

State-Space Models with Endogenous Markov Regime Switching Parameters

KYU HO KANG*

(Department of Economics, Washington University in St. Louis, Box 1208, MO 63130, USA)

This Version: July 2010

Abstract

In this paper, Kim, Piger and Startz' (2008) endogenous Markov-switching model is extended to a general state-space model. This paper also complements Kim's (1991) regime switching dynamic linear models by allowing the discrete regime to be jointly determined with observed or unobserved continuous state variables. An efficient Bayesian MCMC estimation method is developed. It is shown that simulation of the latent state variable controlling the regime shifts enables us to precisely estimate the models without approximation. This method is applied to the estimation of a generalized Nelson-Siegel yield curve model where the unobserved time-varying curvature factor is allowed to be contemporaneously correlated with Markov switching volatility regimes. All techniques are also illustrated using simulated data sets. (JEL classification: C1; C4)

Keywords: Bayesian estimation, Markov switching process, Markov Chain Monte Carlo, Bayes factor, Term structure of interest rates

1 Introduction

State-space models with regime switching parameters are so flexible that they have been commonly used to model heterogenous dynamics of data over time (For instance, Kim (1994), Kim and Piger (2002) and Kuan, Huang, and Tsay (2005)). The flexibility of this modeling approach is due to the fact that it involves two different types of dynamic state variables. One is usually termed as a regime indicator and it takes on discrete values determining the set of state-contingent model parameters at each data point.

**Email:* khkang@wustl.edu

The other is generated from continuous probability distributions over time (for example, ARMA process) and responsible for the dynamics within the discrete state. Those state variables are both responsible for the dynamics of observed quantities. In modeling with such dynamic systems, it is essential to consider contemporaneous feedback among the state variables in many empirical works. The instantaneous interaction between continuous state variables can be modeled by allowing for non-zero correlation of the innovations. (For instance, Morley, Nelson, and Zivot (2003)). Also by estimating the full transition matrix one can consider the co-movement of discrete state variables (i.e. regime indicators).

Although this modeling strategy is convenient and useful in many applications, it is rather restrictive because it only permits instantaneous interaction among the state variables with the same type. A more general approach is to allow for the bi-directional feedback between the two different types of the underlying state variables. In particular, this general approach, which has not been addressed yet, may be important in many macroeconomics or finance models. For example, in Nelson-Siegel yield curve models, the slope and curvature of the term structure, which are modeled as unobserved factors, affect the future risk premia (Cochrane and Piazzesi (2008)). Also they are highly likely to be contemporaneously correlated with Markov switching volatility regimes since the size of shock volatilities determines the magnitude of the risk. It thus seems more sensible to model the joint realization of those two different types of unobserved underlying stochastic processes.

The purpose of this paper is to develop and estimate state space models in which the unobserved discrete regimes and continuous state variables are jointly determined. The regime switches are modeled through a Markov process. In this model formulation, it becomes possible to make the realization of the regimes endogenized through the correlation with the observed or unobserved continuous state variables. Our statistical inference is fully Bayesian. The MCMC method developed by this paper builds on the works of Albert and Chib (1993) and Kim, Piger, and Startz (2008). Kim et al. (2008) propose endogenous Markov-switching linear models which do not involve the dynamics of continuous state variables, and our work extends their work to a general state-space

model. We show that it can be done by introducing and simulating the latent state variables that control regime shifts at each time point. This idea of data augmentation is based on the work of Albert and Chib (1993), in which exact Bayesian methods for modeling categorical response data are developed.¹ In addition, we conduct a Bayesian model choice between endogenous switching and exogenous switching model based on the marginal likelihoods.

In the next section, we lay out a two-regime Markov-switching dynamic factor model with endogenous switching. Section 3 discusses our Bayesian MCMC estimation method. Section 4 gives the results of simulation study and an empirical work. Section 5 has the conclusion.

2 The Class of Models

We study a class of state-space models where at each time point the model parameters are chosen by a discrete state Markov process indexed by s_t , and some of the unobserved factors and the state variable are contemporaneously correlated. The models we discuss are represented by a state space form,

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

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

where \mathbf{y}_t is a q -dimensional vector of dependent variables with finite first and second moments 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 exogenous or predetermined variables, and it is assumed to be covariance-stationary. \mathbf{a}_{s_t} , \mathbf{b}_{s_t} and \mathbf{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 mean and persistence of the factors. $D_{s_t} : q \times q$ and $L : k \times k$ capture the volatilities of the measurement errors and the continuous state shocks. In addition, e_t and ε_t are assumed to be mutually independent, and $\mathbb{E}[e_t|\mathbf{S}_n] = \mathbf{0}$ and $\mathbb{E}[e_t e_t'|\mathbf{S}_n] = \mathbf{I}_q$.

¹Chib and Dueker (2004), as one of related works in the Bayesian approach, develop a non-Markovian regime switching model. In their setup, the regime states depend on the sign of an autoregressive latent variable, which is allowed to be endogenous in sense that regimes are determined jointly with the observed data.

In what follows, we postulate that the number of regimes is two as in Hamilton (1989). The two-regime case is not only convenient to explain and understand, but also popular in many empirical studies. We also assume that the unobserved discrete state variable s_t is governed by a first-order Markov chain with transition probabilities:

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

where \mathbf{z}_t is a vector of covariance-stationary exogenous or predetermined variables, which may include some elements of \mathbf{x}_t . (i.e. the regime-switching point positions are modeled as a Markov process.) 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 . For the purpose a convenient formulation of the Markov process, we introduce another latent variable γ_t , so that the influence of \mathbf{z}_t on the transition probabilities is modeled through a probit specification as in Kim et al. (2008).

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

In other words. the transition probabilities have the form

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

By allowing for non-zero correlation between the regime shock γ_t and the factor shock ε_t we model a bi-directional contemporaneous feedback between the unobserved state variable s_t and the unobserved factors \mathbf{f}_t

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

where $\rho = (\rho_1 \ \rho_2 \ \cdots \ \rho_k)'$. Hence, the presence of some none-zero $\rho'_i s$ ($i = 1, 2, \dots, k$) implies endogenous regime changes, and the realization of regime at the next period is determined jointly with the vector of continuous latent (or observable) variables. This feature distinguishes our work from the existing studies. For example, when all $\rho'_i s$ are zero, we have the Markov switching dynamic factor model with time-varying transition probability of Kim and Nelson (1998). Also it is very useful to notice that

when $k = q = \mathbf{b}_{s_t} = 1$ and $G_{s_t} = \mu_{s_t} = D_{s_t} = 0$, this class of models reduces to that of Kim et al. (2008) in which \mathbf{y}_t is scalar and thereby no unobserved continuous state variable is involved.

