Essays on Bayesian Econometrics

Jaeho Kim

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2014

Reading Committee:
Chang-Jin Kim, Chair
Charles R Nelson
Richard Startz

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Essays on Bayesian Econometrics

Jaeho Kim

Chair of the supervisor committee:

Professor Chang-Jin Kim

Department of Economics

This dissertation explores important macroeconomics issues based on Bayesian Econometrics tools developed. One goal of the first chapter of the dissertation 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 of the first chapter 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).

The second chapter of the dissertation investigates the conventional wisdom that in the case of a flat prior Bayesian inference will not be very different from classical inference, as the

likelihood dominates the posterior density. This chapter shows that there are cases in which this conventional wisdom does not apply. An ARMA model of real GDP growth estimated by Perron and Wada (2009) is an example. While their maximum likelihood estimation of the model implies that real GDP may be a trend stationary process, Bayesian estimation of the same model implies that most of the variations in real GDP can be explained by the stochastic trend component, as in Nelson and Plosser (1982) and Morley et al. (2003). We show such dramatically different results stem from the differences in how the nuisance parameters are handled between the two approaches, especially when the parameter estimate of interest is dependent upon the estimates of the nuisance parameters for small samples. For the maximum likelihood approach, as the number of the nuisance parameters increases, we have higher probability that the moving-average root may be estimated to be one even when its true value is less than one, spuriously indicating that the data is 'over-differenced.' However, the Bayesian approach is relatively free from this pile-up problem, as the posterior distribution is not dependent upon the nuisance parameters.

The last chapter of the dissertation is about Bayesian model comparison. Bayesian model comparison is often achieved by the *Bayes Factor* which is sensitive to prior assumptions. Various alternative *Bayes Factors* such as the *Intrinsic* and *Fractional Bayes Factors* have been proposed to overcome this problem. However, practical problems arise since they include many marginal likelihoods which are not analytically tractable in most cases. An encompassing prior approach (EP) is a recently proposed method to approximate the *Bayes Factor* numerically in comparing nested models. We extend EP approach to the alternative *Bayes Factors* in this chapter. Our method provides a simple and elegant way to conduct robust model comparisons to the prior sensitivity for nested models.

Table of Contents

Chapter 1: Bayesian Inference in Regime-Switching ARMA Models with Absorbing States: The Dynamics of the Ex-Ante Real Interest Rate Under Structural Breaks

Chapter 2: The ‘Pile-up Problem’ in Trend-Cycle Decomposition of Real GDP: Classical and Bayesian Perspectives

Chapter 3: Robust Bayesian Model Selection for Nested Models

List of Tables

List of Figures

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my supervisory committee members: Professor Chang-Jin Kim, Charles Nelson, Richard Startz, Eric Zivot, and Yu-Chin Chen. Especially, I owe my deepest gratitude to professor Chang-Jin Kim for his generous guidance and endless support. This dissertation would never have been accomplished without his invaluable advice.

My wife, Sora Chon, steadfastly supported and encouraged me. She was always with me when I needed her the most. I will never forget our long walks and conversations. Thanks for being my wife and friend. I hope to return her love and favor when she pursues her own Ph.D. degree. Without my mother, Chaemae Ko, my pursuit of this advanced degree would never have been started. Her support, encouragement, and constant love have sustained me throughout my life.

This dissertation is lovingly dedicated to my wife, Sora Chon.

Chapter 1

Bayesian Inference in Regime-Switching ARMA Models with Absorbing States: The Dynamics of the Ex-Ante Real Interest Rate Under Structural Breaks [1]

1.1. 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

---

[1] This chapter is based on a joint work with Chang-Jin Kim.

[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.

(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 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 chapter 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 chapter 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 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 chapter is organized as follows. Section 1.2 present our benchmark econometric model and provides a literature review on the inference of regime-switching

---

[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.

ARMA models. Section 1.3 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 1.4, 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 1.5, the benchmark model in Section 1.2 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 1.6 provides a summary and concluding remarks.

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

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

$$y_t = \mu_{S_t} + \sum_{i=1}^{p} \phi_{i,S_t}(y_{t-i} - \mu_{S_{t-i}}) + e_t - \sum_{j=1}^{q} \theta_{i,S_t} e_{t-j}, \quad e_t \ \sim i.i.d.N(0, \sigma^2), \tag{1.1}$$

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:

$$Pr[S_t = j | S_{t-1} = i] = p_{ij}, \quad \sum_{j=1}^{M} p_{ij} = 1, \quad i, j = 1, 2, ..., M. \tag{1.2}$$

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 1.5, an extended version of this model is applied to the ex-post real interest rate. To deal with the non-i.i.d. nature

---

[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).

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 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

---

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

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

$$\begin{aligned} F(\tilde{S}_T|\tilde{Y}_T) &= \frac{f(\tilde{S}_T)\ f(\tilde{Y}_T|\tilde{S}_T)}{f(\tilde{Y}_T)} \\ &= \frac{f(S_0)\ \prod_{t=1}^{T}\ f(S_t|S_{t-1})\ \prod_{t=1}^{T}\ f(y_t|\tilde{S}_t, \tilde{Y}_{t-1})}{f(\tilde{Y}_T)}. \end{aligned} \tag{1.3}$$

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

$$\begin{aligned} f(S_t|\tilde{S}_{\neq t}, \tilde{Y}_T) &= \frac{f(S_t|\tilde{S}_{\neq t})\ f(\tilde{Y}_T|\tilde{S}_T)}{f(\tilde{Y}_T|\tilde{S}_{\neq t})} \\ &\propto f(S_t|\tilde{S}_{\neq t})\ f(\tilde{Y}_T|\tilde{S}_T) \\ &\propto f(S_{t+1}|S_t)\ f(S_t|S_{t-1})\ \prod_{t=1}^{T} f(y_t|\tilde{S}_T, \tilde{Y}_{t-1}) \\ &= f(S_{t+1}|S_t)\ f(S_t|S_{t-1})\ \prod_{t=1}^{T} f(y_t|\tilde{S}_t, \tilde{Y}_{t-1}) \\ &\propto f(S_{t+1}|S_t)\ f(S_t|S_{t-1})\ \prod_{k=t}^{T} f(y_k|\tilde{S}_k, \tilde{Y}_{k-1}), \end{aligned} \tag{1.4}$$

