Bayesian Inference in Regime-Switching ARMA Models with Absorbing States: The Dynamics of the Ex-Ante Real Interest Rate Under Structural Breaks

Kim, Chang-Jin and Kim, Jaeho

Korea University and University of Washington, University of Washington

August 2013

Bayesian Inference in Regime-Switching ARMA Models with Absorbing States: The Dynamics of the Ex-Ante Real Interest Rate Under Structural Breaks

by

Chang-Jin Kim
University of Washington and Korea university
and
Jaeho Kim [1]
University of Washington

Preliminary Draft
August, 2013

Abstract

One goal of this paper is to develop an efficient Markov-Chain Monte Carlo (MCMC) algorithm for estimating an ARMA model with a regime-switching mean, based on a multi-move sampler. Unlike the existing algorithm of Billio et al. (1999) based on a single-move sampler, our algorithm can achieve reasonably fast convergence to the posterior distribution even when the latent regime indicator variable is highly persistent or when there exist absorbing states.

Another goal is to appropriately investigate the dynamics of the latent ex-ante real interest rate (EARR) in the presence of structural breaks, by employing the econometric tool developed. We argue Garcia and Perron's (1996) conclusion that the EARR rate is a constant subject to occasional jumps may be sample-specific. For an extended sample that includes recent data, Garcia and Perron's (1996) AR(2) model of EPRR may be misspecified, and we show that excluding the theory-implied moving-average terms may understate the persistence of the observed ex-post real interest rate (EPRR) dynamics. Our empirical results suggest that, even though we rule out the possibility of a unit root in the EARR, it may be more persistent and volatile than has been documented in some of the literature including Garcia and Perron (1996).

Key Words: ARMA model with Regime Switching, Multi-move Sampler, Single-Move Sampler, Metropolis-Hastings Algorithm, Absorbing State, Ex-Ante Real Interest Rate.

[1] **Chang-Jin Kim:** Dept. of Economics, University of Washington, Seattle, WA and Dept. of Economics, Korea University, Seoul, Korea (changjin@u.washington.edu); **Jaeho Kim**: Dept. of Economics, Univ. of Washington, Seattle, WA. (jaeho@u.washington.edu).
Chang-Jin Kim acknowledges financial support from the Bryan C. Cressey Professorship at the University of Washington. Jaeho Kim acknowledges the support of the Grover and Creta Ensley Fellowship in Economic Policy from the University of Washington.

1. Introduction

The ex-ante real interest rate (EARR) is a key economic variable which affects economic agents' intertemporal consumption, savings, and investment decisions. Its dynamics play a central role in many theoretical models such as asset pricing models, and macro DSGE models. Thus, understanding the behavior of the EARR has been a crucial issue in the literature, as surveyed in Neely and Rapach (2008).

The seminal article by Fama (1975) provides striking empirical evidence that U.S. EARR is essentially constant. Nelson and Schwert (1977) and Garbade and Wachtel(1978), however, challenge Fama (1975)'s finding by showing that his statistical test is not informative enough to conclude the behavior of the EARR and raise the possibility of a time-varying EARR. In the subsequent studies by Mishkin (1981), Huizinga and Mishkin (1986), Antoncic (1986), they also show that the empirical result of constant U.S. EARR is critically dependent upon a particular sample period and thus, it is hard to confirm Fama (1975)'s argument. Building upon those empirical findings, Rose(1988) even raises the possibility that the EARR may be an I(1) process. Since Rose(1988) has raised the issue, literature has reported mixed results. By applying various unit root and cointegration tests to ex-post real interest rate (EPRR), King et al. (1991), Gali (1992), Mishkin (1992), and Koustas and Serletis (1999) conclude that the EARR is nonstationary with a unit root. [2] On the other hand, Crowder and Hoffman (1996), and Rapach and Weber(2004) argue that the EARR is stationary but highly persistent. Additionally, Sun and Phillips (2004) show that the EARR has mean-reverting dynamics with long-memory properties, based on fractional integration tests.

Another strand of the empirical literature on this issue is to investigate the implications of regime shifts in the real interest rates on the persistence of the EARR. Note that Perron (1990) argues that a failure to account for mean shifts may lead to spurious evidence of high persistence for a series under consideration. Thus, Caporale and Grier (2000), and Bai and Perron (2003) confirm that the unit root hypothesis can be rejected if shifts in the mean are allowed for the ex-post real interest rate, suggesting that the EARR is stationary. By incorporating regime shifts or structural breaks in the mean of EARR in an autoregressive

[2] Under rational expectations, a unit root in the ex-ante real interest rate implies a unit root in the ex-post real interest rate.

model of EPRR, Garcia and Perron (1996) even show that the EARR rate may be a constant subject to occasional jumps caused by important structural events.

One goal of this paper is to appropriately investigate the dynamics of the EARR, in the presence of structural breaks in its mean with unknown break points. Under the maintained hypothesis of rational expectations, if we assume that the EARR follows an AR(2) process then the ex-post real rate follows an ARMA(2,2) process. This is because the ex-post real rate is a sum of an AR(2) process for the EARR and a serially uncorrelated inflation forecast error. We argue that omitting the moving average terms as in Garcia and Perron (1996) may result in misleading inference about the dynamics of the EARR. Furthermore, approximating the moving-average components in the ex-post real interest rate with a finite order autoregressive process would result in size distortions in testing for a unit root. If the ex-post real rate follows an ARMA(2,2) process with a regime-switching mean, however, estimation of the model is not as straightforward as in Garcia and Perron's (1996) regime-switching model, in which the moving average terms implied by the rational expectations theory are omitted.

Thus, another goal of this paper is to develop an efficient Bayesian method for estimating an ARMA model with a regime-switching mean, which will be used as an econometric tool to be employed in achieving the goal of investigating the dynamics of the EARR. In case the variance of the disturbance terms is i.i.d within a regime, the approximate maximum likelihood estimation of the model is readily available based on the state-space representation of the model, as proposed by Kim (1994). However, with heteroscedastic disturbances within a regime, estimation of the model is infeasible within the classical framework, leading us to resort to the Bayesian approach.

Our Bayesian approach builds on the work of Billio et al. (1999) in that we effectively incorporate their Metropolis Hastings algorithm. That is, at each iteration of the Markov-chain Monte Carlo algorithm, the whole sequence of the latent regime indicator variable is drawn from the proposal density which can reasonably approximate the target density, conditional on all the parameters of the model and data. [3] Then, the approximation error in

[3] Throughout the paper, we focus on generating the regime indicator variables S_t, $t = 0, 1, 2, ..., T$, conditional on the parameters of the model. We resort to Chib and Greenberg (1994) and Nakatsuma (2000), for making inferences about the parameters of the model conditional on the regime indicator variables and data.

the proposal density is corrected for by globally accepting or rejecting the newly drawn regime indicator variables according to an appropriately defined acceptance probability. What's different from Billio et al.'s (1999) approach is that we employ a multi-move sampler as opposed to their single-move sampler, when drawing the sequence of the regime-indicator variables.

Note that, as theoretically proven by Liu et al. (1994) and (2002), a multi-move sampler significantly reduces the autocorrelations among successive draws of the regime-indicator variables and other parameters of the model in MCMC iterations. Carter and Kohn (1994), Shephard (1994), and de Jong and Shephard (1995) empirically show that the multi-move samplers are more efficient than the single-move samplers, in the sense that convergence to the posterior distribution will be faster and estimates of the posterior moments will have smaller variances. Actually, there is a case in which the single-move sampler results in no convergence to the posterior distribution at all in a regime-switching ARMA model. This is the case when there exist absorbing states. With absorbing states, correlations between two subsequent latent regime-indicator variables are perfect or almost perfect. As a result, the desired asymptotic posterior distributions are never achieved by the single-move-based algorithm. Garcia and Perron (1996), in their maximum likelihood estimation of a three-state Markov-switching AR model for the ex-post real interest rate, show that their estimates of the transition probabilities imply existence of structural breaks with two absorbing states. Thus, with absorbing states or structural breaks in the mean of our ARMA process for the EPRR, the single-move Gibbs sampler would never achieve convergence. We show that our algorithm based on a multi-move sampler can achieve reasonably fast convergence even in such a case.

The remainder of the paper is organized as follows. Section 2 present our benchmark econometric model and provides a literature review on the inference of regime-switching ARMA models. Section 3 provides a new efficient MCMC algorithm based on a multi-move sampler, for drawing the Markov-switching regime-indicator variables conditional on all parameters of the model. In Section 4, we perform simulation studies in order to evaluate the performance of the proposed Bayesian algorithm. In particular, we show that our multi-move sampler achieves reasonably fast convergence, even in the case in which the single-move sampler fails to converge at all. In section 5, the benchmark model in Section 2 is extended to incorporate stochastic volatility in the disturbance terms, and then the extended model is applied to investigate the dynamics of the latent ex-ante real interest rate by estimating a regime-switching ARMA model for the ex-post real interest rate. Section 5 provides a summary and concluding remarks.