Many other interesting models can also be constructed as special cases such as time-varying coefficient model, dynamic common factor model and unobserved component model and so forth. Specifically, one may set \mathbf{y}_t as a vector of stationary asset returns and \mathbf{f}_t as the dynamic common factors where the factor loadings are regime-specific. So \mathbf{y}_t may endogenously switch between strong and weak co-movement among the observed returns over time. It is also possible to specify and evaluate the endogenous asymmetry in the business cycle using the Friedman's plucking model context.²

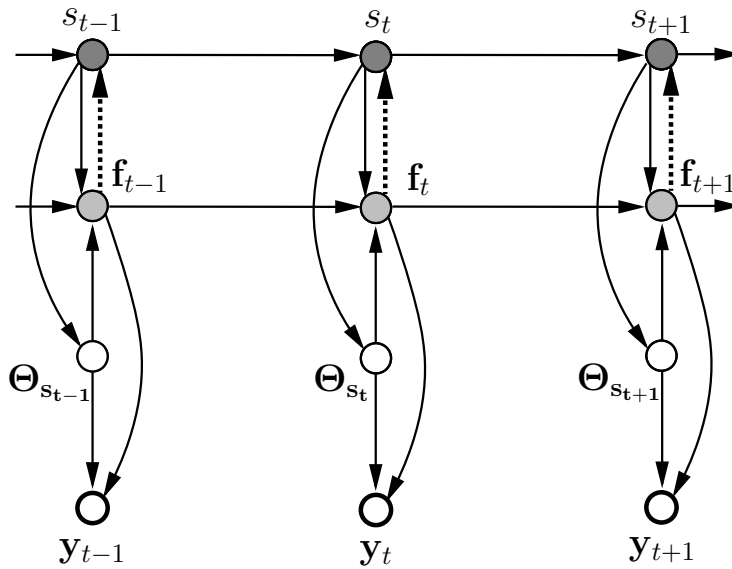


Figure 1: **Directed graph of model linkages.** This is a summary of the data generating process. In the beginning of period t , a regime and a vector of factors occur simultaneously conditioned on f_{t-1} and s_{t-1} . Then given the regime s_t , the corresponding model parameters Θ_{s_t} are taken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

Figure 1 summarizes the way the observations are generated in terms of a directed acyclic graph. In the beginning of period t , a regime and a vector of factors occur

²Sinclair (2010) estimates this model using the quasi-maximum likelihood estimation method, and finds that the asymmetry in the business cycle of the U.S. economy is exogenous.

simultaneously. This realization of the regime at time t is governed with the regime in the previous period and the current factors as indicated by the direction of the two arrows connecting s_{t-1} to s_t and \mathbf{f}_t to s_t . Then given the regime at time t , the corresponding model parameters Θ_{s_t} are taken from the full collection of model parameters. These include \mathbf{a}_{s_t} and \mathbf{b}_{s_t} , for example. Conditioned on the parameters and \mathbf{f}_{t-1} , \mathbf{f}_t is generated by the regime-specific autoregressive process in (2.2). Finally, from (2.1), \mathbf{a}_{s_t} , \mathbf{b}_{s_t} , \mathbf{f}_t and a simulated measurement error e_t at each time point construct the observations \mathbf{y}_t . Notice that in standard state space models with regime switching parameters the dashed line in figure 1 is absent since s_t is assumed to be drawn independently of \mathbf{f}_t .

3 Prior-Posterior Analysis

3.1 Markov Chain Monte Carlo Scheme

Let $\mathbf{Y}_t = \{\mathbf{y}_i\}_{i=1,2,\dots,t}$, $\mathbf{X}_t = \{\mathbf{x}_i\}_{i=1,2,\dots,t}$ and $\mathbf{Z}_t = \{\mathbf{z}_i\}_{i=1,2,\dots,t}$ be observations observed through time t . Similarly, $\mathbf{F}_t = \{\mathbf{f}_i\}_{i=1,2,\dots,t}$, $\mathbf{S}_t = \{s_i\}_{i=1,2,\dots,t}$ and $\gamma_t = \{\gamma_i\}_{i=1,2,\dots,t}$ denote collection of the state variables through time t . Let Θ denote the parameters in the evolution and transition equations and \mathbf{P} is those in the transition probabilities. That is, Θ is the collection of the parameters in $\{\mathbf{a}_{s_t}, \mathbf{b}_{s_t}, \mathbf{H}_{s_t}, D_{s_t}, \mu_{s_t}, G_{s_t}, L_{s_t}\}$ and $\mathbf{P} = \{\alpha_1, \alpha_2, \beta'_1, \beta'_2, \rho_1, \rho_2, \dots, \rho_k\}$. Suppose that we have specified a prior density $\pi(\Theta, \mathbf{P})$ on the parameters and data \mathbf{Y}_n , \mathbf{Z}_n and \mathbf{X}_n are available. A Bayesian state space model with regime switching parameters is defined by a joint distribution over the regime states, continuous latent variables, model parameters and the data. In this context, interest centers on the posterior density $\pi(\Theta, \mathbf{P}, \mathbf{S}_n, \gamma_n, \mathbf{F}_n | \mathbf{Y}_n, \Omega_n)$, and

$$\begin{aligned} & \pi(\Theta, \mathbf{P}, \mathbf{S}_n, \gamma_n, \mathbf{F}_n | \mathbf{Y}_n, \Omega_n) \\ & \propto f(\mathbf{Y}_n | \Omega_n, \mathbf{S}_n, \gamma_n, \mathbf{F}_n, \Theta, \mathbf{P}) \times p(\mathbf{S}_n, \gamma_n, \mathbf{F}_n | \Omega_n, \Theta, \mathbf{P}) \times \pi(\Theta, \mathbf{P}) \end{aligned} \quad (3.1)$$

where $\Omega_t = \mathbf{X}_t \cup \mathbf{Z}_t$ is the collection of exogenous variables at time t whose dynamics are not analyzed through the transition equation in equation (2.2). Note that we use the notation π to denote prior and posterior density functions of (Θ, \mathbf{P}) . We apply our MCMC sampling scheme to the posterior density $\pi(\Theta, \mathbf{P}, \mathbf{S}_n, \gamma_n, \mathbf{F}_n | \mathbf{Y}_n, \Omega_n)$ and obtain

the posterior distribution $\pi(\Theta, \mathbf{P} | \mathbf{Y}_n, \Omega_n)$ by integrating out $(\mathbf{S}_n, \gamma_n, \mathbf{F}_n)$ in a numerical way.

We sample the parameters and the states recursively. In the first step, the parameters in Θ are simulated on $(\mathbf{S}_n, \Omega_n, \mathbf{W}_n, \mathbf{P}, \gamma_n)$ and then \mathbf{P} is sampled in turn. Next, the states \mathbf{S}_n are drawn conditioned on (\mathbf{W}_n, Ω_n) and the other parameters where $\mathbf{W}_t = \mathbf{Y}_t \cup \mathbf{F}_t$. Then we sequentially simulate γ_n and \mathbf{F}_n conditioned on the most recent values of the conditioning variables. Our MCMC algorithm can be summarized as follows.

Algorithm: MCMC sampling

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

Step 2 Sample $\Theta | \mathbf{W}_n, \Omega_n, \mathbf{S}_n, \gamma_n, \mathbf{P}$

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

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

Step 5 Sample $\gamma_n | \mathbf{W}_n, \Omega_n, \mathbf{S}_n, \Theta, \mathbf{P}$

Step 6 Sample $\mathbf{F}_n | \mathbf{Y}_n, \Omega_n, \mathbf{S}_n, \gamma_n, \Theta, \mathbf{P}$

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

Full details of each of these steps are given by the following.