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 (1.4) 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 (1.4) for $t = 0, 1, 2, ..., T$, they propose to generate it from the following individual proposal density:

$$g(S_0|\tilde{S}_{\neq 0}, \tilde{Y}) \propto f(S_1|S_0) f(S_0), \quad for \quad t = 0, \tag{1.5}$$

$$g(S_t|\tilde{S}_{\neq t}, \tilde{Y}) \propto f(S_{t+1}|S_t)\ f(S_t|S_{t-1})\ f(y_t|\tilde{S}_t, \tilde{Y}_{t-1}), \quad for \quad t = 1, ..., T-1. \tag{1.6}$$

$$g(S_T|\tilde{S}_{\neq T}, \tilde{Y}) \propto f(S_T|S_{T-1})\ f(y_T|\tilde{S}_T, \tilde{Y}_{T-1}), \quad for \quad t = T, \tag{1.7}$$

which is an approximation to the individual target density in equation (1.4). 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 (1.4).

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:

$$\alpha(\tilde{S}_T^J, \tilde{S}_T^{J-1}) = \ min[\frac{F(\tilde{S}_T^J|\tilde{Y}_T)}{F(\tilde{S}_T^{J-1}|\tilde{Y}_T)} \frac{G(\tilde{S}_T^{J-1}|\tilde{Y}_T)}{G(\tilde{S}_T^J|\tilde{Y}_T)},\ 1], \tag{1.8}$$

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

$$G(\tilde{S}_T|\tilde{Y}_T) = \prod_{t=0}^{T} \left[\frac{g(S_t|\tilde{S}_{\neq t}, \tilde{Y}_t)}{\sum_{S_t=1}^{M} g(S_t|\tilde{S}_{\neq t}, \tilde{Y}_t)} \right]. \tag{1.9}$$

By substituting equations (1.3) and (1.9) into (1.8) and rearranging terms, Billio et al. (1999) derive the following acceptance probability:

$$\alpha(\tilde{S}_T^J, \tilde{S}_T^{J-1}) = min \left[\prod_{t=1}^{T} \frac{f(S_t^{J-1}|S_{t-1}^{J-1})}{f(S_t^J|S_{t-1}^J)} \prod_{t=0}^{T} \frac{\sum_{S_t} f(S_{t+1}^J|S_t) f(S_t|S_{t-1}^J) f(y_t|\tilde{S}_t^J, \tilde{Y}_{t-1})}{\sum_{S_t} f(S_{t+1}^{J-1}|S_t) f(S_t|S_{t-1}^{J-1}) f(y_t|\tilde{S}_t^{J-1}, \tilde{Y}_{t-1})},\ 1 \right]. \tag{1.8$'$}$$

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 1.3 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.

1.3. 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 $F(\tilde{S}_T|\tilde{Y}_T)$:

$$F(\tilde{S}_T|\tilde{Y}_T) = f(S_T|\tilde{Y}_T)\prod_{t=0}^{T-1} f(S_t|\tilde{S}_{t+1:T},\tilde{Y}_T), \tag{1.10}$$

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

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:

$$\begin{aligned} f(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T) &= f(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_t, \tilde{Y}_{t+1:T}) \\ &= \frac{f(S_t, \tilde{Y}_{t+1:T}|\tilde{S}_{t+1:T}, \tilde{Y}_t)}{f(\tilde{Y}_{t+1:T}|\tilde{S}_{t+1:T}, \tilde{Y}_t)} \\ &\propto f(S_t, \tilde{Y}_{t+1:T}|\tilde{S}_{t+1:T}, \tilde{Y}_t) \\ &= f(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_t)\ f(\tilde{Y}_{t+1:T}|\tilde{S}_{t:T}, \tilde{Y}_t) \\ &\propto f(S_{t+1}|S_t) f(S_t|\tilde{Y}_t) \prod_{k=t+1}^{T} f(y_k|\tilde{S}_{t:k}, \tilde{Y}_{k-1}). \end{aligned} \tag{1.11}$$

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 (1.11):

$$g(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T) \propto f(S_{t+1}|S_t) h(S_t|\tilde{Y}_t), \tag{1.12}$$

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 (1.11). 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 (1.11).

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 (1.12), 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:

$$\alpha(\tilde{S}_T^J, \tilde{S}_T^{J-1}) = \; min\left[\frac{F(\tilde{S}_T^J|\tilde{Y}_T)}{F(\tilde{S}_T^{J-1}|\tilde{Y}_T)}\frac{G(\tilde{S}_T^{J-1}|\tilde{Y}_T)}{G(\tilde{S}_T^J|\tilde{Y}_T)}, 1\right], \tag{1.13}$$

where $F(.|\tilde{Y}_T)$ is given in equation (1.3), as rewritten below:.

$$\begin{aligned} F(\tilde{S}_T|\tilde{Y}_T) &= \frac{f(\tilde{S}_T)\; f(\tilde{Y}_T|\tilde{S}_T)}{f(\tilde{Y}_T)} \\ &= \frac{f(S_0)\; \prod_{t=1}^T \; f(S_t|S_{t-1})\; \prod_{t=1}^T \; f(y_t|\tilde{S}_t, \tilde{Y}_{t-1})}{f(\tilde{Y}_T)}, \end{aligned} \tag{1.3}$$