2. Model Specification and Literature Review on Markov-Switching ARMA Models: Critique

Consider the following ARMA(p,q) model with regime-dependent coefficients: [4]

$$y_t = \mu_{S_t} + \sum_{i=1}^{p} \phi_{i,S_t}(y_{t-i} - \mu_{S_{t-i}}) + e_t - \sum_{j=1}^{q} \theta_{i,S_t} e_{t-j}, \quad e_t \ \sim i.i.d.N(0,\sigma^2), \tag{1}$$

where the subscript S_t suggests that the corresponding coefficient is dependent on a latent regime-indicator variable S_t. We assume that S_t follows an $M-$state first order Markov switching process with the following transition probabilities:

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

Note that, by restricting the transition probabilities of the above regime-switching model appropriately to allow for absorbing states, one can design a model of structural break with unknown break point, as suggested by Chib (1998). Later in Section 5, an extended version of this model is applied to the ex-post real interest rate. To deal with the non-i.i.d. nature of the shocks to ex-post real interest rate within a regime, the model will be extended to allow for stochastic volatility in the disturbance terms. For simplicity of exposition, we stick to the above model specification in this Section.

Due to its non-Markovian nature, the above model is not easy to estimate. Within the classical framework, for example, evaluation of the likelihood function is not feasible without

[4] We focus on generating the regime indicator variables S_t, $t = 0,1,2,...,T$, conditional on the parameters of the model and data. We present the MCMC algorithm for generating the parameters of the model conditional on the regime-indicator variables and data in Appendix A, by complementing those in Chib and Greenberg (1994) and Nakatsuma (2000).

resorting to some sort of approximation. This is because the conditional density of y_t depends upon the entire history of the latent regime-indicator variable up to time t. To get over this problem, we can first cast the above model into a state-space model. We can then employ the approximate Kalman filter algorithm proposed by Kim (1994). The basic idea in Kim (1994) is to employ an approximation to the conditional density of y_t, so that it can be dependent only on $S_t = j$ and $S_{t-1} = i$, $(i, j = 1, 2, ..., M)$ at each iteration of the Kalman filter. His method is easy to implement for the above model with i.i.d. disturbance terms. However, if the above model is extended to deal with stochastic volatility in the disturbance terms, his approach is no longer applicable. Only within the Bayesian framework, is estimation of the extended model feasible.

Within the Bayesian framework, Billio et al. (1999) propose a Markov-Chain Monte Carlo (MCMC) algorithm for sampling the regime-indicator variables S_t, $t = 1, 2, ..., T$, from an appropriate proposal density which can appropriately approximate the target density. Then, they correct for the approximation error in the proposal density by employing the Metropolis Hastings (MH) algorithm. [5] For example, once the whole sequence of the regime indicator variable is drawn from the proposal density, the approximation error is corrected for by globally accepting or rejecting the newly drawn regime indicator variables according to an appropriately defined acceptance probability. In drawing the regime-indicator variables, Billio et al. (1999) resort to a single-move sampler, in which a single indicator variable S_t is drawn one at a time for $t = 1, 2, ..., T$, conditional on the remaining regime-indicator variables S_1, S_2, ..., S_{t-1}, S_{t+1}, ..., S_T. In what follows, we provide a review of Billio et al.'s (1999) algorithm based on a single-move sampler.

Review of MCMC Algorithm based on a Single-Move Sampler

The goal is to generate $\tilde{S}_T = [\, S_0 \quad S_1 \quad \ldots \quad S_T \,]'$ from the target density

[5] Readers are referred to Chib and Greenberg (1995), Gilks et al. (1996), and Koop (2003) for the MH algorithm and references therein.

$$
\begin{aligned}
F(\tilde{S}_T|\tilde{Y}_T) &= \frac{f(\tilde{S}_T)\ f(\tilde{Y}_T|\tilde{S}_T)}{f(\tilde{Y}_T)} \\
&= \frac{f(S_0)\ \prod_{t=1}^{T}\ f(S_t|S_{t-1})\ \prod_{t=1}^{T}\ f(y_t|\tilde{S}_t, \tilde{Y}_{t-1})}{f(\tilde{Y}_T)}.
\end{aligned} \tag{3}
$$

For a direct single-move Gibbs sampler, one can theoretically draw S_t, for $t = 0, 1, 2, ..., T$, from

$$
\begin{aligned}
f(S_t|\tilde{S}_{\neq t}, \tilde{Y}_T) &= \frac{f(S_t|\tilde{S}_{\neq t})\ f(\tilde{Y}_T|\tilde{S}_T)}{f(\tilde{Y}_T|\tilde{S}_{\neq t})} \\
&\propto f(S_t|\tilde{S}_{\neq t})\ f(\tilde{Y}_T|\tilde{S}_T) \\
&\propto f(S_{t+1}|S_t)\ f(S_t|S_{t-1})\ \prod_{t=1}^{T} f(y_t|\tilde{S}_T, \tilde{Y}_{t-1}) \\
&= f(S_{t+1}|S_t)\ f(S_t|S_{t-1})\ \prod_{t=1}^{T} f(y_t|\tilde{S}_t, \tilde{Y}_{t-1}) \\
&\propto f(S_{t+1}|S_t)\ f(S_t|S_{t-1})\ \prod_{k=t}^{T} f(y_k|\tilde{S}_k, \tilde{Y}_{k-1}),
\end{aligned} \tag{4}
$$

where $\tilde{S}_t = [\,S_0 \quad S_1 \quad \ldots \quad S_t\,]'$; $\tilde{S}_{\neq t}$ is $\tilde{S}_T$ excluding S_t; $\tilde{Y}_\tau = [\,y_1 \quad y_2 \quad \ldots \quad y_\tau\,]'$; and $f(S_{t+1}|S_t)$ and $f(S_t|S_{t-1})$ are the transition probabilities. The validity of going from the second line to the third line is ensured by the Markov property of S_t. As we go from the third line to the forth line, all irrelevant future states, S_τ, $\tau = t+1, ..., T$, are dropped. [6]

However, for each generation of S_t one needs to evaluate the individual likelihood functions $f(y_k|\tilde{S}_k, \tilde{Y}_{k-1})$, $k = t, t+1, ..., T$. This means that the sampling scheme requires $O(\frac{T(T+1)}{2})$ operations. Consequently, as the number of regimes or the sample size increases, the algorithm becomes infeasible as computational costs increase exponentially.

In order to get over the problem, Billio et al. (1999) propose a Metropolis Hastings algorithm as an alternative to the direct Gibbs sampling approach. Instead of generating

[6] For an AR(p) process without a moving-average term in Albert and Chib (1993), equation (4) can be simplified as:

$$
f(S_t|\tilde{S}_{\neq t}, \tilde{Y}_T) \propto f(S_{t+1}|S_t)\ f(S_t|S_{t-1})\ \prod_{k=t}^{t+p} f(y_k|\tilde{S}_k, \tilde{Y}_{k-1}).
$$

individual S_t directly from the density in equation (4) for $t = 0, 1, 2, ..., T$, they propose to generate it from the following individual proposal density:

$$g(S_0|\tilde{S}_{\neq 0}, \tilde{Y}) \propto f(S_1|S_0)f(S_0), \quad for \;\; t = 0, \tag{5}$$

$$g(S_t|\tilde{S}_{\neq t}, \tilde{Y}) \propto f(S_{t+1}|S_t)\; f(S_t|S_{t-1})\; f(y_t|\tilde{S}_t, \tilde{Y}_{t-1}), \quad for \;\; t = 1, ..., T-1. \tag{6}$$

$$g(S_T|\tilde{S}_{\neq T}, \tilde{Y}) \propto f(S_T|S_{T-1})\; f(y_T|\tilde{S}_T, \tilde{Y}_{T-1}), \quad for \;\; t = T, \tag{7}$$

which is an approximation to the individual target density in equation (4). As the above density depends only on density of y_t, generating individual S_t is an $O(T)$ algorithm unlike the Gibbs sampling approach of generating individual S_t from equation (4).