3.1.1 Simulation of Θ

We consider the question of simulating Θ conditioned on $(\Omega_n, \mathbf{W}_n, \mathbf{S}_n, \gamma_n, \mathbf{P})$ by the tailored multiple block MH algorithm(Chib and Greenberg (1995)). In this method the parameters in Θ 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 j th iteration, we have g sub-blocks of Θ

$$\Theta_1, \Theta_2, \dots, \Theta_g$$

Then the proposal density $q(\Theta_i | \Theta_{-i}, \mathbf{W}_n, \gamma_n, \mathbf{S}_n, \mathbf{P})$ for the i th block, conditioned on the most current value of the remaining blocks Θ_{-i} , is constructed by a quadratic approximation at the mode of the current target density $\pi(\Theta_i | \Theta_{-i}, \mathbf{W}_n, \gamma_n, \mathbf{S}_n, \mathbf{P})$. In our case, we let this proposal density take the form of a student t distribution with 15 degrees of freedom

$$q(\Theta_i | \Theta_{-i}, \mathbf{P}, \mathbf{W}_n, \gamma_n, \Omega_n, \mathbf{S}_n) = St(\Theta_i | \Theta_i, \mathbf{P}, \mathbf{V}_{\Theta_i}, 15) \quad (3.2)$$

where

$$\begin{aligned} \hat{\Theta}_i &= \arg \max_{\Theta_i} \ln \{ f(\mathbf{W}_n | \gamma_n, \Omega_n, \mathbf{S}_n, \Theta_i, \Theta_{-i}, \mathbf{P}) \pi(\Theta_i) \} \\ \text{and } \mathbf{V}_{\Theta_i} &= \left(- \frac{\partial^2 \ln \{ f(\mathbf{W}_n | \gamma_n, \Omega_n, \mathbf{S}_n, \Theta_i, \Theta_{-i}, \mathbf{P}) \pi(\Theta_i) \}}{\partial \Theta_i \partial \Theta_i'} \right)_{|\Theta_i = \hat{\Theta}_i}^{-1}. \end{aligned} \quad (3.3)$$

We then generate a proposal value Θ_i^\dagger . If Θ_i^\dagger violates any of the constraints in \mathcal{R} , it is immediately rejected. Otherwise, it is accepted as the next value in the chain with probability

$$\begin{aligned} &\alpha(\Theta_i^{(j-1)}, \Theta_i^\dagger | \Theta_{-i}) \\ &= \min \left\{ \frac{f(\mathbf{W}_n | \gamma_n, \Omega_n, \mathbf{S}_n, \Theta_i^\dagger, \Theta_{-i}, \mathbf{P}) \pi(\Theta_i^\dagger)}{f(\mathbf{W}_n | \gamma_n, \Omega_n, \mathbf{S}_n, \Theta_i^{(j-1)}, \Theta_{-i}, \mathbf{P}) \pi(\Theta_i^{(j-1)})}, \frac{St(\Theta_i^{(j-1)} | \Theta_i, \mathbf{P}, \mathbf{V}_{\Theta_i}, 15)}{St(\Theta_i^\dagger | \Theta_i, \mathbf{P}, \mathbf{V}_{\Theta_i}, 15)}, 1 \right\}. \end{aligned} \quad (3.4)$$

The simulation of Θ is complete when all the sub-blocks $\{\Theta_i\}_{i=1,2,\dots,g}$ are sequentially updated as above. On letting $\mathcal{N}_q(x|a, b)$ denote the q -dimensional multivariate normal density of x with mean of a and variance of b , the required joint density of \mathbf{W}_n conditioned on $(\Omega_n, \gamma_n, \mathbf{S}_n, \Theta, \mathbf{P})$ is then a product of conditional predictive densities:

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

where

$$\mathbf{y}_{t|t-1} = \mathbb{E}[\mathbf{y}_t | \mathbf{f}_t, \boldsymbol{\Omega}_n, \mathbf{W}_{t-1}, \mathbf{S}_n, \gamma_n] = \mathbf{a}_{s_t} + \mathbf{b}_{s_t} \mathbf{f}_t + \mathbf{H}_{s_t} \mathbf{x}_t \quad (3.6)$$

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

Since ε_t and γ_t are jointly normally distributed, we have $\mathbb{E}[\varepsilon_t | \gamma_t] = \rho \gamma_t$ and $\mathbb{V}[\varepsilon_t | \gamma_t] = \mathbf{I}_k - \rho \rho'$. It is extremely important to notice 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 addition, without γ_n the density of \mathbf{y}_t conditioned on s_t and s_{t-1} is no longer Gaussian as Kim et al. (2008) pointed out.

When the likelihood function tends to be ill-behaved in these problems because of nonlinearity, one may calculate Θ_i using a suitably designed version of the simulated annealing algorithm like in Chib and Ergashev (2009). Especially when the multimodality of the likelihood surface is severe, this stochastic optimization method is more reliable than the standard Newton-Raphson class of deterministic optimizers. Moreover, if there is no natural way of grouping parameters in the way that parameters in different blocks are not strongly correlated, randomizing blocking scheme introduced by Chib and Ramamurthy (2010) works better than fixed blocking schemes in terms of efficacy and convergence.

3.1.2 Simulation of \mathbf{P}

We now discuss sampling \mathbf{P} . In our formulation its full conditional $p(\mathbf{P} | \mathbf{S}_n, \boldsymbol{\Omega}_n, \mathbf{F}_n, \Theta)$ 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 factors, it depends on \mathbf{F}_n . Also since 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 Θ , 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 the its prior

$$p(\mathbf{P} | \mathbf{S}_n, \mathbf{F}_n, \boldsymbol{\Omega}_n, \Theta) \propto p(\mathbf{S}_n, \mathbf{F}_n | \boldsymbol{\Omega}_n, \Theta, \mathbf{P}) \pi(\mathbf{P}) \quad (3.8)$$

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

$$p(\mathbf{S}_n, \mathbf{F}_n | \boldsymbol{\Omega}_n, \boldsymbol{\Theta}, \mathbf{P}) = p(s_0) f(\mathbf{f}_0) \prod_{t=1}^n p[s_t, \mathbf{f}_t | s_{t-1}, \mathbf{f}_{t-1}, \mathbf{z}_t, \boldsymbol{\Theta}] \quad (3.9)$$

$$\propto \prod_{t=1}^n p[s_t, \mathbf{f}_t | s_{t-1}, \mathbf{f}_{t-1}, \mathbf{z}_t, \boldsymbol{\Theta}] \quad (3.10)$$

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.