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

$$\begin{aligned} G(\tilde{S}_T|\tilde{Y}_T) &= \prod_{t=0}^{T}\left[\frac{g(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T)}{\sum_{S_t} g(S_t|\tilde{S}_{t+1:T}, \tilde{Y}_T)}\right] \\ &= \prod_{t=0}^{T}\left[\frac{f(S_{t+1}|S_t)h(S_t|\tilde{Y}_t)}{\sum_{S_t} f(S_{t+1}|S_t)h(S_t|\tilde{Y}_t)}\right]. \\ &= \prod_{t=0}^{T}\left[\frac{f(S_{t+1}|S_t)h(S_t|\tilde{Y}_t)}{h(S_{t+1}|\tilde{Y}_t)}\right] \end{aligned} \tag{1.14}$$

By substituting equations (1.3) and (1.14) into equation (1.13), we can derive the following acceptance probability:

$$\alpha(\tilde{S}_T^J, \tilde{S}_T^{J-1}) = \; min\left[\prod_{t=1}^{T}\frac{f(y_t|\tilde{S}_t^J, \tilde{Y}_{t-1})}{f(y_t|\tilde{S}_t^{J-1}, \tilde{Y}_{t-1})}\prod_{t=1}^{T}\frac{h(S_t^{J-1}|\tilde{Y}_t)}{h(S_t^J|\tilde{Y}_t)}\prod_{t=0}^{T-1}\frac{h(S_{t+1}^J|\tilde{Y}_t)}{h(S_{t+1}^{J-1}|\tilde{Y}_t)}, 1\right], \tag{1.13$'$}$$

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 (1.5). 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 (1.13'). Then, we accept or reject $\tilde{S}_T^J$ according to this acceptance probability.

1.4. 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 1.1, 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 1.1.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 1.1.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 1.1.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 1.2. 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 1.2.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 1.2.B and 1.2.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 1.3. Posterior means or modes of the parameters are close to the true values. In Figure 1.3.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 1.3.B and 1.3.C, also confirm fast convergence of our multi-move sampler.

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

1.5.1. Model Specification for Ex-Post Real Interest Rate

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

$$i_t = r_t^{EA} + E[\pi_t|I_{t-1}] \tag{1.16}$$

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:

$$r_t^{EP} = r_t^{EA} - \varepsilon_t, \tag{1.17}$$

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:

$$\phi(L)(r_t^{EA} - \mu_{S_t}) = \nu_t, \tag{1.18}$$

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 (1.17) 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:

$$r_t^{EP} = \mu_{S_t} + \phi_1(r_{t-1}^{EP} - \mu_{S_{t-1}}) + \phi_2(r_{t-2}^{EP} - \mu_{S_{t-2}}) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, \tag{1.19}$$

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:

$$Pr[S_t = j | S_{t-1} = i] = p_{ij}, \quad \sum_{j=1}^{3} p_{ij} = 1; \quad i, j = 1, 2, 3. \tag{1.20}$$

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]

$$e_t \sim N(0, \sigma_t^2), \tag{1.21}$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \omega_t, \quad \omega_t \sim N(0, \sigma_\omega^2), \tag{1.22}$$

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.

$$E(r_t^{EP}|I_{t-1}) = E(\mu_{S_t}|I_{t-1}) + E(u_t|I_{t-1}), \tag{1.23}$$

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 1.2, 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.

1.5.2. 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 1.4.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 1.4.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 1.4.B.) Figures 1.4.A.1 and 1.4.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 1.4.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 1.5.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 1.5.A.1), the posterior distribution of the sum of AR coefficients (Figure 1.5.B.1), the plot of EARR (the first panel of Figure 1.5.C), the measure of time-varying volatility (the first panel of Figure 1.5.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 1.4.B) and that for our ARMA(2,2) model (the second panel of Figure 1.5.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 1.5.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

1.5.C suggest that we cannot reject the null hypothesis that the standardized prediction errors and their squares are white noise processes.

1.6. Summary and Conclusion

In this chapter 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 1.3 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).

Table 1.1. Simulation Study for Case #1 (Benchmark Case):

$[\,\boldsymbol{p_{11} = 0.9}\,;\; \boldsymbol{p_{22} = 0.96}\,; \boldsymbol{\mu_1 = 0.4}\,;\; \boldsymbol{\mu_2 = 0};\; \boldsymbol{\phi = 0.3}\,;\; \boldsymbol{\theta = 0.6}\,]$

$$y_t = \mu_{S_t} + \phi\left(y_{t-1} - \mu_{S_{t-1}}\right) + e_t - \theta\, e_{t-1},\ t = 1,2,\ldots,300$$

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

$$\Pr[\,S_t = 1 |\, S_t = 1\,] = p_{11}\,, \Pr[\,S_t = 2 |\, S_t = 2\,] = p_{22}$$

		Prior		Posterior (Proposed Algorithm)			Posterior (Billio et. al (1999))		
Parameters	True	Mean	SD	Mean	Median	SD	Mean	Median	SD
p_{11}	0.9	0.9	0.09	0.867	0.870	0.040	0.870	0.874	0.039
p_{22}	0.96	0.9	0.09	0.945	0.947	0.017	0.946	0.948	0.017
μ_1	0.4	0.5	0.3	0.393	0.392	0.021	0.393	0.393	0.020
μ_2	0	0	0.3	-0.001	-0.001	0.011	-0.001	-0.001	0.011
ϕ	0.3	0.5	0.3	0.340	0.342	0.158	0.302	0.304	0.153
θ	0.6	0.5	0.3	0.564	0.574	0.148	0.520	0.527	0.139
σ	0.2	0.5	0.2	0.208	0.208	0.009	0.208	0.208	0.009
e_0	-	0	0.5	-0.197	-0.200	0.427	-0.199	-0.202	0.426
Correct Assignment Rate					0.949			0.950	
Proportion of Acceptance					0.240			0.076	
Burn-in /Total iterations					5,000/15,000			40,000/50,000	

Note: A correct assignment rate is a ratio of the number of correctly assigned MS states to the total number of generated MS states.