As the above individual proposal densities are based on approximations, Billio et al. (1999) propose to employ the Metropolis Hastings algorithm. Once a set of candidate $\tilde{S}$ is drawn from the individual candidate densities, the approximation errors can be corrected for by globally accepting or rejecting the generated $\tilde{S}_T$ according to an appropriately defined acceptance probability. By defining $\tilde{S}_T^J$ to be the newly generated set of $\tilde{S}_T$ and $\tilde{S}_T^{J-1}$ to be an accepted set of $\tilde{S}_T$ at the previous iteration of the sampler, the acceptance probability is defined as:

$$\alpha(\tilde{S}_T^J, \tilde{S}_T^{J-1}) = \; min[\frac{F(\tilde{S}_T^J|\tilde{Y}_T)}{F(\tilde{S}_T^{J-1}|\tilde{Y}_T)} \frac{G(\tilde{S}_T^{J-1}|\tilde{Y}_T)}{G(\tilde{S}_T^J|\tilde{Y}_T)}, \; 1], \tag{8}$$

where, by considering the normalizing constants, the proposal density $G(\tilde{S}_T|\tilde{Y}_T)$ is given by:

$$G(\tilde{S}_T|\tilde{Y}_T) = \prod_{t=0}^{T} \left[\frac{g(S_t|\tilde{S}_{\neq t}, \tilde{Y}_t)}{\sum_{S_t=1}^{M} g(S_t|\tilde{S}_{\neq t}, \tilde{Y}_t)} \right]. \tag{9}$$

By substituting equations (3) and (9) into (8) and rearranging terms, Billio et al. (1999) derive the following acceptance probability:

$$\alpha(\tilde{S}_T^J, \tilde{S}_T^{J-1}) = min \left[\prod_{t=1}^{T} \frac{f(S_t^{J-1}|S_{t-1}^{J-1})}{f(S_t^J|S_{t-1}^J)} \prod_{t=0}^{T} \frac{\sum_{S_t} f(S_{t+1}^J|S_t) f(S_t|S_{t-1}^J) f(y_t|\tilde{S}_t^J, \tilde{Y}_{t-1})}{\sum_{S_t} f(S_{t+1}^{J-1}|S_t) f(S_t|S_{t-1}^{J-1}) f(y_t|\tilde{S}_t^{J-1}, \tilde{Y}_{t-1})}, \; 1 \right]. \tag{8'}$$

As discussed in Liu et al. (1994) and Scott (2002), however, a potential weakness of the single-move sampler is that its performance gets worse with slower mixing as the persistence of the latent state variable increases. [7] Furthermore, slower mixing for the regime-indicator variables translates into slower mixing for the parameters of the model as well, according to a duality principle introduced by Diebolt and Robert (1994). Actually, our simulation study in Section 3 shows that there are cases in which the single-move sampler results in no convergence to the posterior distribution at all. This happens when the Markov-switching regime indicator variable is highly persistent or when there exists an absorbing state, as in Garcia and Perron (1996). Note that Garcia and Perron (1996), in their maximum likelihood estimation of a Markov-switching AR model for the ex-post real interest rate, show that some of the transition probabilities are estimated to be close to zero. We show that the efficient algorithm based on a multi-move sampler proposed in the next section can achieve reasonably fast convergence even in these cases.

3. A New Efficient MCMC Algorithm based on a Multi-Move Sampler

In this section, we attempt to get over the weaknesses of the above-mentioned single-move sampler by implementing an efficient Metropolis Hastings algorithm based on a multi-move sampler. A successful implementation of the Metropolis Hastings algorithm depends critically upon the appropriate derivation of a candidate density that reasonably approximates the target density. We thus consider the following decomposition of the target density $F(\tilde{S}_T|\tilde{Y}_T)$:

$$F(\tilde{S}_T|\tilde{Y}_T) = f(S_T|\tilde{Y}_T)\prod_{t=0}^{T-1} f(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T), \tag{10}$$

where $\tilde{S}_{t+1:T} = [\, S_{t+1} \quad S_{t+2} \quad \dots \quad S_T \,]'$.

Theoretically, the above decomposition suggests that one can sequentially generate S_T from $f(S_T|\tilde{Y}_T)$, and then S_t from the conditional density $f(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T)$, for $t = T-1, ..., 0$.

[7] In probability theory, the mixing time of a Markov chain means the time until the Markov chain reaches the steady-state distribution. The mixing time determines the running time for simulation.

By defining $\tilde{Y}_t = [\,y_1 \quad y_2 \quad \ldots \quad y_t\,]'$ and $\tilde{Y}_{t+1:T} = [\,y_{t+1} \quad y_{t+2} \quad \ldots \quad y_T\,]'$, this conditional density can be derived as:

$$\begin{aligned}
f(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T) &= f(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_t, \tilde{Y}_{t+1:T}) \\
&= \frac{f(S_t, \tilde{Y}_{t+1:T}|\tilde{S}_{t+1:T}, \tilde{Y}_t)}{f(\tilde{Y}_{t+1:T}|\tilde{S}_{t+1:T}, \tilde{Y}_t)} \\
&\propto f(S_t, \tilde{Y}_{t+1:T}|\tilde{S}_{t+1:T}, \tilde{Y}_t) \\
&= f(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_t)\; f(\tilde{Y}_{t+1:T}|\tilde{S}_{t:T}, \tilde{Y}_t) \\
&\propto f(S_{t+1}|S_t) f(S_t|\tilde{Y}_t) \prod_{k=t+1}^{T} f(y_k|\tilde{S}_{t:k}, \tilde{Y}_{k-1}).
\end{aligned} \tag{11}$$

However, evaluating the above density is not feasible in the presence of a non-trivial moving-average structure. Thus, we propose to sequentially generate S_t, $t = T, T-1, ..., 1, 0$, from the individual proposal density given below, as an approximation to the density in equation (11):

$$g(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T) \propto f(S_{t+1}|S_t) h(S_t|\tilde{Y}_t), \tag{12}$$

where $f(S_{t+1}|S_t)$ is the transition probability and the $h(S_t|\tilde{Y}_t)$ term is an approximation to the $f(S_t|\tilde{Y}_t)$ term in equation (11). The nature of approximation in the $h(S_t|\tilde{Y}_t)$ term is discussed below. An additional approximation involved is that we ignore $\prod_{k=t+1}^{T} f(y_k|\tilde{S}_{t:k}, \tilde{Y}_{k-1})$ from equation (11).

Building upon ideas in Hamilton (1988, 1989), Cosslett and Lee (1985) and Harrison and Stevens (1976), Kim (1994) presents filtering and smoothing algorithms for a state-space model with Markov switching, along with maximum likelihood estimation of the unknown parameters of the model. In particular, by combining the Hamilton filter (1989) and an approximate Kalman filter, he provides an algorithm for obtaining $h(S_t|\tilde{Y}_t)$ as an approximation to $f(S_t|\tilde{Y}_t)$ for a general state-space model with Markov switching. Note that an ARMA model with Markov switching can always be cast into a state-space model with Markov switching. For details of Kim's (1994) approximate Kalman filter and algorithm for calculating $h(S_t|\tilde{Y}_t)$ as an approximation to $f(S_t|\tilde{Y}_t)$, readers are referred to Appendix B.

Once $\tilde{S}_T$ is generated from the multi-move candidate density in equation (12), we follow Billio et al. (1999) in adopting a global Metropolis-Hastings approach in order to correct for the approximations involved in our candidate density. We accept or reject globally the whole sequence of $S_0, S_1, ..., S_T$, using an appropriate acceptance probability. Let $\tilde{S}_T^J$ and $\tilde{S}_T^{J-1}$ be the sequences of $S_0, S_1, ..., S_T$ generated at the current and the previous iterations of the MCMC algorithm, respectively. Then, the acceptance probability is given by:

$$\alpha(\tilde{S}_T^J, \tilde{S}_T^{J-1}) = \ min\left[\frac{F(\tilde{S}_T^J|\tilde{Y}_T)}{F(\tilde{S}_T^{J-1}|\tilde{Y}_T)}\frac{G(\tilde{S}_T^{J-1}|\tilde{Y}_T)}{G(\tilde{S}_T^J|\tilde{Y}_T)}, 1\right], \tag{13}$$

where $F(.|\tilde{Y}_T)$ is given in equation (3), as rewritten below:.

$$\begin{aligned} F(\tilde{S}_T|\tilde{Y}_T) &= \frac{f(\tilde{S}_T)\ f(\tilde{Y}_T|\tilde{S}_T)}{f(\tilde{Y}_T)} \\ &= \frac{f(S_0)\ \prod_{t=1}^T\ f(S_t|S_{t-1})\ \prod_{t=1}^T\ f(y_t|\tilde{S}_t, \tilde{Y}_{t-1})}{f(\tilde{Y}_T)}, \end{aligned} \tag{3}$$

and $G(.|\tilde{Y}_T)$ is the multi-move candidate density defined below:

$$\begin{aligned} G(\tilde{S}_T|\tilde{Y}_T) &= \prod_{t=0}^T\left[\frac{g(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T)}{\sum_{S_t} g(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T)}\right] \\ &= \prod_{t=0}^T\left[\frac{f(S_{t+1}|S_t)h(S_t|\tilde{Y}_t)}{\sum_{S_t} f(S_{t+1}|S_t)h(S_t|\tilde{Y}_t)}\right]. \\ &= \prod_{t=0}^T\left[\frac{f(S_{t+1}|S_t)h(S_t|\tilde{Y}_t)}{h(S_{t+1}|\tilde{Y}_t)}\right] \end{aligned} \tag{14}$$

By substituting equations (3) and (14) into equation (13), we can derive the following acceptance probability:

$$\alpha(\tilde{S}_T^J, \tilde{S}_T^{J-1}) = \ min\left[\prod_{t=1}^T\frac{f(y_t|\tilde{S}_t^J, \tilde{Y}_{t-1})}{f(y_t|\tilde{S}_t^{J-1}, \tilde{Y}_{t-1})}\prod_{t=1}^T\frac{h(S_t^{J-1}|\tilde{Y}_t)}{h(S_t^J|\tilde{Y}_t)}\prod_{t=0}^{T-1}\frac{h(S_{t+1}^J|\tilde{Y}_t)}{h(S_{t+1}^{J-1}|\tilde{Y}_t)}, 1\right], \tag{13$'$}$$

where $h(S_t|\tilde{Y}_t)$ can be obtained by applying the approximate filter of Kim (1994) to the state-space model representation of the Markov-switching ARMA model; and $f(y_t|\tilde{S}_t, \tilde{Y}_{t-1})$ can be evaluated by applying the conventional Kalman filter to the state-space model. What follows describes a brief summary of the Metropolis Hastings algorithm for generating $\tilde{S}_T$.

