, respectively, indicating a well-mixing, efficient sampler. Also, the sampler converges quickly

parameters		$s_t = 1$				$s_t = 2$		
	mean	95% C.I.		ineff.	mean	95% C.I.		ineff.
$\mu_{s_t}^{(1)}$	0.041	-0.046	0.142	38.99	0.332	0.219	0.450	31.09
$\mu_{s_t}^{(2)}$	-0.140	-0.199	-0.080	10.19	0.139	0.040	0.240	10.64
$\mu_{s_t}^{(3)}$	-0.049	-0.107	0.009	4.26	-0.342	-0.529	-0.159	15.07
$d_{s_t}^{*(1)}$	1.136	1.040	1.240	11.23	0.725	0.642	0.815	4.70
$d_{s_t}^{*(2)}$	0.663	0.407	1.066	94.51	0.814	0.706	0.926	4.07
$d_{s_t}^{*(3)}$	3.641	3.341	3.956	3.89	0.426	0.293	0.582	21.61
$d_{s_t}^{*(4)}$	0.590	0.364	0.875	59.02	0.369	0.265	0.492	23.47
$d_{s_t}^{*(5)}$	1.712	1.566	1.869	3.40	5.288	4.696	5.918	3.46
$d_{s_t}^{*(6)}$	0.493	0.328	0.702	29.44	4.204	3.748	4.688	3.31
$d_{s_t}^{*(7)}$	1.458	1.340	1.584	2.26	0.691	0.397	1.129	35.75
$d_{s_t}^{*(8)}$	1.530	1.407	1.659	2.19	1.193	1.065	1.329	3.02
$d_{s_t}^{*(9)}$	1.328	1.221	1.445	2.84	2.694	2.413	2.998	2.16
$10 \times L_{s_t}^{(1)}$	3.880	3.546	4.234	6.29	4.870	4.359	5.437	7.01
$10 \times L_{s_t}^{(2)}$	4.176	3.823	4.551	6.64	5.228	4.702	5.815	4.74
$10 \times L_{s_t}^{(3)}$	4.586	4.194	4.999	9.38	7.433	6.645	8.288	5.51
α_{s_t}	0.557	0.402	0.724	15.75	-0.138	-0.343	0.057	19.77

Table 4: **Estimates of regime-independent parameters (Unrestricted model ($\rho \neq 0$))** *This table presents the posterior mean, 95% credibility interval and inefficiency factor (ineff.) of the regime-independent parameters based on 50,000 MCMC draws beyond a burn-in of 5,000.*

to the same region of the parameter space regardless of the starting values. Finally, as one can see in Figure 4, the posterior densities of the parameters are generally different from the prior given in Table 2, which implies that the data carry information distinct from that contained in the prior.

Table 3 summarizes the posterior distribution of the regime-independent parameters providing a strong evidence of endogenous regime switching. The 95% credibility intervals of ρ_1 and ρ_2 are entirely negative. The negative value of ρ_2 indicates that the conditional volatility and the risk of long-term bonds holdings, not surprisingly, move in the same direction because the estimated $\mathbf{f}_t^S$ is indeed the negative spread as shown in Figure 6. The negative ρ_1 means that when the level becomes lower, it tends to be associated with the higher volatility because the short rate falls relatively more rapidly. Therefore, one can conclude that the regime switching in volatility is endogenous and