Table 1.2. Simulation Study for Case #2 (The Case of High Transition Probabilities):
$[\,\boldsymbol{p_{11} = 0.95};\ \boldsymbol{p_{22} = 0.99}\,;\boldsymbol{\mu_1 = 0.4}\,;\ \boldsymbol{\mu_2 = 0};\ \boldsymbol{\phi = 0.3}\,;\ \boldsymbol{\theta = 0.6}\,]$

$$y_t = \mu_{S_t} + \phi\left(y_{t-1} - \mu_{S_{t-1}}\right) + e_t - \theta\, e_{t-1},\ t = 1,2,\dots,300$$

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

$$\Pr[\,S_t = 1 |\, S_t = 1\,] = p_{11}\,, \Pr[\,S_t = 2 |\, S_t = 2\,] = p_{22}$$

	Prior			Posterior (Proposed Algorithm)			Posterior (Billio et. al (1999))	
Parameters	True	Mean	SD	Mean	Median	SD	Mean	Median	SD
p_{11}	0.95	0.9	0.09	0.938	0.942	0.029	0.940	0.944	0.028
p_{22}	0.99	0.9	0.09	0.986	0.987	0.007	0.986	0.988	0.007
μ_1	0.4	0.5	0.3	0.379	0.378	0.017	0.398	0.398	0.019
μ_2	0	0	0.3	-0.004	-0.004	0.008	0.000	0.000	0.009
ϕ	0.3	0.5	0.3	0.376	0.380	0.123	0.209	0.213	0.147
θ	0.6	0.5	0.3	0.637	0.649	0.106	0.494	0.505	0.128
σ	0.2	0.5	0.2	0.212	0.211	0.008	0.217	0.216	0.008
e_0	-	0	0.5	-0.213	-0.217	0.392	0.098	0.097	0.409
Correct Assignment Rate					0.995			0.993	
Proportion of Acceptance					0.620			0.590	
Burn-in /Total iterations					5,000/15,000			140,000/150,000	

Note: A correct assignment rate is a ratio of the number of correctly assigned MS states to the total number of generated MS states.

Table 1.3. Simulation Study for Case #3 (The Case of a Structural Break with an Absorbing State): $[\,\boldsymbol{p_{11} = 0.993}\,;\ \boldsymbol{p_{22} = 1}\,; \boldsymbol{\mu_1 = 0.4}\,;\ \boldsymbol{\mu_2 = 0};\ \boldsymbol{\phi = 0.3}\,;\ \boldsymbol{\theta = 0.6}\,]$

$$y_t = \mu_{S_t} + \phi\left(y_{t-1} - \mu_{S_{t-1}}\right) + e_t - \theta\, e_{t-1},\ t = 1,2,\ldots,300$$

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

$$\Pr[\,S_t = 1 |\, S_t = 1\,] = p_{11}\,, \Pr[\,S_t = 2 |\, S_t = 2\,] = 1$$

		Prior		Posterior (Proposed Algorithm)		
Parameters	True	Mean	SD	Mean	Median	SD
p_{11}	0.993	0.9	0.09	0.987	0.989	0.008
p_{22}	1	-	-	-	-	-
μ_1	0.4	0.5	0.3	0.397	0.397	0.010
μ_2	0	0	0.3	-0.011	-0.011	0.010
ϕ	0.3	0.5	0.3	0.363	0.368	0.131
θ	0.6	0.5	0.3	0.615	0.627	0.116
σ	0.2	0.5	0.2	0.211	0.211	0.008
e_0	-	0	0.5	-0.167	-0.165	0.399
Correct Assignment Rate				0.998		
Proportion of Acceptance				0.962		
Burn-in /Total iterations				5,000/15,000		

Note: A correct assignment rate is a ratio of the number of correctly assigned MS states to the total number of generated MS states.

Table 1.4.A. Posterior Moments: AR (2) Model with Markov-Switching Mean [Garcia and Perron (1996) / 1960:1~1986:2]

$$r_t^{EP} = \mu_{S_t} + \phi_1\left(r_{t-1}^{EP} - \mu_{S_{t-1}}\right) + \phi_1\left(r_{t-2}^{EP} - \mu_{S_{t-2}}\right) + e_t$$

$$e_t \sim N\,(0, \sigma_t^2)$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t, \;\; \varepsilon_t \sim i.i.d.N\,(0, \sigma_\varepsilon^2)$$

$$\Pr[S_t = j\,|S_{t-1} = i] = p_{ij,} \qquad \Sigma_{j=1}^{3}\, p_{ij,} = 1, \;\; i,j = 1,2,3$$

Parameters	Prior		Posterior			
	Mean	SD	Mean	Median	SD	90 % HPD
p_{11}	0.98	0.04	0.970	0.977	0.026	[0.935, 0.999]
p_{12}	0.01	0.03	0.003	0.000	0.009	[0.000, 0.008]
p_{21}	0.01	0.03	0.018	0.013	0.017	[0.002, 0.035]
p_{22}	0.98	0.04	0.980	0.985	0.018	[0.955, 0.999]
p_{31}	0.01	0.03	0.003	0.000	0.010	[0.000, 0.009]
p_{32}	0.01	0.03	0.003	0.000	0.010	[0.000, 0.009]
μ_1	0	1	-1.403	-1.417	0.345	[-1.935, -0.851]
μ_2	2	1	1.405	1.404	0.190	[1.119, 1.715]
μ_3	4	1	4.938	4.949	0.406	[4.323, 5.630]
$\phi_1 + \phi_2$	0	0.5	0.062	0.061	0.153	[-0.175 , 0.317]
ϕ_2	0	0.5	0.111	0.111	0.108	[-0.059, 0.293]
σ_ε^2	0.02	0.1	0.015	0.011	0.014	[0.003, 0.031]
e_0	0	2	-0.803	-0.851	2.049	[-4.041, 2.607]
Largest AR Root	-	-	0.378	0.386	0.141	[0.141, 0.604]
Acceptance Rate for S_t					0.932	

Note: 1. Burn-in / Total iterations = 5,000 / 25,000

2. SD refers to standard deviation.

3. A highest posterior density (HPD) region is a posterior density interval, the narrowest one possible with a chosen probability.