Summary of Metropolis Hastings Algorithm for Generating $\tilde{S}_T$ at the $J-th$ Iteration

i) We cast the Markov-switching ARMA model into in a state-space form, conditional on all the parameters. For a state-space representation of the model, readers are referred to Appendix B.

ii) We apply the approximate filter in Kim (1994) to the state-space representation of the model in order to evaluate and save $h(S_t|\tilde{Y}_t)$ and $h(S_{t+1}|\tilde{Y}_t)$. In this step, we also calculate and save $h(S_t^{J-1}|\tilde{Y}_t)$ and $h(S_{t+1}^{J-1}|\tilde{Y}_t)$, where S_t^{J-1} and S_{t+1}^{J-1} refer to the regime indicator variables generated at the previous iteration of the Sampler.

iii) Using $h(S_t|\tilde{Y}_t)$ and $h(S_{t+1}|\tilde{Y}_t)$ saved from ii), we generate S_t sequentially in the backward direction for $t = T, T-1, ..., 1, 0$, based on the individual proposal density in equation (5). In this step, we save $h(S_t^J|\tilde{Y}_t)$ and $h(S_{t+1}^J|\tilde{Y}_t)$, where S_t^J and S_{t+1}^J refer to the regime indicator variables generated.

iv) We apply the conventional Kalman filter again to the state-space model representation of the model conditional on $\tilde{S}_T = \tilde{S}_T^J$, in order to evaluate and save $f(y_t|\tilde{S}_t^J, \tilde{Y}_{t-1})$, $t = 1, 2, ..., T$.

v) We apply the conventional Kalman filter to the state-space model representation of the model conditional on $\tilde{S}_T = \tilde{S}_T^{J-1}$, in order to evaluate and save $f(y_t|\tilde{S}_t^{J-1}, \tilde{Y}_{t-1})$.

vi) Using the output from ii)-v), we calculate the acceptance probability as in equation (13'). Then, we accept or reject $\tilde{S}_T^J$ according to this acceptance probability.

4. Performance of the Proposed Algorithm: Simulation Study

In this section, we compare the performances of the proposed multi-move sampler and Billio et al.'s (1999) single-move sampler. For this purpose, we consider the following ARMA(1,1) model with a Markov-switching mean as the data generating process:

$$y_t = \mu_{S_t} + \phi(y_{t-1} - \mu_{S_{t-1}}) + e_t - \theta e_{t-1},$$

$$e_t \sim i.i.d.N(0, \sigma^2),$$

$$Pr[S_t = j|S_{t-1} = i] = p_{ij}, \ i, j = 1, 2,$$

$$t = 1, 2, ..., 300.$$

We generate three sets of data from three alternative cases with different sets of parameters. We first consider a case in which both the proposed multi-move sampler and the single-move sampler achieve fast convergences. We then consider a case in which the single-move sampler converges much slower than the proposed multi-move sampler. This is the case in which more persistent dynamics for the latent regime-indicator variable with the transition probabilities being closer to 1. In our third case, the single-move sampler never achieves convergence, while the proposed multi-move sampler continues to have reasonably fast convergence. This is the case a structural break with an absorbing state. The three alternative cases with different parameters are given by:

Case #1: Benchmark Case

$$\mu_1 = 0.4;\ \mu_2 = 0;\ p_{11} = 0.9;\ p_{22} = 0.96;$$

$$\phi = 0.3;\ \theta = 0.6;\ \sigma = 0.2,$$

Case #2: Higher Persistence for the Regime-Indicator Variable

$$\mu_1 = 0.4;\ \mu_2 = 0;\ p_{11} = 0.95;\ p_{22} = 0.99;$$

$$\phi = 0.3;\ \theta = 0.6;\ \sigma = 0.2,$$

Case #3: The Case of a Structural Break with an Absorbing State

$$\mu_1 = 0.4;\ \mu_2 = 0;\ p_{11} = 0.993;\ p_{22} = 1;$$

$$\phi = 0.3;\ \theta = 0.6;\ \sigma = 0.2,$$

where the values of the ϕ, θ, and σ parameters are the same for all the cases. In order to generate data, we need $\tilde{S}_T$ generated using the transition probabilities. We use $\tilde{S}_T$ the elements of which are assigned according to expected durations of the regimes calculated based the assigned transition probabilities. The sample size T is 300. In implementing the two alternative MCMC algorithms, we employ the same prior distributions for the parameters for all cases.

In Table 1, the prior and the posterior moments of the parameters for Case #1 are reported. For both algorithms, convergence is achieved after reasonable numbers of iterations. The posterior means or medians of the parameters are close to their true values for the two algorithms. The posterior standard deviations of the parameters are almost the same for the two algorithms. The posterior probabilities of regime 2, which are depicted in Figure 1.A against the shared true periods of regime 2, are also almost the same for the two alternative algorithms, with the correct assignment rates of the regimes being close to one. As depicted in Figure 1.B, the cumulative averages of the MCMC samples for selected parameters from the two algorithms converge reasonably fast to the true parameters. The autocorrelations of the MCMC samples depicted in Figure 1.C also die out fast for both algorithms. To sum up, both the proposed multi-move sampler and the single-move sampler perform equally well for the benchmark case, with satisfactory mixing properties.

In Case #2, we increase the transition probabilities closer to 1, while maintaining the other parameters the same as in Case #1. Simulation results are reported in Table 2. As in the benchmark case, the posterior means or medians of the parameters are close to their true values. The posterior standard deviations of the parameters are almost the same for the two algorithms. Furthermore, the posterior probabilities of regime 2, which are depicted in Figure 2.A, are almost the same for the two algorithms with the correct assignment rates of the regimes being close to one. However, notice that the posterior moments of the parameters and the posterior regime probabilities for the proposed multi-move algorithm are calculated based on the 10,000 MCMC samples after 5,000 burn-in's, while those for the single-move algorithm are calculated based on 10,000 MCMC samples after 140,000 burn-in's. That is, with higher transition probabilities, the convergence of the single-move sampler is extremely slow. Such extremely slow convergence of the single-move sampler is shown in Figures 2.B and 2.C, which depict the cumulative averages and the autocorrelations of the MCMC samples for selected parameters. For example, while the autocorrelations of the MCMC samples for the multi-move sampler die out very quickly, those for the single-move sampler remain very high even at the lag of 1000. For Case #2, the single-mover sampler has a considerably inferior mixing property than the proposed multi-move sampler.

In Case #3, we deal with the case in which the single-move sampler never achieves convergence. For a model with a structural break in the sample, where state 2 is an absorbing state with $Pr[S_t = 1|S_{t-1} = 1] = 1$, correlations between two subsequent states within regime 2 are perfect. As a result, the desired asymptotic posterior distributions are never achieved by the single-move-based algorithm. Therefore, we report only the results from the proposed multi-move sampler. The prior and the posterior moments of the parameters are summarized in Table 3. Posterior means or modes of the parameters are close to the true values. In Figure 3.A, the posterior probabilities of regime 2 is depicted against the shaded true period of regime 2. Our multi-move sampler does an excellent job of inferring the regimes, with correct assignment rate being about 98%. Furthermore, convergence of the sampler is achieved within 20,000 iterations. Cumulative averages and Autocorrelations of MCMC Samples for selected parameters, depicted in Figures 3.B and 3.C, also confirm fast convergence of our multi-move sampler.

5. Uncovering the Dynamics of U.S. Ex-Ante Real Interest Rate Under Regime Shifts: 1960Q1-2008Q2

5.1. Model Specification for Ex-Post Real Interest Rate

Consider the following expression for the nominal interest rate (i_t):

$$i_t = r_t^{EA} + E[\pi_t|I_{t-1}] \tag{16}$$

where r_t^{EA} denotes the EARR; π_t denotes the inflation rate; and $E[\pi_t|I_{t-1}]$ refers to economic agents' rational expectation of π_t conditional on all the available information up to period $t-1$. Then the ex-post real interest rate (r_t^{EP}) is given by:

$$r_t^{EP} = r_t^{EA} - \varepsilon_t, \tag{17}$$

where $\varepsilon_t = \pi_t - E[\pi_t|I_{t-1}]$ is inflation forecast error, which is serially uncorrelated under the rational expectations assumption.