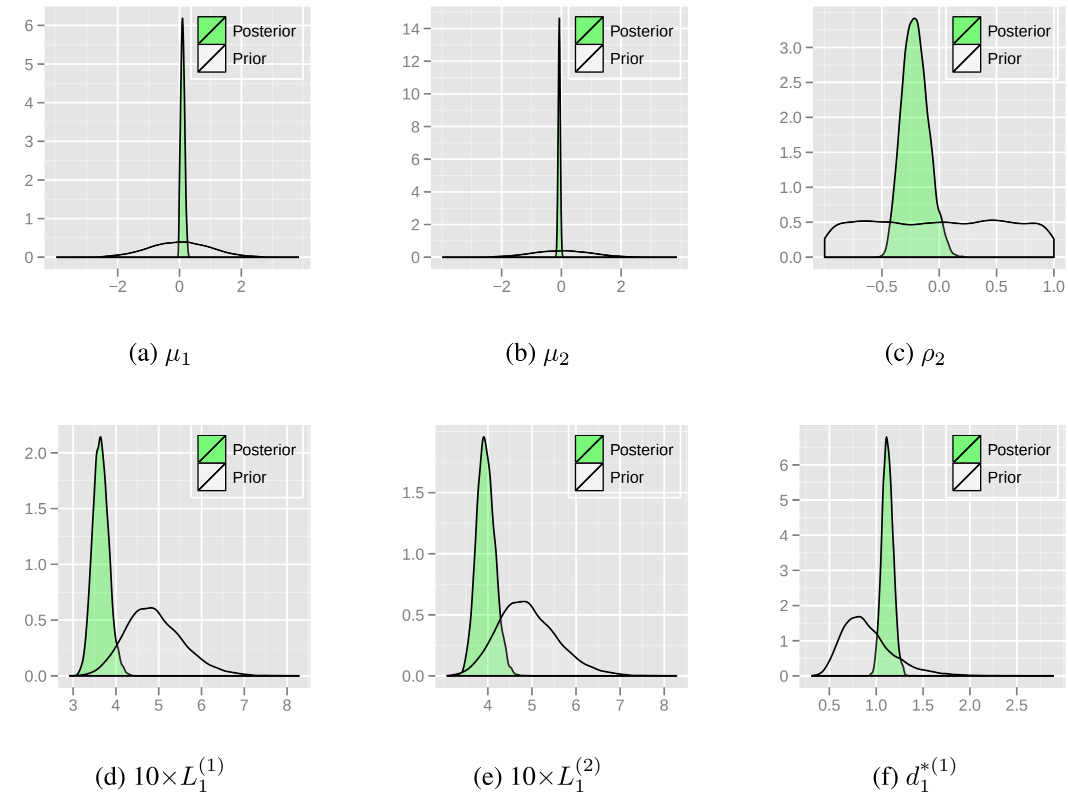
parameters		$s_t = 1$				$s_t = 2$		
	mean	95% C.I.		ineff.	mean	95% C.I.		ineff.
$\mu_{s_t}^{(1)}$	0.132	0.027	0.244	13.87	0.064	-0.044	0.174	12.90
$\mu_{s_t}^{(2)}$	-0.049	-0.108	0.007	2.50	-0.114	-0.212	-0.016	6.98
$\mu_{s_t}^{(3)}$	-0.052	-0.109	0.005	2.55	-0.303	-0.463	-0.146	8.35
$d_{s_t}^{*(1)}$	1.115	1.024	1.214	7.72	0.733	0.651	0.823	2.71
$d_{s_t}^{*(2)}$	0.690	0.420	1.064	82.75	0.831	0.727	0.943	6.01
$d_{s_t}^{*(3)}$	3.615	3.315	3.933	3.31	0.384	0.273	0.520	31.27
$d_{s_t}^{*(4)}$	0.580	0.364	0.860	64.02	0.358	0.253	0.479	32.15
$d_{s_t}^{*(5)}$	1.706	1.561	1.861	2.97	5.190	4.608	5.809	3.87
$d_{s_t}^{*(6)}$	0.503	0.327	0.725	38.31	4.152	3.709	4.615	3.73
$d_{s_t}^{*(7)}$	1.461	1.339	1.589	2.86	0.676	0.376	1.110	55.44
$d_{s_t}^{*(8)}$	1.531	1.406	1.662	1.66	1.185	1.058	1.319	2.13
$d_{s_t}^{*(9)}$	1.330	1.221	1.449	1.72	2.672	2.397	2.970	2.17
$10\times L_{s_t}^{(1)}$	3.625	3.334	3.935	9.23	4.473	4.016	4.974	2.37
$10\times L_{s_t}^{(2)}$	3.917	3.601	4.266	12.75	5.011	4.510	5.573	6.04
$10\times L_{s_t}^{(3)}$	4.536	4.150	4.954	8.86	7.424	6.650	8.266	3.03
α_{s_t}	1.156	0.961	1.355	10.80	-0.835	-1.063	-0.610	12.15

Table 5: **Estimates of regime-independent parameters (Restricted model (ρ =0))** *This table presents the posterior mean, 95% credibility interval and inefficiency factor (ineff.) of the regime-independent parameters based on 50,000 MCMC draws beyond a burn-in of 5,000.*

influenced by the shocks to the long-term bond risk.