$$\begin{aligned} p(s_t = 1, \mathbf{f}_t | s_{t-1} = j, \mathbf{f}_{t-1}, \mathbf{z}_t, \boldsymbol{\Theta}) &= p(s_t = 1, \varepsilon_t | s_{t-1} = j, \mathbf{z}_t, \boldsymbol{\Theta}) \quad (3.11) \\ &= p(s_t = 1 | \varepsilon_t, s_{t-1} = j, \mathbf{z}_t, \boldsymbol{\Theta}) f[\varepsilon_t | s_{t-1} = j, \mathbf{z}_t] \\ &= \Pr[\gamma_t < \alpha_j + \beta'_j \mathbf{z}_t | \varepsilon_t, s_{t-1} = j, \mathbf{z}_t] \phi_k(\varepsilon_t) \text{ since } f[\varepsilon_t | s_{t-1} = j, \mathbf{z}_t] = \Pr[\varepsilon_t] \\ &\propto \Pr\left[\rho' \varepsilon_t + \sqrt{1 - \rho' \rho} \omega_t < \alpha_j + \beta'_j \mathbf{z}_t | \varepsilon_t, \mathbf{z}_t\right] \text{ with } \omega_t \sim \text{i.i.d. } \mathcal{N}(0, 1) \\ &= \Pr\left[\omega_t < \frac{\alpha_j + \beta'_j \mathbf{z}_t - \rho' \varepsilon_t}{\sqrt{1 - \rho' \rho}} | \varepsilon_t, \mathbf{z}_t\right] \\ &= \Phi\left(\frac{\alpha_j + \beta'_j \mathbf{z}_t - \rho' \varepsilon_t}{\sqrt{1 - \rho' \rho}}\right) \end{aligned}$$

where $\varepsilon_t = L_{s_t}^{-1}(\mathbf{f}_t - \mu_{s_t} - G_{s_t} \mathbf{f}_{t-1})$ and $\phi_k(\cdot)$ is a multivariate standard normal density and Φ is the c.d.f of the standard normal distribution.

In exactly the same way,

$$\begin{aligned} p[s_t = 2, \varepsilon_t | s_{t-1} = j, \boldsymbol{\Omega}_t] &\quad (3.12) \\ \propto \left[1 - \Phi\left(\frac{\alpha_j + \beta'_j \mathbf{z}_t - \rho' \varepsilon_t}{\sqrt{1 - \rho' \rho}}\right)\right] &= \Phi\left(\frac{-\alpha_j - \beta'_j \mathbf{z}_t + \rho' \varepsilon_t}{\sqrt{1 - \rho' \rho}}\right) \end{aligned}$$

Now $p(\mathbf{S}_n, \mathbf{F}_n | \boldsymbol{\Theta}, \mathbf{P})$ yields the likelihood density for \mathbf{P} conditioned on $(\mathbf{S}_n, \boldsymbol{\Omega}_n, \mathbf{F}_n, \boldsymbol{\Theta})$ without the normalizing constant, which enables us to sample \mathbf{P} by the MH method.

One can see that when all elements in ρ and β_j are zeros for $j = 1, 2$ (i.e. $p_{ij}(\mathbf{z}_t) = p_{ij}$), 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 regard to the sampling model. Therefore, an exogenous Markov regime switching model with constant transition probabilities is a special cases of our formulation.

3.1.3 Simulation of $\{s_t\}$

In order to sample \mathbf{S}_n we provide a generalized version of the method of Chib (1996), where exogenous Markov mixture models are analyzed by MCMC methods. The objective in this subsection is to draw a sequence of values of $\{s_t\}_{t=0,1,2,\dots,n}$ 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 $p(s_t|\mathbf{W}_n, \boldsymbol{\Omega}_n, s_{t+1}, \boldsymbol{\Theta}, \mathbf{P})$, and by Bayes theorem

$$p(s_t|\mathbf{W}_n, \boldsymbol{\Omega}_n, s_{t+1}, \boldsymbol{\Theta}, \mathbf{P}) \propto p(s_t|\mathbf{W}_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})p(s_{t+1}|s_t, \boldsymbol{\Omega}_{t+1}, \mathbf{f}_{t+1}, \boldsymbol{\Theta}, \mathbf{P}) \quad (3.13)$$

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 $p(s_t|\mathbf{W}_n, \boldsymbol{\Omega}_n, \boldsymbol{\Theta}, \mathbf{P})$ by calculating

$$p(s_t|\mathbf{W}_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \quad (3.14)$$

$$= \frac{\sum_{j=1}^2 f(\mathbf{y}_t, \mathbf{f}_t|\mathbf{W}_{t-1}, s_{t-1} = j, s_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})p(s_{t-1} = j, s_t|\boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})}{f(\mathbf{y}_t, \mathbf{f}_t|\mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})} \quad (3.15)$$

Then the each term in the numerator can be written as

$$f(\mathbf{y}_t, \mathbf{f}_t|s_{t-1}, s_t, \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})p(s_{t-1}, s_t|\mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \quad (3.16)$$

$$= f(\mathbf{y}_t|\mathbf{f}_t, s_t, s_{t-1}, \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})f(\mathbf{f}_t|s_t, s_{t-1} = j, \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \quad (3.17)$$

$$\times p(s_t|s_{t-1} = j, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})p(s_{t-1}|\mathbf{W}_{t-1}, \boldsymbol{\Omega}_{t-1}, \boldsymbol{\Theta}, \mathbf{P})$$

The first term of equation (3.17) is given by

$$f(\mathbf{y}_t|\mathbf{f}_t, \mathbf{W}_{t-1}, s_{t-1}, s_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) = \mathcal{N}_q(\mathbf{y}_t|\mathbf{y}_{t|t-1}, D_{st}D'_{st}) \quad (3.18)$$

Assuming $p(s_{t-1}|\mathbf{W}_{t-1}, \boldsymbol{\Omega}_{t-1}, \boldsymbol{\Theta}, \mathbf{P})$ given, Kim et al. (2008) provide the remaining terms for non-zero values of ρ as follows.

$$f(\mathbf{f}_t|s_t, s_{t-1} = j, \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})p(s_t|s_{t-1} = j, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \quad (3.19)$$

$$= \frac{\phi_k(L_{st}^{-1}(\mathbf{f}_t - \mu_{st} - G_{st}\mathbf{f}_{t-1}))}{|L_{st}|} \Phi\left(\frac{\alpha_j + \beta'_j \mathbf{z}_t - \rho'(L_{st}^{-1}(\mathbf{f}_t - \mu_{st} - G_{st}\mathbf{f}_{t-1}))}{\sqrt{1 - \rho'\rho}}\right)$$

These calculations are initialized at $t = 1$ by treating the initial probability $p(s_0|\mathbf{W}_0, \boldsymbol{\Omega}_0, \boldsymbol{\Theta}, \mathbf{P})$ as an additional parameter to be estimated or approximating it by the unconditional probability $p(s_0|\boldsymbol{\Theta}, \mathbf{P})$. With the numerator at hand, the conditional joint density of \mathbf{y}_t and \mathbf{f}_t which is the denominator of equation (3.14) is given by the law of total probability

$$\begin{aligned} & f(\mathbf{y}_t, \mathbf{f}_t | \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \\ &= \sum_{s_t=1}^2 \sum_{s_{t-1}=1}^2 f(\mathbf{y}_t, \mathbf{f}_t | s_{t-1}, s_t, \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) p(s_{t-1}, s_t | \mathbf{W}_{t-1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \end{aligned} \quad (3.20)$$