We assume that r_t^{EA} follows an AR(2) process with a regime-shifting mean, as given below:

$$\phi(L)(r_t^{EA} - \mu_{S_t}) = \nu_t, \tag{18}$$

where $\phi(L) = (1 - \phi_1 L - \phi_2 L^2)$; the roots of $\phi(L) = 0$ lie outside the complex unit circle; ν_t is serially uncorrelated with $E(\nu) = 0$; the subscript S_t refers to a latent regime-indicator variable. Then, by subtracting μ_{S_t} from both sides of equation (17) and multiplying both sides of the resulting equation by $\phi(L)$, it is straightforward to show that the resulting ex-post real interest rate follows an ARMA(2,2) process with a Markov-switching Mean, as given below:

$$r_t^{EP} = \mu_{S_t} + \phi_1(r_{t-1}^{EP} - \mu_{S_{t-1}}) + \phi_2(r_{t-2}^{EP} - \mu_{S_{t-2}}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, \tag{19}$$

where the roots of $(1 - \theta_1 L - \theta_2 L^2) = 0$ lie outside the complex unit circle. Following Garcia and Perron (1996), we further assume that the latent regime-indicator variable S_t follow a three-state, first-order Markov-switching process with the following transition probabilities:

$$Pr[S_t = j | S_{t-1} = i] = p_{ij}, \quad \sum_{j=1}^{3} p_{ij} = 1; \quad i, j = 1, 2, 3. \tag{20}$$

In order to complete the model by accommodating the heteroscedastic nature of the shocks to the ex-post real interest rate, we assume the following stochastic volatility for e_t: [8]

$$e_t \sim N(0, \sigma_t^2), \tag{21}$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2), \tag{22}$$

where ω_t is independent of e_t.

Given the above model, we construct the EARR series by taking a conditional expectation of the ex-post real interest rate:

[8] While Garcia and Perron (1996) assume a Markov-switching variance for e_t, we employ a random-walk stochastic volatility, which is much more flexible than a Markov-switching variance. In order to estimate the stochastic volatility, we implement the procedure proposed by Kim et al. (1998) in our MCMC algorithm.

$$E(r_t^{EP}|I_{t-1}) = E(\mu_{S_t}|I_{t-1}) + E(u_t|I_{t-1}), \tag{23}$$

where $u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$ and I_{t-1} refers to information up to time $t-1$, which consists of all the current and past history of ex-post real interest rate in the sample.

In this section, we employ the Bayesian econometric tool developed in Section 2, in estimating the above model for the U.S. ex-post real interest rate. We use quarterly data on ex-post real interest, which is constructed by subtracting the CPI inflation rate from the three-month Treasury bill rate. We extend Garcia and Perron's (1996) sample to cover recent observations right before the financial crisis, and thus our sample covers the period of 1960:I-2008:II. All the inferences are based on 25,000 Markov-Chain Monte Carlo (MCMC) outputs, after 5,000 burn-in's.

5.2. Empirical Results

We first estimate an AR(2) model by constraining $\theta_1 = \theta_2 = 0$, as in Garcia and Perron (1996). Both Garcia and Perron's sample (1960Q1-1986Q2) and our extended sample (1960Q1-2008Q2) are investigated. Table 4.A reports the posterior moments of the parameters for the Garcia and Perron sample. As in Garcia and Perron, once regime shifts in mean are taken account, the posterior mean of the sum of AR coefficients ($\phi_1 + \phi_2$) is close to zero, suggesting that persistence of the EARR is close to zero. Thus, the EARR may be regarded as a constant subject to occasional jumps caused by important structural events. For the extended sample, however, Table 4.B shows that the posterior mean of the sum of AR coefficients increases to 0.34 with the 90% highest posterior density (HPD) being [0.215,0.550]. (For a comparison of the posterior distributions of the sum of AR coefficients for the two samples, refer to Figure 4.B.) Figures 4.A.1 and 4.A.2 depict regime probabilities for the two samples. In the early 2000s, we have a decline in the EARR from a medium mean regime to a low mean regime. The estimated EARR's are plotted in Figure 4.C, which reveal that the variability of the EARR is higher for the extended sample.

However, ignoring the moving average terms in the ex-post real interest rate may result in misleading inference about the dynamics of the EARR. This is confirmed in Table 4.C, where we report the results for diagnostic checks. We perform white noise tests for the standardized prediction errors and their squares, as implied by the AR(2) model for ex-post real interest rate. Even though we cannot reject the null that they are white noise processes for the Garcia and Perron sample, the null is rejected at a 5% significance level for the extended sample. This evidence suggests that an AR(2) model with a Markov-switching mean for the ex-post real interest rate is misspecified for an extended sample period of 1960Q1-2008Q2.

When moving average (MA) terms are included for the Garcia and Perron sample (1960Q1-1986Q2), the posterior moments of the parameters reported in Table 5.A suggest that the results are almost the same as in the case of Garcia and Perron's (1996) AR(2) model. The posterior mean of the sum of AR coefficients, as well as that of the sum of MA coefficients, is close to zero. Furthermore, the regime probabilities (Figure 5.A.1), the posterior distribution of the sum of AR coefficients (Figure 5.B.1), the plot of EARR (the first panel of Figure 5.C), the measure of time-varying volatility (the first panel of Figure 5.D) are all very close to from those for an AR(2) model.

For the extended sample (1960Q1-2008Q2), however, the dynamics of the EARR implied by our ARMA(2,2) model are drastically different from those for an AR(2) model of Garcia and Perron (1996). The posterior median of the sum of AR coefficients is 0.732, with the 90% highest posterior Density (HPD) being [0.299,0.999]. Note that the posterior median of the AR coefficient sum in an AR(2) model is only 0.330 the 90% highest posterior Density (HPD) being [0.125,0.550]. If we compare the posterior distribution of the sum of AR coefficients for an AR(2) model (the second panel of Figure 4.B) and that for our ARMA(2,2) model (the second panel of Figure 5.B.1), the differences in the persistence dynamics of the EARR as implied by the two model are clearer. That is, omitting MA terms in the model of ex-post real interest rate considerably underestimates the persistence of the EARR for the extended sample. The plot of EARR in the lower panel of Figure 5.C show that EARR varies considerably within each regime, in contrast to the conclusion of Garcia and Perron (1996). Furthermore, for our ARMA(2,2) model, the results for diagnostic checks reported in Table 5.C suggest that we cannot reject the null hypothesis that the standardized prediction errors and their squares are white noise processes.

6. Summary and Conclusion

In this paper we provide an efficient MCMC algorithm for making inference of regime-switching ARMA models, based on a multi-move sampler. Our approach builds on the work of Billio et al. (1999), who propose an MCMC algorithm based on a single-move sampler. As discussed in Liu et al. (1994, 1995) and Scott (2002), one potential weakness of the algorithm based on a single-move sampler is that, its performance gets worse with slower mixing as the persistence of the latent state variable increases. However, our simulation study in Section 3 shows that the proposed algorithm based on a multi-move sampler achieves reasonably fast convergence to the posterior distribution, even when the latent regime indicator variable is highly persistent or even when there exist absorbing states.

We apply the proposed model and the algorithm to U.S. data on ex-post real interest rate (EPRR), in order to investigate the dynamics of the latent ex-ante real interest rate (EARR) under regime shifts. The rational expectations assumption implies the EPRR follows an ARMA process, if we assume that the latent EARR follows an AR process. We argue Garcia and Perron's (1996) conclusion that the EARR rate is a constant subject to occasional jumps may be sample-specific. For an extended sample that includes recent data, Garcia and Perron's (1996) AR(2) model of EPRR may be misspecified, and we show that excluding the theory-implied moving-average terms may understate the persistence of the EARR dynamics. Our empirical results suggest that, even though we rule out the possibility of a unit root in the EARR, it may be more persistent and volatile than has been documented in some of the literature including Garcia and Perron (1996).

Appendix A. Generating ARMA Parameters, Ψ, Conditional on MS states, $\tilde{S}_T$

Recursive data transformation schemes developed by Chib and Greenberg (1994) are introduced in this section, which produces simple linear regression relationships for μ, ϕ, and e_0. They successfully yield full conditional densities under a general ARMA(p,q) model and are employed for posterior Gibbs sampling. However, the posterior simulation of θ is.is complicated since its conditional posterior does not belong to standard families of distributions. Chib and Greenberg (1994) suggest suggest employing an MH algorithm for θ to successfully implement their Bayesian approach. While they provide a candidate density function for θ, which requires an additional estimation step, we, instead, utilize a random walk candidate density function. This particular class of MH algorithm with a random walk density is referred to as a random-walk chain Metropolis-Hastings algorithm. (see Koop (2003).) In the case of low acceptance probabilities, Chib and Greenberg(1994)'s algorithm can be employed as an alternative.

1. Generating Transition Probabilities conditional on $\tilde{Y}_T$, $\tilde{S}_T$, and other parameters

Assuming an independent Dirichlet distribution for the prior of $P_i = [p_{i1}\ p_{i2}\ ...\ p_{iM}]'$, the $i-th$ column of the matrix of the transition probabilities, P, we have:

$$Prior : P_i \sim Dirichlet(u_{i1}, u_{i2}, ..., u_{i,M}), \tag{A.1}$$

$$Posterior : P_i|\tilde{Y}_T, \tilde{S}_T, \Psi_{-q} \sim Dirichlet(u_{i1} + n_{i1}, u_{i2} + n_{i2}, ..., u_{iM} + n_{iM}),$$

where u_{ij} for $j = 1, 2, ..., M$, are known hyper parameters of the priors; n_{ij} refers to the number of the transitions from state i to j in $\tilde{S}_T$, which can be easily counted.