As one can see in both Tables 4 and 5, the estimated posterior means of the regime-specific parameters are markedly different across the regimes. Most of the diagonal elements in L_{s_t} and D_{s_t} have switched over time between high and low values, suggesting regime switching volatilities of the yield curve. Moreover, according to the estimates for α_{s_t}, the regime changes are asymmetric because these posterior means imply that the transition probability from regime 1 (regime 2) to regime 2 (regime 1) is 22.6% (14.5%).

The most notable feature that can be observed from the tables is that the first two elements of μ_{s_t}, denoted by $\mu_{s_t}^{(1)}$ and $\mu_{s_t}^{(2)}$, are substantially different between the two competing models whereas the third element of μ_{s_t} is almost the same. In particular, the difference in $\mu_{s_t}^{(2)}$ is 0.446, which is actually large in terms of the unconditional mean of the slope factor due to its high persistence. The correlation coefficients ρ_1



(a) μ_1 (b) μ_2 (c) ρ_2

(d) $10\times L_1^{(1)}$ (e) $10\times L_1^{(2)}$ (f) $d_1^{*(1)}$

Figure 4: **Marginal posterior plots for some selected parameters** *These graphs are based on 50,000 simulated draws of the posterior simulation.*

and ρ_2 are significantly negative and hence the posterior estimates for $\mu_{s_t}^{(1)}$ and $\mu_{s_t}^{(2)}$ from the restricted model are inaccurate. As a result, the endogenous switching model outperforms the exogenous model in terms of the out-of-sample forecasting. The 12 months of 2008 are forecasted using the data until 2007:M12, and the forecasts from the proposed model are found to be more accurate at all horizons than those from the exogenous regime switching model although we do not report the result in the interest of space.[6]

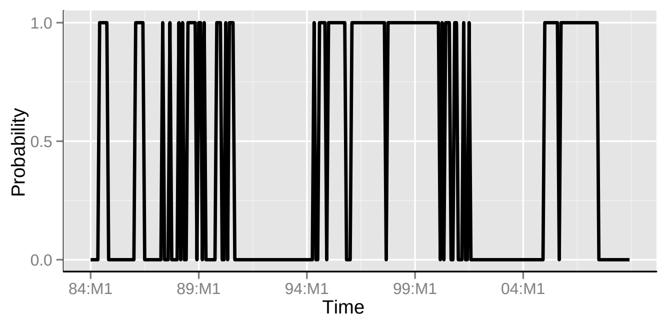
Table 6 confirms the endogeneity of the regime changes based on the marginal likelihood which reflects an automatic penalty for model complexity. As can be seen in this table, the endogenous regime switching model is most supported by the data. Finally, Figure 5 shows the persistence of the volatility regimes revealing that the high volatility

[6]We measure the predictive accuracy of the forecasts in terms of the posterior predictive criterion proposed by Gelfand and Ghosh (1998).

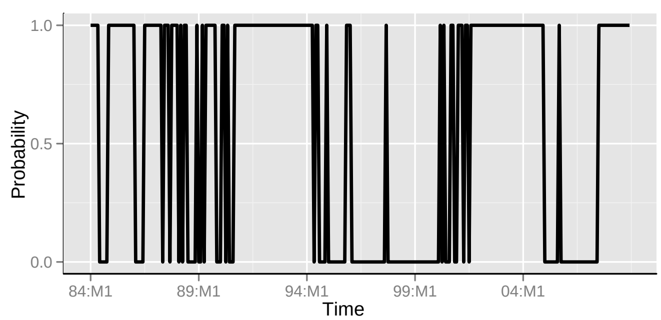
model	lnL (n.s.e.)	lnPost (n.s.e)	lnML (n.s.e.)
Endogenous RS model ($\rho \neq 0$)	4744.31 (1.14)	-75.88 (0.02)	4813.35 (1.18)
Exogenous RS model ($\rho = 0$)	4687.47 (1.11)	-66.27 (0.03)	4754.85 (1.08)
Non-Switching model	3859.47 (0.00)	-109.78 (0.02)	3969.25 (0.02)

Table 6: **Log likelihood (lnL), log posterior ordinate (lnPost), log marginal likelihood (lnML) and numerical standard error(n.s.e)**

regime is far less persistent than the low volatility regime.