According to the scheme described in equation (3.13), the backward recursion can be done by the method of composition. First, one simulate s_n from $s_n | \mathbf{W}_n, \boldsymbol{\Omega}_n, \boldsymbol{\Theta}, \mathbf{P}$. Then given s_{t+1}, s_t ($t = 1, 2, \dots, t-1$) can be sampled based on its posterior mass function

$$\begin{aligned} & p(s_t = j | s_{t+1}, \mathbf{W}_{t+1}, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \\ &= \frac{p(s_{t+1} | s_t = j, \mathbf{W}_{t+1}, \boldsymbol{\Omega}_{t+1}, \boldsymbol{\Theta}, \mathbf{P}) p(s_t = j | \mathbf{W}_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})}{\sum_{j=1}^2 p(s_{t+1} | s_t = j, \mathbf{W}_{t+1}, \boldsymbol{\Omega}_{t+1}, \boldsymbol{\Theta}, \mathbf{P}) p(s_t = j | \mathbf{W}_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})} \end{aligned} \quad (3.21)$$

where

$$\begin{aligned} & p(s_{t+1} | s_t = j, \mathbf{W}_{t+1}, \boldsymbol{\Omega}_{t+1}, \boldsymbol{\Theta}, \mathbf{P}) \\ &= \Phi \left(\frac{\alpha_j + \beta'_j \mathbf{z}_{t+1} - \rho' (L_{s_{t+1}}^{-1} (\mathbf{f}_{t+1} - \mu_{s_{t+1}} - G_{s_{t+1}} \mathbf{f}_t))}{\sqrt{1 - \rho' \rho}} \right) \\ &\neq \Phi (\alpha_j + \beta'_j \mathbf{z}_{t+1}) \end{aligned} \quad (3.22)$$

This is equivalent to updating s_t by combining $p(s_t | \mathbf{W}_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P})$ and information contained in the generated s_{t+1} and the given \mathbf{f}_{t+1} . The Monte Carlo estimate of $p(s_t | \mathbf{W}_n, \boldsymbol{\Omega}_n)$ can be obtained by take an average of s_t over the MCMC iterations. It is interesting to see that when ρ is a vector of zeros, this sampling procedure reduces to the method of Chib (1996).

3.1.4 Simulation of $\{\gamma_t\}$

One of the most important features in our algorithm is to introduce an auxiliary variable γ_t into the sampling like in Albert and Chib (1993). Suppose that $s_t = 1$ conditioned on $s_{t-1} = j$ and $\varepsilon_t = L_{s_t}^{-1} (\mathbf{f}_t - \mu_{s_t} - G_{s_t} \mathbf{f}_{t-1})$, then from Bayes theorem,

$$f(\gamma_t | s_t = 1, s_{t-1} = j, \mathbf{W}_t, \boldsymbol{\Omega}_t, \boldsymbol{\Theta}, \mathbf{P}) \propto \mathcal{N}(\gamma_t | \rho' \varepsilon_t, 1 - \rho' \rho) \times I[\gamma_t < \gamma_t^*] \quad (3.23)$$

where $\gamma_t^* = \alpha_j + \beta_j' \mathbf{z}_t$ and $I[\cdot]$ is an indicator function. The information $s_t = 1$ simply serves to truncate the support of γ_t , which depends on ε_t due to the correlation. By a similar agreement, the support of γ_t is (γ_t^*, ∞) conditioned on the event $s_t = 2$ and $s_{t-1} = j$. Each of these truncated normal distributions is simulated as follows.

$$\begin{aligned}\gamma_t | \varepsilon_t, s_t = 1, s_{t-1} = j &\sim \mathcal{TN}_{(-\infty, \gamma_t^*]}(\rho' \varepsilon_t, \mathbf{1} - \rho' \rho) \\ \gamma_t | \varepsilon_t, s_t = 2, s_{t-1} = j &\sim \mathcal{TN}_{(\gamma_t^*, \infty)}(\rho' \varepsilon_t, \mathbf{1} - \rho' \rho)\end{aligned}\quad (3.24)$$

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

We complete our MCMC sampling scheme by sampling \mathbf{F}_n conditioned on $(\mathbf{Y}_n, \mathbf{S}_n, \boldsymbol{\Omega}_n, \boldsymbol{\gamma}_n, \boldsymbol{\Theta}, \mathbf{P})$ where this sampling stage is not necessary when the continues state variables are observable. For this, we modify and use the multi-move Gibbs-sampling suggested by Carter and Kohn (1994) that generates the whole time series of the unobserved factors 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 the factors \mathbf{f}_t based on the information up to time t as follows.

$$\mathbf{f}_{t|t} = \mathbb{E}[\mathbf{f}_t | \mathbf{Y}_t, \mathbf{S}_n, \boldsymbol{\Omega}_n, \boldsymbol{\gamma}_n, \boldsymbol{\Theta}, \mathbf{P}] = \mathbf{f}_{t|t-1} + K_t \eta_{t|t-1} \quad (3.25)$$

$$\mathbf{P}_{t|t} = \mathbb{V}[\mathbf{f}_t | \mathbf{Y}_t, \mathbf{S}_n, \boldsymbol{\Omega}_n, \boldsymbol{\gamma}_n, \boldsymbol{\Theta}, \mathbf{P}] = (\mathbf{I}_k - K_t \bar{\mathbf{b}}_{s_t}) \mathbf{P}_{t|t-1} \quad (3.26)$$

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

$$\mathbf{f}_{t|t-1} = \mu_{s_t} + G_{s_t} \mathbf{f}_{t-1|t-1} + L_{s_t} \rho \gamma_t \quad (3.27)$$

$$\mathbf{P}_{t|t-1} = G_{s_t} \mathbf{P}_{t-1|t-1} G_{s_t}' + L_{s_t} (\mathbf{I}_k - \rho \rho') L_{s_t}' \quad (3.28)$$

$$\eta_{t|t-1} = \mathbf{y}_t - \mathbf{a}_{s_t} - \mathbf{b}_{s_t} \mathbf{f}_{t|t-1} - \mathbf{H}_{s_t} \mathbf{x}_t \quad (3.29)$$

$$f_{t|t-1} = \mathbf{b}_{s_t} \mathbf{P}_{t|t-1} \mathbf{b}_{s_t}' + D_{s_t} D_{s_t}' \quad (3.30)$$

$$K_t = \mathbf{P}_{t|t-1} \mathbf{b}_{s_t}' f_{t|t-1}^{-1} \quad (3.31)$$

One distinguishing feature from the basic filter is that the last term of equation (3.27) and (3.28) exist due to the non-zero correlation ρ . We initialize those recursions 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 (3.25) and (3.26), the backward updating can be done by the standard sampling procedure.

$$\mathbf{f}_t | \mathbf{Y}_t, \mathbf{S}_n, \mathbf{f}_{t+1}, \boldsymbol{\Omega}_n, \boldsymbol{\gamma}_n, \boldsymbol{\Theta}, \mathbf{P} \sim \mathcal{N}_k(\mathbf{f}_{t|t, \mathbf{f}_{t+1}}, \mathbf{P}_{t|t, \mathbf{f}_{t+1}}) \quad (3.32)$$

where

$$\mathbf{f}_{t|t, \mathbf{f}_{t+1}} = \mathbf{f}_{t|t} + \mathbf{P}_{t|t} G'_{s_{t+1}} (\mathbf{P}_{t+1|t}^*)^{-1} (\mathbf{f}_{t+1} - \mathbf{f}_{t+1|t}^*) \quad (3.33)$$

$$\mathbf{P}_{t|t, \mathbf{f}_{t+1}} = \mathbf{P}_{t|t} - \mathbf{P}_{t|t} G'_{s_{t+1}} (\mathbf{P}_{t+1|t}^*)^{-1} G_{s_{t+1}} \mathbf{P}_{t|t} \quad (3.34)$$

$$\mathbf{f}_{t+1|t}^* = \mu_{s_{t+1}} + G_{s_{t+1}} \mathbf{f}_{t|t} + L_{s_t} \rho \gamma_{t+1} \quad (3.35)$$

$$\mathbf{P}_{t+1|t}^* = G_{s_{t+1}} \mathbf{P}_{t|t} G'_{s_{t+1}} + L_{s_{t+1}} (\mathbf{I}_k - \rho \rho') L'_{s_{t+1}} \quad (3.36)$$