2. Generating ϕ conditional on $\tilde{Y}_T$, $\tilde{S}_T$, and other parameters $\Psi_{-\phi}$

The following is the necessary data transformation step for generating ϕ:

$$\bar{Y} = \bar{X}\phi + e, \tag{A.2}$$

$$\bar{y}_t = y_t - \mu_{S_t} - \Sigma_{j=1}^{q}\ \theta_i\ \bar{y}_{t-j}\ ,$$

$$\bar{x}_t = [\, \bar{y}_{t-1}\ \bar{y}_{t-2}\ ...\ \bar{y}_{t-p}\,]\,,$$

where $\bar{Y} = [\bar{y}_1\ \bar{y}_2\ ...\ \bar{y}_T]'$; $\bar{X} = [\bar{x}'_1\ \bar{x}'_2\ ...\ \bar{x}'_T]'$; $e = [e_1, e_2, ..., e_T]'$; $\bar{y}_t = 0$ for $t < 0$, $\bar{y}_0 = e_0$. The above derivation of data transformation can be easily shown by the fact that $e_t = \bar{y}_t - \bar{x}_t\phi$. The transformed data $\bar{Y}$ and $\bar{X}$ yield a desirable linear regression equation in terms of ϕ, which is employed for constructing the following conventional normal posterior:

$$Prior: \quad \phi \sim\ N(\underline{\phi}, \underline{\Phi})I_{\phi}, \tag{A.3}$$

$$Likelihood: \quad f(\tilde{Y}_T|\tilde{S}_T, \Psi_{-\phi}) = \prod_{t=1}^{T} \frac{1}{\sqrt(2\pi\sigma^2)}\ exp(-\frac{(\bar{y}_t - \bar{x}_t\phi)^2}{2\sigma^2}),$$

$$Posterior: \quad \phi|\tilde{Y}_T, \tilde{S}_T, \Psi_{-\phi} \sim\ N(\bar{\phi}, \bar{\Phi})I_{\phi},$$

where $\underline{\phi}$ and $\underline{\Phi}$ are a prior mean and a prior variance, respectively; I_{ϕ} is an indication function for stationarity; $\bar{\phi} = \bar{\Phi}(\underline{\Phi}^{-1}\underline{\phi} + \sigma^{-2}\bar{X}'\bar{Y})$ and $\bar{\Phi}(\underline{\Phi}^{-1} + \sigma^{-2}\bar{X}'\bar{X})^{-1}$, which are a posterior mean and a posterior variance, respectively.

3. Generating μ conditional on $\tilde{Y}_T$, $\tilde{S}_T$, and other parameters $\Psi_{-\mu}$

First, we show recursive data transformations for generating μ:

$$Y^* = X^*\mu + e, \tag{A.4}$$

$$y^*_t = y_t - \Sigma^p_{i=1}\phi_i\ y_{t-i} - \Sigma^q_{j=1}\theta_i\ y^*_{t-j},$$

$$x^*_t = x_t - \Sigma^p_{i=1}\phi_i\ x_{t-i} - \Sigma^q_{j=1}\theta_j\ x^*_{t-j},$$

where $Y^* = [y^*_1\ y^*_2\ ...\ y^*_T]'$; $X^* = [x'^*_1\ x'^*_2\ ...\ x'^*_T]'$; $e = [e_1, e_2, ..., e_T]'$; $x_t = [I_{S_t=1}\ I_{S_t=2}\ ...\ I_{S_t=M}]$ and I_{S_t} is an indication function of each MS state; $y_t = y^*_t = 0$ for $t < 0$ and $y_0 = y^*_0 = e_0$; the vectors $x_t = x^*_t = 0$ for $t \leq 0$. The above derivation of data transformation can be easily shown by the fact that $e_t = y^*_t - x^*_t\mu$.

The generated data sets, Y^* and X^* have a conventional linear regression relationship as well. Therefore, the prior and the posterior densities of μ are given by:

$$Prior: \quad \mu \sim N(\underline{\mu}, \underline{\Omega}_{\mu})I_{\mu}, \tag{A.5}$$

$$Likelihood: \quad f(\tilde{Y}_T|\tilde{S}_T, \Psi_{-\mu}) = \prod_{t=1}^{T} \frac{1}{\sqrt(2\pi\sigma^2)}\ exp(-\frac{(y^*_t - x^*_t\mu)^2}{2\sigma^2}),$$

$$Posterior: \quad \mu|\tilde{Y}_T, \tilde{S}_T, \Psi_{-\mu} \sim N(\bar{\mu}, \bar{\Omega}_\mu) I_\mu,$$

where $\underline{\mu}$ and $\underline{\Omega}_\mu$ are a prior mean and a prior variance, respectively; I_μ is the indication function for identification of MS regimes; $\bar{\mu} = \bar{\Omega}_\mu(\underline{\Omega}_\mu{}^{-1}\underline{\mu} + \sigma^{-2}X^{*\prime}Y^*)$ and $\bar{\Omega}_\mu = (\underline{\Omega}_\mu^{-1} + \sigma^{-2}X^{*\prime}X^*)^{-1}$, which are a posterior mean and a posterior covariance matrix, respectively.

4. Generating θ conditional on $\tilde{Y}_T$, $\tilde{S}_T$, and other parameters $\Psi_{-\theta}$

In order to generate θ, the MH algorithm is inevitable as the error term, e_t, is not a linear function of θ. Chib and Greenberg (1994) suggested a candidate density of θ based on the first-order Taylor expansion and the non-linear least-squares estimation which requires additional classical estimation and data transformation steps. We, instead, take advantage of a random walk chain MH as an alternative to simplify these steps. (See Koop (2003).) In the procedure, a candidate density is defined as:

$$\theta^* = \theta^{m-1} + \varepsilon \tag{A.6}$$

where θ^* is a new candidate sample; θ^{m-1} is a previously accepted θ in the previous MCMC iteration; ε is an increment random variable. The corresponding acceptance probability is given by:

$$\alpha(\theta^*, \theta^{m-1}) = min[\ \frac{f[\theta^*|\tilde{S}_T, \tilde{Y}_T, \Psi_{-\theta}]}{f[\theta^{m-1}|\tilde{S}, \tilde{Y}, \Psi_{-\theta}]}\ ,\ 1\ ] \tag{A.7}$$

where $f[\theta|\tilde{Y}_T, \tilde{S}_T, \Psi_{-\theta}]$ is the conditional posterior density of θ. Note that a choice of density for ε completes the candidate density. We take a common choice of ε which is a multivariate normal with mean 0 and a variance-covariance, Σ_c. Σ_c^2 is appropriately chosen to get an acceptance probability between 0.2 and 0.5 which is the range advocated by Koop (2003).

The posterior simulation on θ is conducted with the candidate generating function in equation (A.6), where the prior and the posterior are given by:

$$Prior: \ \theta \sim N(\underline{\theta}, \underline{\Omega}_\theta) I_\theta, \tag{A.8}$$

$$Posterior: \ \theta|\tilde{Y}_T, \tilde{S}_T, \Psi_{-\theta} \ \propto \ \prod_{t=1}^{T} \ exp[-\frac{1}{2\sigma^2}\ e_t(\theta)^2] \ \times \ exp[-\frac{1}{2}(\theta - \underline{\theta})'\underline{\Omega}_\theta{}^{-1}(\theta - \underline{\theta})] I_\theta,$$

where $\underline{\theta}$ and $\underline{\Omega}_\theta$ are a prior mean and a prior variance, respectively; I_θ is the indication function for invertibility; $\bar{\theta}$ and $\bar{\Omega}_\theta$ are a posterior mean and a posterior variance; $e_t(\theta) = (y_t - \mu_{S_t}) - \phi(y_{t-1} - \mu_{S_{t-1}}) - \theta e_{t-1} = \bar{y}_t - \bar{x}_t\phi = y_t^* - x_t^*\mu$, which can be obtained in the preceding transformations.

5. Generating e_0 conditional on $\tilde{Y}_T$, $\tilde{S}_T$, and other parameters Ψ_{-e_0}

Chib and Greenberg (1994) proposed a method to estimate e_0 based on the Kalman filter and the backward recursions with the Moore-Penrose inverse. We follow an efficient alternative by Nakatsuma (2000) to avoid the complexity. The following is the required data transformation step which generates a simple linear regression equation as other parameters:

$$\hat{Y} = \hat{X}e_0 + e, \tag{A.9}$$

$$\hat{y}_t = y_t - \mu_{S_t} - \Sigma_{i=1}^{p}\phi_i\ (y_{t-i} - \mu_{S_{t-i}}) - \Sigma_{j=1}^{q}\ \theta_j\ \hat{y}_{t-j},$$

$$\hat{x}_t = (\phi_t + \theta_t) - \Sigma_{j=1}^{q}\ \theta_j\ \hat{x}_{t-j},$$

where $\hat{Y} = [\hat{y}_1\ \hat{y}_2\ ...\ \hat{y}_T]'$; $\hat{X} = [\hat{x}_1\ \hat{x}_2\ ...\ \hat{x}_T]'$; $(y_t - \mu_{S_t}) = \hat{y}_t = \hat{x}_t = 0$ for $t \leq 0$; $\phi_t = 0$ for $t > p$ and $\theta_t = 0$ for $t > q$. The above derivation of data transformation can be easily shown by the fact that $e_t = \hat{y}_t - \hat{x}_t e_0$.