(a) Low volatility regime ($s_t = 1$)

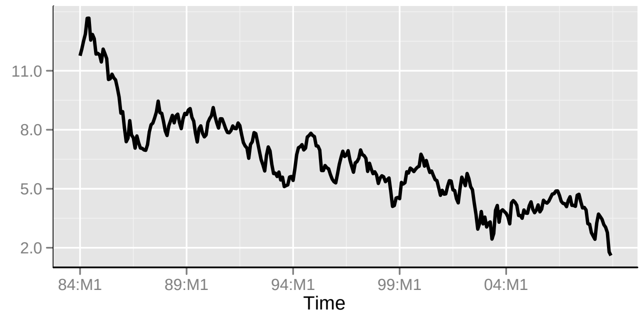


(b) High volatility regime ($s_t = 2$)

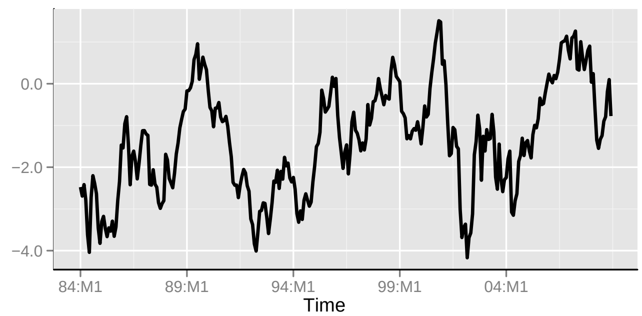
Figure 5: **Probabilities of Regimes** *These graphs plot the estimates of the probabilities of regimes. These graphs are based on 50,000 simulated draws of the posterior simulation.*

5 Conclusion

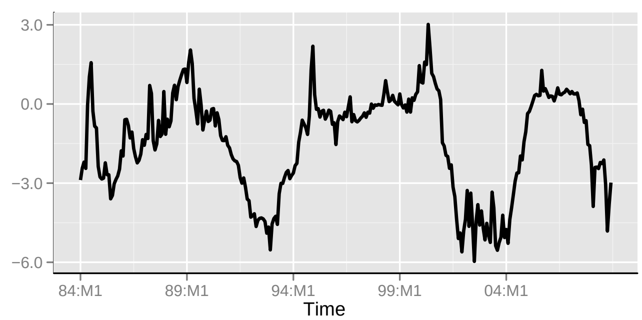
In this study the regime switching linear state space models is proposed, in which the regime switches are endogenously determined through the correlation with the observed or unobserved continuous state variables. This work is an extension of that by Kim



(a) Level



(b) Slope



(c) Curvature

Figure 6: **Factors** *These graphs plot the estimates of the factors. These graphs are based on 50,000 simulated draws of the posterior simulation.*

et al. (2008) to a general state space model that has been widely used in modeling in the area of macroeconomics and finance. This paper also provides an efficient Bayesian MCMC estimation method. The key idea is to simulate the latent state variable that controls the regime shifts. Thus one can estimate the models without approximation and inaccuracy. Furthermore, the validity of our method is also demonstrated by simulation study and application to a generalized Nelson-Siegel yield curve model with endogenous Markov switching volatility regime shifts.

References

Albert, J. H. and Chib, S. (1993), "Bayesian-analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 88, 669–679.

Ang, A., Bekaert, G., and Wei, M. (2008), "The term structure of real rates and expected inflation," *Journal of Finance*, 63, 797–849.

Arnold, B. C. and Beaver, R. J. (2002), "Skewed multivariate models related to hidden truncation and/or selective reporting," *Test*, 11(1), 7–54.

Bansal, R. and Zhou, H. (2002), "Term structure of interest rates with regime shifts," *Journal of Finance*, LVII, 463–473.