Table 1.4.B. Posterior Moments: AR (2) Model with Markov-Switching Mean [Garcia and Perron (1996) / 1960:1~2008:2]

$$r_t^{EP} = \mu_{S_t} + \phi_1\left(r_{t-1}^{EP} - \mu_{S_{t-1}}\right) + \phi_1\left(r_{t-2}^{EP} - \mu_{S_{t-2}}\right) + e_t$$

$$e_t \sim N\,(0, \sigma_t^2)$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t,\;\; \varepsilon_t \sim i.i.d.N\,(0, \sigma_\varepsilon^2)$$

$$\Pr[S_t = j\,|S_{t-1} = i] = p_{ij,} \qquad \Sigma_{j=1}^{3}\, p_{ij,} = 1, \quad i,j = 1,2,3$$

	Prior		Posterior			
Parameters	Mean	SD	Mean	Median	SD	90 % HPD
p_{11}	0.98	0.04	0.977	0.983	0.022	[0.948, 0.999]
p_{12}	0.01	0.03	0.003	0.000	0.009	[0.000, 0.009]
p_{21}	0.01	0.03	0.017	0.014	0.014	[0.001, 0.034]
p_{22}	0.98	0.04	0.982	0.985	0.015	[0.966, 0.999]
p_{31}	0.01	0.03	0.003	0.000	0.011	[0.000, 0.009]
p_{32}	0.01	0.03	0.035	0.026	0.033	[0.001, 0.083]
μ_1	0	1	-1.054	-1.112	0.530	[-1.849, -0.222]
μ_2	2	1	1.676	1.672	0.233	[1.329, 2.065]
μ_3	4	1	4.598	4.667	0.651	[3.590, 5.605]
$\phi_1 + \phi_2$	0	0.5	0.336	0.330	0.133	[0.125, 0.550]
ϕ_2	0	0.5	0.231	0.229	0.088	[0.098, 0.377]
σ_ε^2	0.02	0.1	0.021	0.017	0.013	[0.005, 0.040]
e_0	0	2	-0.809	-0.821	1.920	[-3.927, 2.340]
Largest AR Root	-	-	0.533	0.537	0.110	[0.363, 0.716]
Acceptance Rate for S_t				0.749		

Note: 1. Burn-in / Total iterations = 5,000 / 25,000

2. SD refers to standard deviation.

3. A highest posterior density (HPD) region is a posterior density interval, the narrowest one possible with a chosen probability.

Table 1.4.C. Diagnostic Checks: AR (2) Model with Markov-Switching Mean [Garcia and Perron (1996)]

	1960:1~1986:2			**1960:1~2008:2**		
			Standardized Forecasting Error			
Lags	Autocorrelation	Q-Statistics	P-Value	Autocorrelation	Q-Statistics	P-Value
1	-0.020	0.0419	0.838	-0.102	2.0443	0.153
2	0.007	0.0471	0.977	-0.051	2.5610	0.278
3	0.170	3.2512	0.354	0.232	13.246	0.004
4	0.082	4.0056	0.405	-0.012	13.276	0.010
5	0.147	6.4455	0.265	0.072	14.322	0.014
6	0.042	6.6499	0.354	0.076	15.476	0.017
			Squared Standardized Forecasting Error			
Lags	Autocorrelation	Q-Statistics	P-Value	Autocorrelation	Q-Statistics	P-Value
1	-0.009	0.0083	0.927	0.041	0.3284	0.567
2	-0.043	0.2112	0.900	-0.058	0.9876	0.610
3	-0.066	0.6958	0.874	-0.071	1.9893	0.575
4	0.018	0.7306	0.948	0.017	2.0493	0.727
5	-0.025	0.7996	0.977	-0.040	2.3720	0.796
6	0.153	3.4809	0.747	0.042	2.7269	0.842

Note:
1. Burn-in / Total iterations = 5,000 / 25,000
2. Forecasting errors are standardized with estimated stochastic volatilities.
3. Autocorrelations are calculated with the posterior averages of the standardized forecasting errors and the squatted forecasting errors.

Table 1.5.A. Posterior Moments: ARMA(2,2) Model with Markov-Switching Mean [Proposed Model/ 1960:1~1986:2]

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

$$e_t \sim N\,(0, \sigma_t^2)$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t,\;\; \varepsilon_t \sim i.i.d.N\,(0, \sigma_\varepsilon^2)$$

$$\Pr[S_t = j\,|S_{t-1} = i] = p_{ij,} \quad \Sigma_{j=1}^3\, p_{ij,} = 1, \quad i, j = 1,2,$$