The generated data have a conventional linear regression relationship conditional on Ψ_{-e_0}. Therefore, it is now straightforward to draw e_0 from the following conditional posterior density:

$$Prior:\ e_0 \sim N(\underline{e}_0, \underline{\Omega}_{e_0}), \tag{A.10}$$

$$Likelihood:\quad f(\tilde{Y}_T|\tilde{S}_T, \Psi_{-e_0}) = \prod_{t=1}^{T} \frac{1}{\sqrt{(2\pi\sigma^2)}}\ exp(-\frac{(\hat{y}_t - \hat{x}_t e_0)^2}{2\sigma^2}),$$

$$Posterior:\ e_0|\tilde{Y}_T, \tilde{S}_T, \Psi_{-e_0} \sim\ N(\bar{e}_0, \bar{\Omega}_{e_0}),$$

where $\underline{e}_0$ and $\underline{\Omega}_{e_0}$ are a prior mean and a prior variance, respectively; $\bar{e}_0$ and $\bar{\Omega}_{e_0}$ are a posterior mean and a posterior variance; $\bar{e_0} = \bar{\Omega}_{e_0}(\underline{\Omega}_{e_0}{}^{-1}\underline{e_0} + \sigma^{-2}\hat{X}'\hat{Y})$; $\bar{\Omega}_{e_0} = (\underline{\Omega}_{e_0}^{-1} + \sigma^{-2}\hat{X}'\hat{X})^{-1}$.

6. Generating σ^2 conditional on $\tilde{Y}$, $\tilde{S}$, and all the other parameters $\Psi_{-\sigma^2}$

The posterior simulation on σ^2 is straightforward given one of the previously transformed data sets. The posterior samples on σ^2 are drawn from the following conditional posterior density:

$$Prior: \ \sigma^2 \sim IG(\frac{\nu}{2}, \frac{\delta}{2}), \tag{A.11}$$

$$Posterior: \ \sigma^2 | \tilde{Y}_T, \tilde{S}_T, \Psi_{-\sigma^2} \sim IG(\frac{\bar{\nu}}{2}, \frac{\bar{\delta}}{2}),$$

where $\underline{\nu}$ and $\underline{\delta}$ are a prior degree of freedom and a prior scale parameter, respectively; $\bar{\nu} = \underline{\nu} + T$; $\bar{\delta} = \underline{\delta} + d$ where $d = \prod_{t=0}^{T} \ e_t^2 = \prod_{t=0}^{T}(\bar{y}_t - \bar{x}_t\phi)^2$. Note that alternatively, other transformed data sets (Y^*, X^*) or $(\hat{Y}, \hat{X})$ can be used to calculate d. While different choices of how to calculate d would lead to slightly different values of d, this would not make significant differences on the Bayesian estimates. This step completes the MCMC algorithm of an ARMA (p,q) model with a Marokov-switching mean conditional on $\tilde{S}$.

Appendix B. Generating $\tilde{S}_T$ from the Proposed Multi-move Candidate Density

1. State Space Representation of ARMA(p,q) Models

Consider the following general MS-ARMA (p,q) model:

$$y_t = \mu_{S_t} \; + \sum_{i=1}^{p} \phi_{i,S_t} \; (y_{t-i} - \mu_{S_{t-i}}) \; + \; e_t \; + \sum_{j=1}^{q} \theta_{i,S_t} \; e_{t-j}, \quad t = 1, 2, ..., T \qquad (B.1)$$

where e_t follows independent $N(0, \sigma^2_{S_t})$. The parameters of the ARMA (p,q) model are dependent upon M discrete unobserved states ($S_t = 1, 2, ..., M$) at each time period. We assume stationarity and invertibility under all regimes. The ARMA (p,q) model of equation (B.1) can be equivalently expressed as the following:

$$\begin{aligned} y_t &= \mu_{S_t} + u_t, \\ u_t &= \sum_{i=1}^{p} \phi_i \; u_{t-i} \; + \; e_t \; + \sum_{i=1}^{q} \theta_i \; e_{t-j}. \end{aligned} \qquad (B.2)$$

Finally, equation (B.2) has the following state-space representation:

Measurement Equation

$$y_t = \mu_{S_t} + H\alpha_t, \qquad (B.3)$$

where $H = [\,1 \; 0 \; ... \; 0\,]$ is a $1 \times (p+q)$ matrix;

Transition Equation

$$\begin{bmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t-p+1} \\ e_t \\ e_{t-1} \\ \vdots \\ e_{t-q+1} \end{bmatrix} = \begin{bmatrix} \phi_{S_t,1} & \phi_{S_t,2} & \dots & \phi_{S_t,p} & \theta_{S_t,1} & \theta_{S_t,2} & \dots & \theta_{S_t,q} \\ & & & 0 & 0 & 0 & \dots & 0 \\ & I_{p-1} & & \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & & & & 0 \\ \vdots & \ddots & \vdots & \vdots & & I_{q-1} & & \vdots \\ 0 & \dots & 0 & 0 & & & & 0 \end{bmatrix} \begin{bmatrix} u_{t-1} \\ u_{t-2} \\ \vdots \\ u_{t-p} \\ e_{t-1} \\ e_{t-2} \\ \vdots \\ e_{t-q} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} e_t, \qquad (B.4)$$

$$(\alpha_t = F\alpha_{t-1} + Ge_t,)$$

where the model parameters in $\mu_{S_t}, F_{S_t}, \sigma^2_{S_t}$ are dependent upon an unobserved, discrete-valued, M state Markov-switching variable S_t ($S_t = 1, 2, ..., M$); I_m is a $m \times m$ identity matrix. Transition probabilities are given in equation (1.2).

2. Conditional Kalman Filter

Here in the state-space model with Markov switching, we need to use the conventional Kalman filter conditional on $S_{t-1} = i$ and $S_t = j$ for i,j = 1,2,...,M which is given by:

$$\alpha^{(i,j)}_{t|t-1} = F_j \alpha^{(i)}_{t-1|t-1}, \qquad (B.5)$$

$$P^{(i,j)}_{t|t-1} = F_j P^{(i)}_{t-1|t-1} F_j' + GG' \sigma^2_j, \qquad (B.6)$$

$$\eta^{(i,j)}_{t|t-1} = y_t - \mu_j - H\alpha^{(i,j)}_{t|t-1}, \qquad (B.7)$$

$$f^{(i,j)}_{t|t-1} = HP^{(i,j)}_{t|t-1} H', \qquad (B.8)$$

$$\beta^{(i,j)}_{t|t} = \beta^{(i,j)}_{t|t-1} + P^{(i,j)}_{t|t-1} H' [f^{(i,j)}_{t|t-1}]^{-1} \eta^{(i,j)}_{t|t-1}, \qquad (B.9)$$

$$P^{(i,j)}_{t|t} = (I - P^{(i,j)}_{t|t-1} H' [f^{(i,j)}_{t|t-1}]^{-1} H) P^{(i,j)}_{t|t-1}, \qquad (B.10)$$

where $\alpha^{(i)}_{t-1|t-1}$ is an inference on α_{t-1} based on $\tilde{Y}_{t-1}$ and given $S_{t-1} = i$; $\alpha^{(i,j)}_{t|t-1}$ is an inference on α_t based on $\tilde{Y}_{t-1}$ and given $S_t = j$ and $S_{t-1} = i$; $P^{(i,j)}_{t|t-1}$ is the mean squared error matrix of $\alpha^{(i,j)}_{t|t-1}$ conditional on $S_t = j$ and $S_{t-1} = i$; $\eta^{(i,j)}_{t|t-1}$ is the conditional forecast error of y_t based on $\tilde{Y}_{t-1}$, given $S_{t-1} = i$ and $S_t = j$; and $f^{(i,j)}_{t|t-1}$ is the conditional variance of forecast error $\eta^{(i,j)}_{t|t-1}$.

However, notice that each iteration of the above Kalman filter produces an M-fold increase in the number of cases to consider. For example, M^{10} cases should be considered by the time, t = 10. As a result, without some approximation, the above Kalman filter is not operable. Kim (1994) proposed a algorithm to complete the above Kalman filter by collapsing ($M \times M$) posteriors ($\alpha^{(i,j)}_{t|t}$, $P^{(i,j)}_{t|t}$) into M posteriors ($\alpha^{(j)}_{t|t}$, $P^{(j)}_{t|t}$).