Carter, C. K. and Kohn, R. (1994), "On Gibbs sampling for state space models," *Biometrika*, 81, 541–553.

Chib, S. (1996), "Calculating posterior distributions and modal estimates in Markov mixture models," *Journal of Econometrics*, 75, 79–97.

Chib, S. and Dueker, M. (2004), "Non-Markovian Regime Switching with Endogenous States and Time-Varying State Strengths," Working Paper 2008-31, Federal Reserve Bank of St. Louis.

Chib, S. and Ergashev, B. (2009), "Analysis of multi-factor affine yield curve Models," *Journal of the American Statistical Association*, 104, 1324–1337.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings algorithm," *American Statistician*, 49, 327–335.

Chib, S. and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96, 270–281.

Chib, S. and Kang, K. H. (2010), "Change Points in Affine Term-Structure Models: Pricing, Estimation and Forecasting," *Manuscript*.

Chib, S., Nardari, F., and Shephard, N. (2002), "Markov chain Monte Carlo methods for stochastic volatility models," *Journal of Econometrics*, 108, PII S0304–4076(01)00137–3.

Chib, S. and Ramamurthy, S. (2010), "Tailored Randomized-block MCMC Methods with Application to DSGE Models," *Journal of Econometrics*, 155, 19–38.

Cochrane, J. H. and Piazzesi, M. (2008), "Decomposing the Yield Curve," *Manuscript*.

Diebold, F., Lee, J.-H., and Weinbach, G. C. (1994), "Regime Switching with Time-Varying Transition Probabilities," *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press, 283–302.

Diebold, F. X. and Li, C. L. (2006), "Forecasting the term structure of government bond yields," *Journal of Econometrics*, 130, 337–364.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006), "The macroeconomy and the yield curve: a dynamic latent factor approach," *Journal of Econometrics*, 131, 309–338.

Filardo, A. J. and Gordon, S. F. (1994), "Business-Cycle Phases and Their Transitional Dynamics," *Journal of Business and Economic Statistics*, 12(3), 299–308.

— (1998), "Business cycle durations," *Journal of Econometrics*, 85(1), 99–123.

Gelfand, A. E. and Ghosh, S. K. (1998), "Model choice: A minimum posterior predictive loss approach," *Biometrika*, 85, 1–11.

Gray, S. F. (1996), "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42(1), 27–62.

Gurkaynak, R. S., Sack, B., and Wright, J. H. (2007), "The U.S. treasury yield curve: 1961 to the present," *Journal of Monetary Economics*, 54, 2291–2304.

Hamilton, J. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–84.

Kaufmann, S. (2000), "Measuring business cycles with a dynamic Markov switching factor model: an assessment using Bayesian simulation methods," *The Econometrics Journal*, 3(1), 39–65.

Kim, C.-J. (1993), "Unobserved-Component Time Series Models with Markov-Switching Heteroscedasticity: Changes in Regime and the Link between Inflation Rates and Inflation Uncertainty," *Journal of Business and Economic Statistics*, 11(3), 341–49.

Kim, C. J. (1994), "Dynamic linear models with Markov-switching," *Journal of Econometrics*, 60, 1–22.

Kim, C. J. and Piger, J. (2002), "Dynamic linear models with Markov-switching," *Journal of Monetary Economics*, 49, 1189–1211.

Kim, C. J., Piger, J., and Startz, R. (2008), "Estimation of Markov regime-switching regression models with endogenous switching," *Journal of Econometrics*, 143, 263–273.

Kuan, C.-M., Huang, Y.-L., and Tsay, R. S. (2005), "An Unobserved-Component Model With Switching Permanent and Transitory Innovations," *Journal of Business and Economic Statistics*, 23, 443–454.

Morley, J., Nelson, C. R., and Zivot, E. (2003), "Why Are Unobserved Component and Beveridge-Nelson Trend-Cycle Decompositions of GDP So Different," *Review of Economics and Statistics*, 85, 235–243.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), "Stochastic volatility with leverage: fast and efficient likelihood inference," *Journal of Econometrics*, 140, 425–449.

Pitt, M. K. and Shephard, N. (1999), "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94(446), 590–599.