Parameters	Prior		Posterior			
	Mean	SD	Mean	Median	SD	90 % HPD
p_{11}	0.98	0.04	0.970	0.977	0.027	[0.932 , 0.999]
p_{12}	0.01	0.03	0.003	0.000	0.010	[0.000, 0.008]
p_{21}	0.01	0.03	0.018	0.013	0.017	[0.000, 0.044]
p_{22}	0.98	0.04	0.980	0.985	0.018	[0.956, 0.999]
p_{31}	0.01	0.03	0.003	0.000	0.010	[0.000, 0.008]
p_{32}	0.01	0.03	0.003	0.000	0.011	[0.001, 0.008]
μ_1	0	1	-1.361	-1.395	0.394	[-1.961, -0.777]
μ_2	2	1	1.417	1.411	0.216	[1.112, 1.761]
μ_3	4	1	4.929	4.946	0.421	[4.270, 5.599]
$\phi_1 + \phi_2$	0	0.5	0.128	0.126	0.379	[-0.460, 0.840]
ϕ_2	0	0.5	0.076	0.085	0.278	[-0.315, 0.536]
$\theta_1 + \theta_2$	0	0.5	0.065	0.063	0.323	[-0.451, 0.612]
$\theta_{2,1}$	0	0.5	-0.021	-0.018	0.278	[-0.440, 0.461]
σ_ε^2	0.02	0.1	0.013	0.010	0.012	[0.003, 0.025]
e_0	0	2	-0.550	-0.590	2.110	[-4.032, 3.013]
Largest AR Root	-	-	0.563	0.561	0.209	[0.248, 0.932]
Largest MA Root	-	-	0.535	0.525	0.209	[0.190, 0.883]
Acceptance Rate for S_t					0.903	

Note: 1. Burn-in / Total iterations = 5,000 / 25,000

2. SD refers to standard deviation.

3. A highest posterior density (HPD) region is a posterior density interval, the narrowest one possible with a chosen probability.

Table 1.5.B. Posterior Moments: ARMA(2,2) Model with Markov-Switching Mean [Proposed Model/ 1960:1~2008:2]

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

$$e_t \sim N\,(0, \sigma_t^2)$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t,\ \ \varepsilon_t \sim i.i.d.N\,(0, \sigma_\varepsilon^2)$$

$$\Pr[S_t = j\,|S_{t-1} = i] = p_{ij,} \quad \Sigma_{j=1}^3 p_{ij,} = 1, \quad i, j = 1,2,$$

Parameters	Prior		Posterior			
	Mean	SD	Mean	Median	SD	90 % HPD
p_{11}	0.98	0.04	0.973	0.983	0.032	[0.943, 0.999]
p_{12}	0.01	0.03	0.005	0.000	0.014	[0.000, 0.014]
p_{21}	0.01	0.03	0.017	0.013	0.018	[0.001, 0.035]
p_{22}	0.98	0.04	0.981	0.986	0.020	[0.961, 0.999]
p_{31}	0.01	0.03	0.005	0.000	0.014	[0.000, 0.023]
p_{32}	0.01	0.03	0.032	0.021	0.034	[0.001, 0.080]
μ_1	0	1	-0.707	-0.822	0.746	[-1.795, 0.705]
μ_2	2	1	1.696	1.682	0.381	[1.077, 2.283]
μ_3	4	1	4.307	4.452	0.882	[2.655, 5.695]
$\phi_1 + \phi_2$	0	0.5	0.679	0.732	0.229	[0.299, 0.999]
ϕ_2	0	0.5	0.220	0.228	0.275	[--0.242, 0.678]
$\theta_1 + \theta_2$	0	0.5	0.337	0.363	0.221	[-0.050 , 0.675]
$\theta_{2,1}$	0	0.5	-0.014	-0.028	0.278	[-0.462, 0.440]
σ_ε^2	0.02	0.1	0.021	0.017	0.018	[0.006, 0.038]
e_0	0	2	-0.904	-0.912	1.892	[-4.089, 2.229]
Largest AR Root	-	-	0.757	0.795	0.168	[0.501, 0.999]
Largest MA Root	-	-	0.534	0.541	0.163	[0.253, 0.796]
Acceptance Rate for S_t				0.517		

Note: 1. Burn-in / Total iterations = 5,000 /25,000
2. SD refers to standard deviation.
3. A highest posterior density (HPD) region is a posterior density interval, the narrowest one possible with a chosen probability.

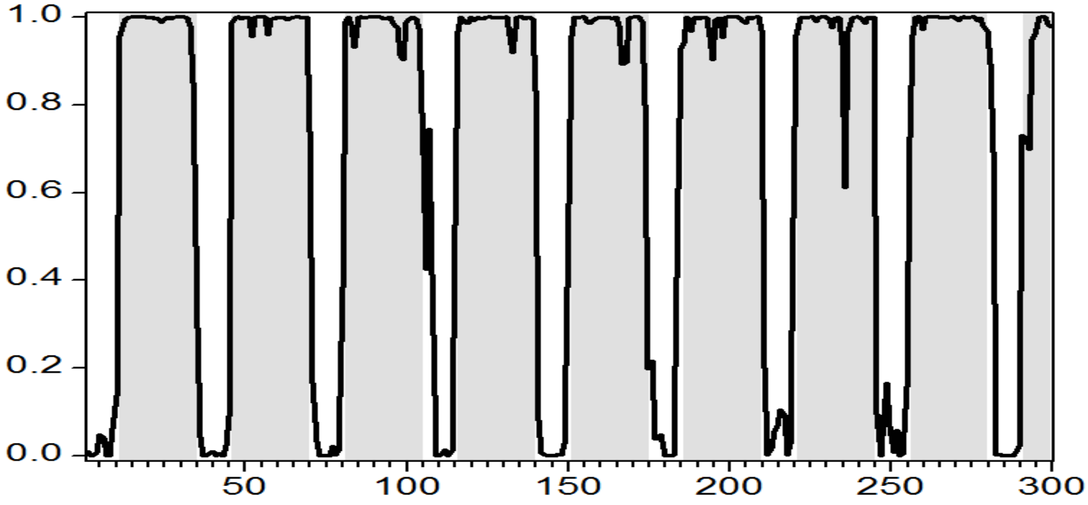
Table 1.5.C. Diagnostic Checks: ARMA (2,2) Model with Markov-Switching Mean [Proposed Model]