3. Approximated Filtering Algorithm by Kim (1994)

The approximated filtering algorithm by Kim (1994) is a combination of extended versions of the Kalman filter and the Hamilton filter, along with appropriate approximations. It starts with initial values $\alpha_{0|0}^{(j)}$, $P_{0|0}^{(j)}$ and $f[S_0]$, which are the unconditional mean and covariance matrix of the unobserved process of α_t conditional on state $S_0 = j$ and the steady state probability of the Markov chain process, respectively. The filtering algorithm contains the following steps:

1) Run the Kalman filter given in equation (B.5)-(B.10) for $i, j = 1, 2, ..., M$ to get the followings:

$$\alpha_{t|t-1}^{(i,j)}, P_{t|t-1}^{(i,j)}, \eta_{t|t-1}^{(i,j)}, f_{t|t-1}^{(i,j)}, \alpha_{t|t}^{(i,j)}, P_{t|t}^{(i,j)} \qquad (B.11)$$

2) Calculate $Pr[S_t = j|\tilde{Y}_t]$ for $j = 1, 2, ..., M$ through the following Hamilton filter:

$$f(S_t, S_{t-1}|\tilde{Y}_{t-1}) = f(S_t|S_{t-1})f(S_{t-1}|\tilde{Y}_{t-1}) \qquad (B.12)$$

$$f(y_t|\tilde{Y}_{t-1}) = \sum_{S_t} \sum_{S_{t-1}} f(y_t|\tilde{Y}_{t-1}, S_t, S_{t-1})f(S_t, S_{t-1}|\tilde{Y}_{t-1}) \qquad (B.13)$$

$$f(S_t, S_{t-1}|\tilde{Y}_t) = \frac{f(y_t, S_t, S_{t-1}|\tilde{Y}_{t-1})}{f(y_t|\tilde{Y}_{t-1})} = \frac{f(y_t|S_t, S_{t-1}, \tilde{Y}_{t-1})f(S_t, S_{t-1}|\tilde{Y}_{t-1})}{f(y_t|\tilde{Y}_{t-1})} \qquad (B.14)$$

$$Pr[S_t = j|\tilde{Y}_t] = \sum_{S_{t-1}} f(S_t = j, S_{t-1}|\tilde{Y}_t) \qquad (B.15)$$

3) Using the conditional probabilities from the Hamilton filter, collapse $M \times M$ posteriors in equations (B.9) and (B.10) into M posteriors using the following approximations:

$$\alpha_{t|t}^{(j)} \approx \frac{\sum_{S_{t-1}} f(S_t = j, S_{t-1}|\tilde{Y}_t)\ \alpha_{t|t}^{(i,j)}}{Pr[S_t = j|\tilde{Y}_t]} \qquad (B.16)$$

$$P_{t|t}^{(j)} \approx \frac{\sum_{S_{t-1}} f(S_t = j, S_{t-1}|\tilde{Y}_t)\ [\ P_{t|t}^{(i,j)} + (\alpha_{t|t}^{(j)} - \alpha_{t|t}^{(i,j)})(\alpha_{t|t}^{(j)} - \alpha_{t|t}^{(i,j)})\ ]}{Pr[S_t = j|\tilde{Y}_t]} \qquad (B.17)$$

4) Repeat step1-step3 and save approximated $Pr[S_t = j|\tilde{Y}_t]$ for $j = 1, 2, .., M$ in each iteration for Bayesian inference on $\tilde{S}_T$.

The resulting $Pr[S_t|\tilde{Y}_t]$ for t = 1,2,...,T with the above approximations is used as $h(S_t|\tilde{Y}_t)$ for the candidate density function in equation (12).

References

[1] Albert, J.H. and S. Chib, 1993, Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics,* 11, 1-15.

[2] Antoncic, Madelyn, 1986, High and Volatile Real Interest Rates: Where Does the Fed Fit In? *Journal of Money, Credit, and Banking,* 18(1), 18-27.

[3] Bai, Jushan and Pierre Perron, 2003, Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics,* 18(1), 1-22.

[4] Billio, M., A. Monfort and C.P. Robert, 1999, Bayesian estimation of switching ARMA models. *Journal of Econometrics,* 93, 229-255.

[5] Caporale, Tony and Kevin B. Grier, 2000, Political Regime Change and the Real Interest Rate. *Journal of Money, Credit, and Banking,* 32(3), 320-34.

[6] Caporale, Tony, and Kevin B. Grier, 2005, Inflation, presidents, Fed Chairs and regime shifts in the US real interest rate. *Journal of Money, Credit, and Banking,* 37, 1153?163.

[7] Carter, C. K. and P. Kohn, 1994, On Gibbs sampling for state space models. *Biometrica* 81, 541-553.

[8] Chib, S., 1998, Estimation and comparison of multiple change-point models. *Journal of Econometrics,* 86, 221-241.

[9] Chib, S., and E. Greenberg, 1994, Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics,* 64, 183-206.

[10] Chib, S., and E. Greenberg, 1995, Understanding the Metropolis-Hastings Algorithm. *The American Statistician,* 49, 327-335.

[11] Cosslett, S.R. and L.-F. Lee, 1985, Serial correlation in latent discrete variable models. *Journal of Econometrics 27*, 79-97.

[12] Crowder, William J. and Dennis L. Hoffman, 1996, The Long-Run Relationship Between Nominal Interest Rates and Inflation: The Fisher Equation Revisited. *Journal of Money, Credit, and Banking,* 28(1), 102-18.

[13] De Jong, P. and N. Shephard, 1995, The simulation smoother for time series models. *Biometrika,* 82, 339-50.

[14] Diebolt, J. and Robert C.P., 1994, Estimation of finite mixture distributions by Bayesian sampling. *Journal of the Royal Statistical Society (Series B),* 56, 363-365.

[15] Fama, Eugene F., 1975, Short-Term Interest Rates as Predictors of Inflation. *American Economic Review,* 63(3), 269-82.

[16] Gali, Jordi, 1992, How Well Does the IS-LM Model Fit Postwar U.S. Data? *Quarterly Journal of Economics,* 107(2), 709-38.

[17] Garbade, Kenneth, and Paul Wachtel, 1978, Time Variation in the Relationship between Inflation and Interest Rates.*Journal of Monetary Economics*, 4, 755-765.

[18] Garcia, Rene and Pierre Perron, 1996, An Analysis of the Real Interest Rate Under Regime Shifts. *Review of Economics and Statistics,* 78(1), 111-25.

[19] Hamilton, J.D., 1988, Rational expectations econometric analysis of changes in regimes: An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control,* 12, 385-432.

[20] Hamilton, J.D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica,* 57, 357-84.

[21] Harrison, P.J. and C.F. Stevens, 1976, Bayesian forecasting. *Journal of the Royal Statistical Society B,* 38, 205-247.

[22] Huizinga, John and Frederic S. Mishkin, 1986, Monetary Regime Shifts and the Unusual Behavior of Real Interest Rates. *Carnegie-Rochester Conference Series on Public Policy*, 24, 231-74.

[23] Kim, C.-J., 1994, Dynamic linear models with Markov-Switching. *Journal of Econometrics,* 60, 1-22.

[24] Kim, S., N. Shepard and S. Chib, 1998, Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies,* 65, 361-93.

[25] King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson, 1991, Stochastic Trends and Economic Fluctuations, *American Economic Review,* 81(4), 819-40.

[26] Koop, G., 2003, Bayesian econometrics. *John Wiley and Sons.*

[27] Koustas, Zisimos and Apostolos Serletis, 1999, On the Fisher Effect. *Journal of Monetary Economics,* 44(1), 105-30.

[28] Liu, J.S., W.H. Wong and A. Kong, 1994, Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika,* 81, 27-40.

[29] Mishkin, Frederic S., 1981, The Real Rate of Interest: An Empirical Investigation. *CarnegieRochester Conference Series on Public Policy,* 15, 151-200.

[30] Mishkin, Frederic S., 1992, Is the Fisher Effect for Real? A Reexamination of the Relationship Between Inflation and Interest Rates. *Journal of Monetary Economics,* 30(2), 195-215.

[31] Nakatsuma, T., 2000, Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach. *Journal of Econometrics,* 95, 57-69.

[32] Neely, C. J. and D. E. Rapach, 2008, Real interest rate persistence: evidence and implications. *Working Paper 2008-018A, Federal Reserve Bank, St. Louis.*

[33] Nelson, Charles R., and G. William Schwert, 1977, Short-Term Interest Rates as Predictors of Inflation: On Testing the Hypothesis that the Real Interest Rate is Constant. *American Economic Review,* 67(3), 478-86.

[34] Perron, Pierre, 1990, Testing for a Unit Root in a Time Series with a Changing Mean. *Journal of Business and Economic Statistics,* 8(2), 153-162.

[35] Rapach, David E., and Christian E. Weber, 2004, Are Real Interest Rates Really Nonstationary? New Evidence from Tests with Good Size and Power. *Journal of Macroeconomics,* 26(3), 409-30.

[36] Rose, Andrew K., 1988, Is the Real Interest Rate Stable? *Journal of Finance*, 43(5), 1095-112.

[37] Scott, S.L., 2002, Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century. *Journal of the American Statistical Association,* 97, 337-351.

[38] Shephard, N., 1994, Partial non-Gaussian state space. *Biometrika,* 81, 115-131.

[39] Sun, Yixiao, and Peter C.B. Phillips, 2004, Understanding the Fisher Equation. *Journal of Applied Econometrics,* 19(7), 869-86.