3.2 Marginal Likelihood Calculation

One of our goals 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 M in our notation since all our MCMC computations must be repeated for all competing models. The marginal likelihood of any given model $m(\mathbf{Y}_n | \boldsymbol{\Omega}_n)$ is obtained as

$$m(\mathbf{Y}_n | \boldsymbol{\Omega}_n) = \int p(\mathbf{Y}_n | \boldsymbol{\Omega}_n, \boldsymbol{\Theta}, \mathbf{P}) \pi(\boldsymbol{\Theta}, \mathbf{P}) d(\boldsymbol{\Theta}, \mathbf{P}) \quad (3.37)$$

Because 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 we rely on a simulation-based approach. As is well known, provided we have an estimate of posterior ordinate $\pi(\boldsymbol{\Theta}^*, \mathbf{P}^* | \mathbf{Y}_n)$ the marginal likelihood can be estimated on the log scale as

$$\ln \hat{m}(\mathbf{Y}_n | \boldsymbol{\Omega}_n) = \ln f(\mathbf{Y}_n | \boldsymbol{\Omega}_n, \boldsymbol{\Theta}^*, \mathbf{P}^*) + \ln \pi(\boldsymbol{\Theta}^*, \mathbf{P}^*) - \ln \hat{\pi}(\boldsymbol{\Theta}^*, \mathbf{P}^* | \mathbf{Y}_n) \quad (3.38)$$

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

Notice that the first term in this expression is the likelihood evaluated at a single point.

$$f(\mathbf{Y}_n | \boldsymbol{\Omega}_n, \boldsymbol{\Theta}^*, \mathbf{P}^*) \quad (3.39)$$

$$= \int f(\mathbf{Y}_n, \mathbf{F}_n, \mathbf{S}_n, \gamma_n | \Omega_n, \Theta^*, \mathbf{P}^*) p(\mathbf{F}_n, \mathbf{S}_n, \gamma_n | \Omega_n, \Theta^*, \mathbf{P}^*) d(\mathbf{F}_n, \mathbf{S}_n, \gamma_n)$$

This integration is obviously infeasible by direct means. Using the posterior draws $(\mathbf{F}_n^{(g)}, \mathbf{S}_n^{(g)}, \gamma_n^{(g)})$ from the g th MCMC iteration, one can estimate it as

$$\hat{f}(\mathbf{Y}_n | \Omega_n, \Theta^*, \mathbf{P}^*) = \left(\frac{1}{n_1} \sum_{g=1}^{n_1} \frac{1}{f(\mathbf{Y}_n, \mathbf{F}_n^{(g)}, \mathbf{S}_n^{(g)}, \gamma_n^{(g)} | \Omega_n, \Theta^*, \mathbf{P}^*)} \right)^{-1}, \quad (3.40)$$

which is a simulation-consistent estimate of the likelihood. The calculation of the prior density is straightforward. Finally, the third term is obtained from a marginal-conditional decomposition following Chib and Jeliazkov (2001). It completes the calculation of the marginal likelihood.

4 Examples

4.1 Simulation Study

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

$$\mathbf{y}_t = \mathbf{f}_t + D_{s_t} e_t \quad (4.1)$$

$$\mathbf{f}_t = \mu_{s_t} + L \varepsilon_t \quad (4.2)$$

where $e_t \sim \text{i.i.d. } \mathcal{N}_2(0, \mathbf{I}_2)$ and $\varepsilon_t \sim \text{i.i.d. } \mathcal{N}_1(0, \mathbf{I}_1)$. To keep the discussion simple, the unobserved scalar, \mathbf{f}_t is a serially uncorrelated process with a regime switching mean, 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, $q = 2$, $k = 1$, $h = 0$, $\mathbf{a}_{s_t} = \mathbf{H}_{s_t} = G_{s_t} = 0$, $\beta_{s_t} = 0$ and $\mathbf{b}_{s_t} = 1$.

Table 1 shows the results of the simulation experiment examining the proposed MCMC method of the endogenous-switching 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 tables 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. 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. It 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.

parameters	$s_t = 1$				$s_t = 2$			
	true	mean	2.5%	97.5%	true	mean	2.5%	97.5%
μ_{s_t}	4.000	4.169	3.748	4.596	-4.000	-4.302	-4.979	-3.604
$d_{s_t}^{(1)}$	0.500	0.711	0.654	0.759	0.500	0.146	0.041	0.295
$d_{s_t}^{(2)}$	9.000	9.067	8.474	9.664	9.000	8.888	8.324	9.501
α_{s_t}	0.500	0.661	0.498	0.828	-0.500	-0.526	-0.699	-0.362
β_{s_t}	1.000	1.112	0.934	1.316	-1.000	-1.082	-1.217	-0.949
L	5.000	5.016	4.771	5.268				
ρ	0.600	0.676	0.560	0.772				

(a) when ρ is estimated

parameters	$s_t = 1$				$s_t = 2$			
	true	mean	2.5%	97.5%	true	mean	2.5%	97.5%
μ_{s_t}	4.000	3.291	2.899	3.696	-4.000	-3.359	-3.967	-2.718
$d_{s_t}^{(1)}$	0.500	0.501	0.208	0.680	0.500	0.477	0.294	0.703
$d_{s_t}^{(2)}$	9.000	9.027	8.440	9.657	9.000	8.898	8.325	9.526
α_{s_t}	0.500	0.555	0.389	0.724	-0.500	-0.580	-0.788	-0.387
β_{s_t}	1.000	0.989	0.826	1.155	-1.000	-1.058	-1.231	-0.895
L	5.000	4.859	4.623	5.100				

(b) when ρ is constrained to be zero

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.*

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

We now consider the Nelson-Siegel model that originally motivates the proposed modeling approach. The Nelson-Siegel model is widely 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 $y_t(\tau)$

$$\mathbf{y}_t = (y_t(\tau_1) \quad y_t(\tau_2) \quad \cdots \quad y_t(\tau_N))$$

is statistically modeled by

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

where

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

$$\mathbf{f}_t = (\mathbf{f}_t^L \quad \mathbf{f}_t^S \quad \mathbf{f}_t^C)' \tag{4.5}$$

$$e_t = (e_t(\tau_1) \quad e_t(\tau_2) \quad \cdots \quad e_t(\tau_N))' \tag{4.6}$$

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.

$$\mathbf{f}_t = \mu + G\mathbf{f}_{t-1} + \mathbf{L}\varepsilon_t \tag{4.7}$$

The coefficient λ , referred to the shape parameter, determines the exponential decay rate of the factor loadings, $\mathbf{\Lambda}$. In order to deal with the potential heteroscedasticity we assume that the volatility of the measurement errors, denoted by a diagonal matrix D_{s_t} , varies over time depending on a regime indicator s_t . 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 our prior. 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.