	1960:1~1986:2			**1960:1~2008:2**		
			Standardized Forecasting Error			
Lags	Autocorrelation	Q-Statistics	P-Value	Autocorrelation	Q-Statistics	P-Value
1	-0.031	0.1059	0.745	-0.091	1.6382	0.201
2	0.025	0.1774	0.915	-0.056	2.2524	0.324
3	0.144	2.4891	0.477	0.131	5.6644	0.129
4	0.077	3.1530	0.533	-0.041	5.9949	0.200
5	0.148	5.6355	0.343	0.064	6.8100	0.235
6	0.035	5.7732	0.449	0.042	7.1621	0.306
			Squared Standardized Forecasting Error			
Lags	Autocorrelation	Q-Statistics	P-Value	Autocorrelation	Q-Statistics	P-Value
1	0.001	0.0002	0.989	0.053	0.5588	0.455
2	-0.036	0.1451	0.930	-0.056	1.1779	0.555
3	-0.071	0.7023	0.873	-0.107	3.4751	0.324
4	0.010	0.7134	0.950	0.015	3.5209	0.475
5	-0.030	0.8125	0.976	-0.015	3.5652	0.614
6	0.164	3.9029	0.690	0.015	3.6101	0.729

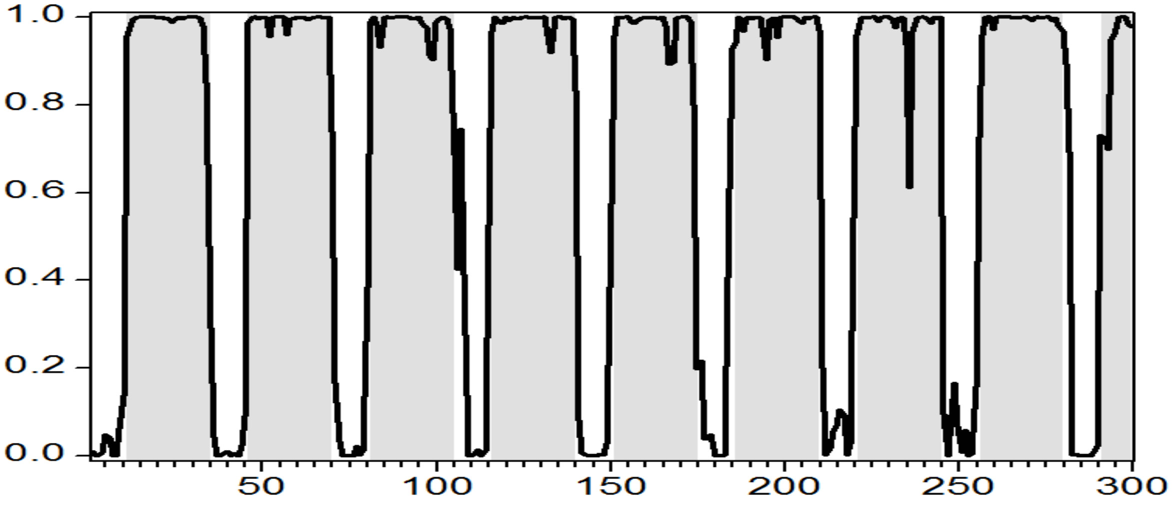
Note: 1. Burn-in / Total iterations = 5,000 / 25,000
2. Forecasting errors are standardized with estimated stochastic volatilities.
3. Autocorrelations are calculated with the posterior averages of the standardized forecasting errors and the squatted forecasting errors.

Figure 1.1.A. Posterior Probabilities of Regime 2: Simulation Study for Case #1

$[\, \boldsymbol{p_{11} = 0.9}\,;\; \boldsymbol{p_{22} = 0.96}\,; \boldsymbol{\mu_1 = 0.4}\,;\; \boldsymbol{\mu_2 = 0};\; \boldsymbol{\phi = 0.3}\,;\; \boldsymbol{\theta = 0.6}\,]$



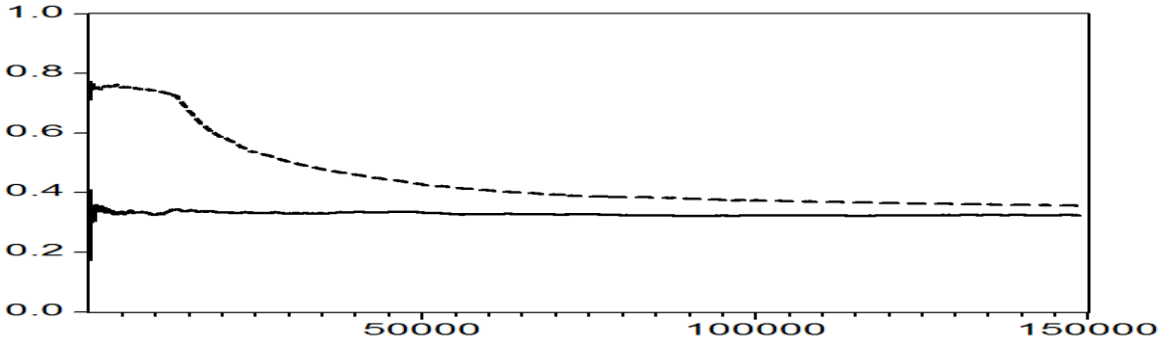
Proposed Multi-move Algorithm
(Burn-in / Total iterations: 5,000 / 15,000)