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

Theoretically speaking, the riskiness of the long term bonds that 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 it has been ignored by the literatures in the context of Nelson-Siegel models. To do that, we make regime switching endogenous by allowing for the contemporaneous correlation between the factor shocks ε_t and the latent variable γ_t whereas ε_t and e_t are independently normally distributed. These assumptions can be summarized by

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

where $\rho = (\rho_1 \ \rho_2 \ \rho_3)'$.

Notice 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 before and 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 1986:M1 to 2008:M12. This data is available online from the Board of Governors of the Federal Reserve System (Gurkaynak, Sack, and Wright (2007))

We first discuss the posterior estimates of the parameters. Table 2 summarizes the posterior distribution of the regime-specific parameters based on 50,000 of the MCMC algorithm beyond a burn-in of 5,000. We measure the efficiency of the MCMC sampling in terms of the acceptance rate in the M-H step and the inefficiency factors (Chib (2001)). These values on average are 62.3% and 28.3, respectively, indicating a well mixing, efficient sampler. Also, the sampler converges quickly to the same region of the parameter space regardless of the starting values. Finally, as one can see in Figure 2, the posterior densities of the regime-specific parameters are markedly different across the regimes. This can be also found in Table 2. From this table, we see that all the diagonal

parameters	$s_t = 1$			$s_t = 2$		
	mean	95% C.I.	ineff.	mean	95% C.I.	ineff.
$d_{s_t}^{(1)}$	-2.718	[-2.795, -2.645]	9.30	-1.587	[-1.724, -1.436]	51.19
$d_{s_t}^{(2)}$	-3.987	[-4.057, -3.915]	9.66	-2.928	[-3.044, -2.802]	36.28
$d_{s_t}^{(3)}$	-12.146	[-12.61, -11.73]	224.15	-12.292	[-12.86, -11.58]	40.61
$d_{s_t}^{(4)}$	-5.321	[-5.407, -5.234]	89.61	-4.302	[-4.428, -4.187]	30.64
$d_{s_t}^{(5)}$	-8.914	[-9.127, -8.590]	132.80	-10.459	[-11.18, -9.746]	150.19
$d_{s_t}^{(6)}$	-5.046	[-5.111, -4.979]	99.19	-3.830	[-3.918, -3.723]	91.15
$d_{s_t}^{(7)}$	-6.911	[-7.023, -6.778]	103.08	-3.219	[-3.362, -3.089]	191.37
$d_{s_t}^{(8)}$	-3.323	[-3.380, -3.276]	163.39	-6.922	[-7.124, -6.703]	231.69
$d_{s_t}^{(9)}$	-2.406	[-2.449, -2.371]	87.55	-2.313	[-2.369, -2.259]	90.42
α_{s_t}	0.956	[0.788, 1.126]	3.17	0.189	[-0.066, 0.450]	1.00

Table 2: **Regime-specific parameters** *This table presents the posterior mean, 95% credibility interval and inefficiency factor (ineff.) of the regime-dependent parameters based on 50,000 MCMC draws beyond a burn-in of 5,000. $d_{s_t}^{(i)}$ denotes i th diagonal element of D_{s_t}*

parameters	mean	95% C.I.	ineff.
λ	0.248	[0.248, 0.248]	260.32
$\mu^{(1)}$	0.063	[-0.021, 0.200]	9.86
$\mu^{(2)}$	-0.037	[-0.100, 0.008]	3.22
$\mu^{(3)}$	-0.052	[-0.149, 0.010]	8.55
$G^{(1)}$	0.985	[0.965, 0.998]	9.04
$G^{(2)}$	0.963	[0.940, 0.983]	2.87
$G^{(3)}$	0.921	[0.883, 0.954]	2.73
$L^{(1)}$	-0.990	[-1.060, -0.911]	144.30
$L^{(2)}$	-0.908	[-0.970, -0.867]	46.39
$L^{(3)}$	-0.371	[-0.425, -0.320]	166.26
$\mathbf{f}_0^{(1)}$	10.99	[10.374, 11.621]	1.45
$\mathbf{f}_0^{(2)}$	-4.140	[-4.857, -3.439]	0.48
$\mathbf{f}_0^{(3)}$	-0.732	[-2.214, 0.185]	1.59
ρ_1	0.041	[-0.160, 0.241]	0.87
ρ_2	-0.122	[-0.304, 0.076]	1.00
ρ_3	-0.264	[-0.385, -0.136]	2.29

Table 3: **Regime-independent parameters** *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.*

elements in D_{s_t} except for the 3rd and the 9th yield are regime-specific. According to the estimates for α_{s_t} , the regime changes are asymmetric because the transition probability from regime 1 (regime 2) to regime 2 (regime 2) is 83% (43%).

More importantly, Table 3 provides a strong evidence of endogenous regime switching. In particular, notice that the 95% credibility intervals of ρ_3 are entirely negative and the estimated \mathbf{f}_t^C is indeed the negative curvature as seen in the Figure 4. This indicates that the conditional volatility and the riskiness of the long term bond holdings, not surprisingly, move in the same direction. It may be interpreted as the regime switching in volatility is influenced by the shocks to the long term bond risk.

model	lnL	lnML	n.s.e.
Endogenous RS model ($\rho \neq 0$)	9670.21	to be done	to be done
Exogenous RS model ($\rho = 0$)	9063.84	to be done	to be done
Non-Switching model	8993.51	to be done	to be done

Table 4: **Marginal likelihoods**

Table 4 confirms the endogeneity of the regime changes based on the marginal likelihoods. As can be seen in this table, the endogenous regime switching model is most supported by the data. Finally, the Figure 3 shows the persistence of the volatility regimes. This figure reveals that the high volatility regime is far less persistent than the low volatility regime.

5 Conclusion

In this paper we propose regime switching linear state space models in which the regime switches are endogenously determined through the correlation with the observed or unobserved continuous state variables. Our work is an extension of Kim, Piger and Startz' (2008) to a general state space model. 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. By doing this we are able to estimate the models without approximation and inaccuracy. It also demonstrates the validity of our method by simulation study and application to a generalized Nelson-Siegel yield curve model with

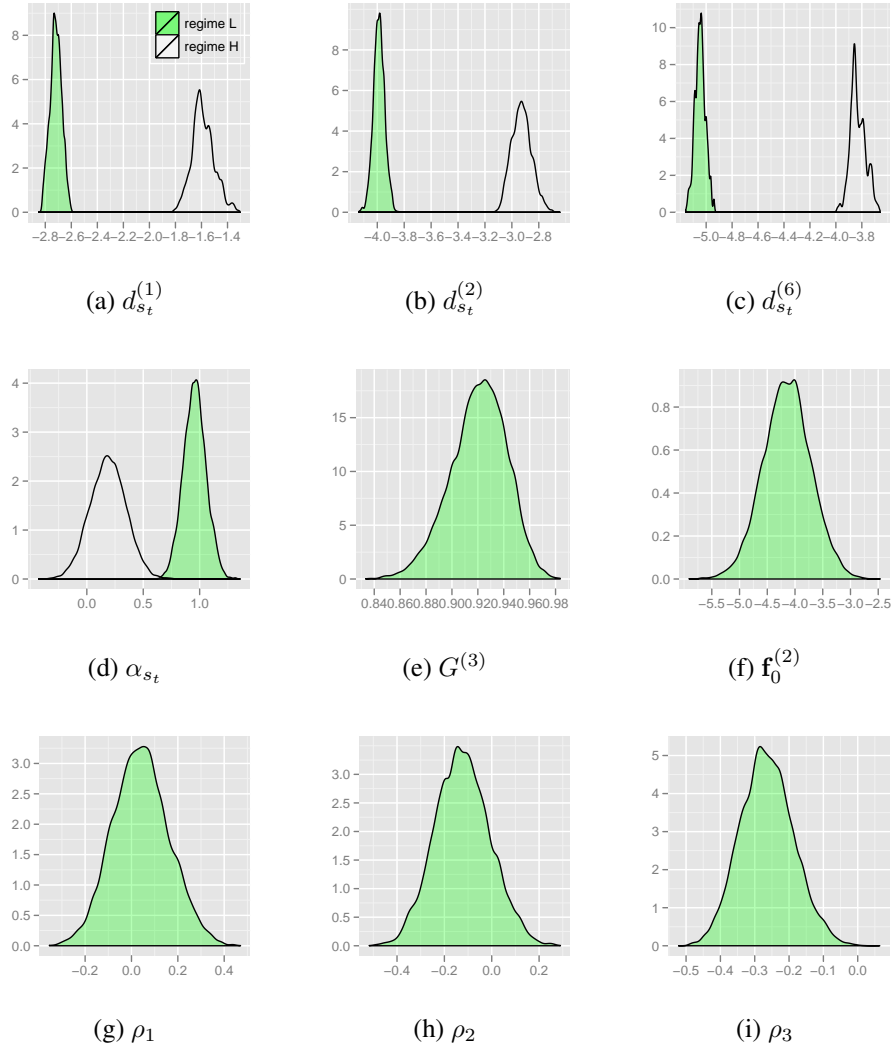
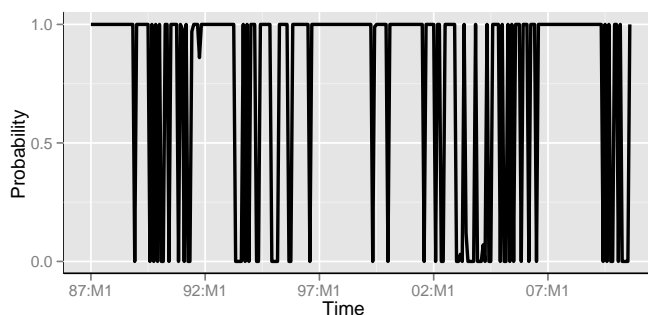


Figure 2: **Marginal posterior plots for some selected parameters** *These graphs are based on 50,000 simulated draws of the posterior simulation.*

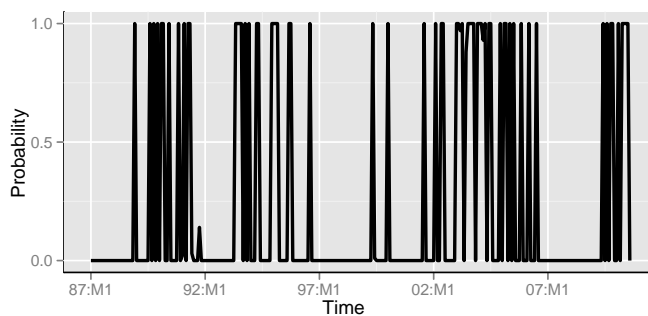
endogenous Markov switching volatility regime shifts.

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



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



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

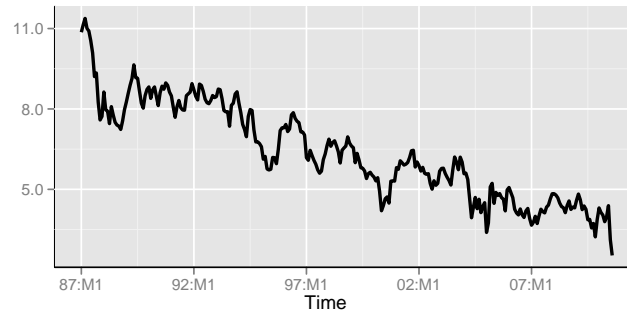
Figure 3: Probabilities of Regimes *These graphs plot the estimates of the probabilities of regimes. These graphs are based on 50,000 simulated draws of the posterior simulation.*

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

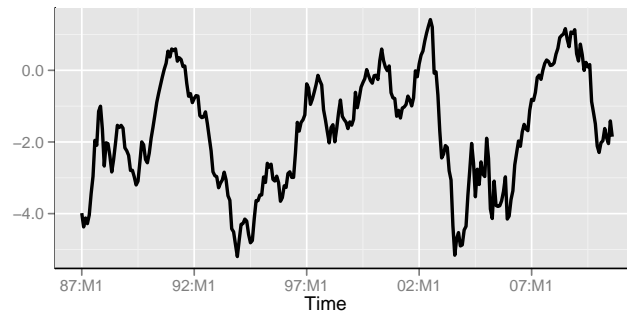
— (2001), “Markov chain Monte Carlo methods: computation and inference,” in *Handbook of Econometrics*, eds. Heckman, J. and Leamer, E., North Holland, Amsterdam, vol. 5, pp. 3569–3649.

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.

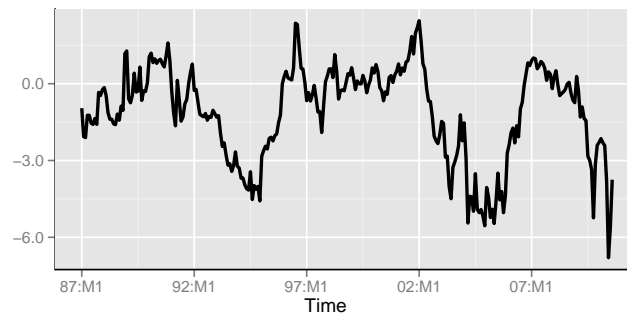
Chib, S. and Ergashev, B. (2009), “Analysis of multi-factor affine yield curve Models,” *Journal of the American Statistical Association*, 104, 1324–1337.



(a) Level



(b) Slope



(c) Curvature

Figure 4: **Factors** *These graphs plot the estimates of the factors. These graphs are based on 50,000 simulated draws of the posterior simulation.*

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 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. X. and Li, C. L. (2006), “Forecasting the term structure of government bond yields,” *Journal of Econometrics*, 130, 337–364.
- 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.
- Kim, C. J. (1994), “Dynamic linear models with Markov-switching,” *Journal of Econometrics*, 60, 1–22.
- Kim, C. J. and Nelson, C. R. (1998), “Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model With Regime Switching,” *Review of Economics and Statistics*, 80, 188–201.
- 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.

Sinclair, T. M. (2010), "Asymmetry in the Business Cycle: Friedman's Plucking Model with Correlated Innovations," *Studies in Nonlinear Dynamics and Econometrics*, 14, 235–243.