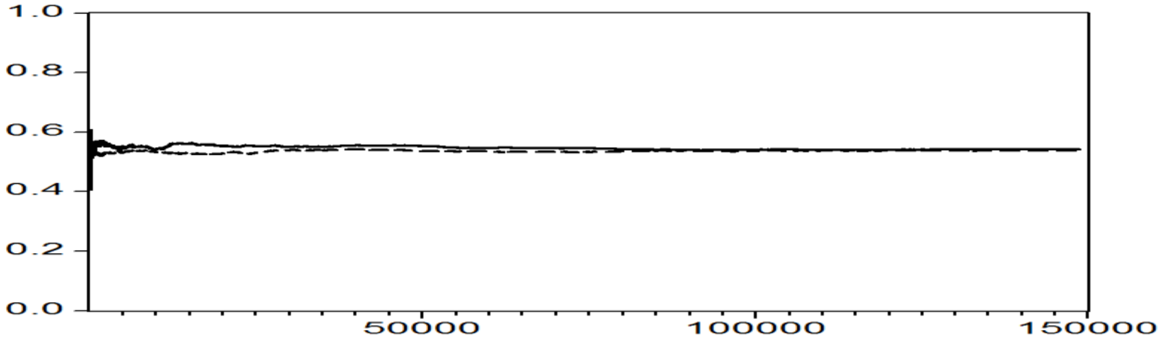
Single-move Algorithm by Billio et. al (1999)
(Burn-in / Total iterations: 40,000 / 50,000)

Note The shaded areas represent the true periods of Regime 2. ($S_t = 2$).

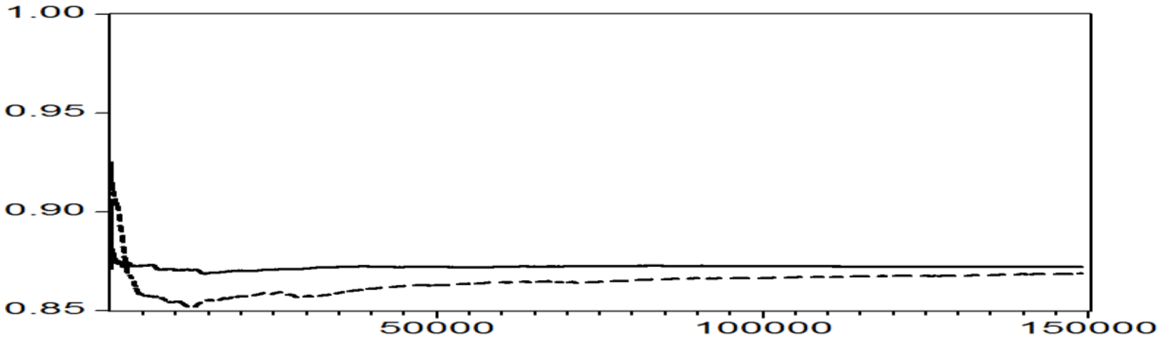
Figure 1.1.B. Cumulative Averages of MCMC Samples: Simulation Study for Case #1 $[\,p_{11}=0.9\,;\ p_{22}=0.96\,;\mu_1=0.4\,;\ \mu_2=0;\ \phi=0.3\,;\ \theta=0.6\,]$



Cumulative Averages of AR Coefficient



Cumulative Averages of MA Coefficient



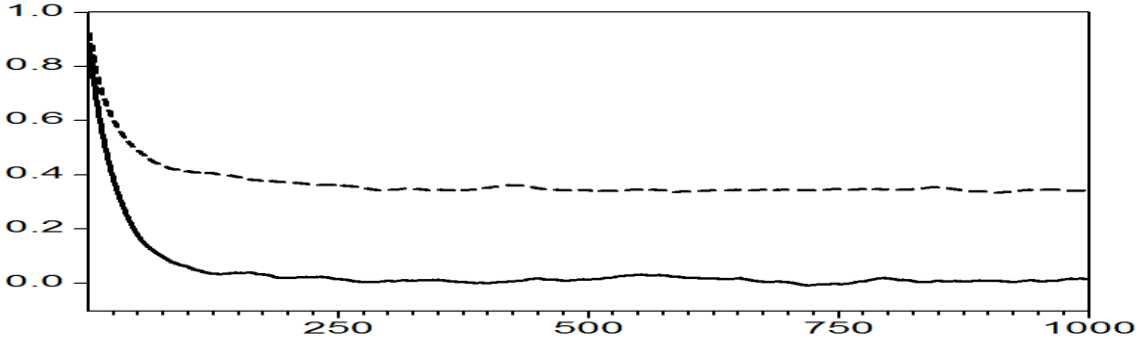
Cumulative Averages of Transition Probability for Regime 1

Note

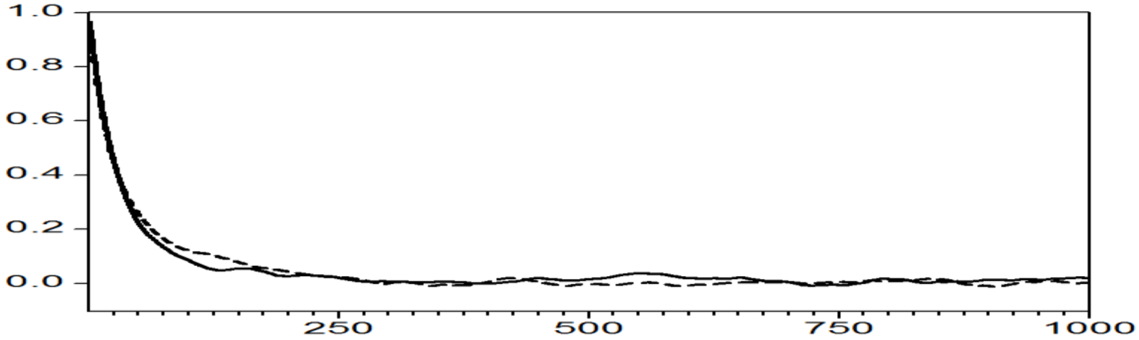
1. Bold lines are associated with the proposed algorithm and dotted lines are based on Billio et. al (1999)'s single-move algorithm.
2. Burn-in / Total iterations = 1,000 / 150,000

Figure 1.1.C. Autocorrelations of MCMC Samples: Simulation Study for Case #1

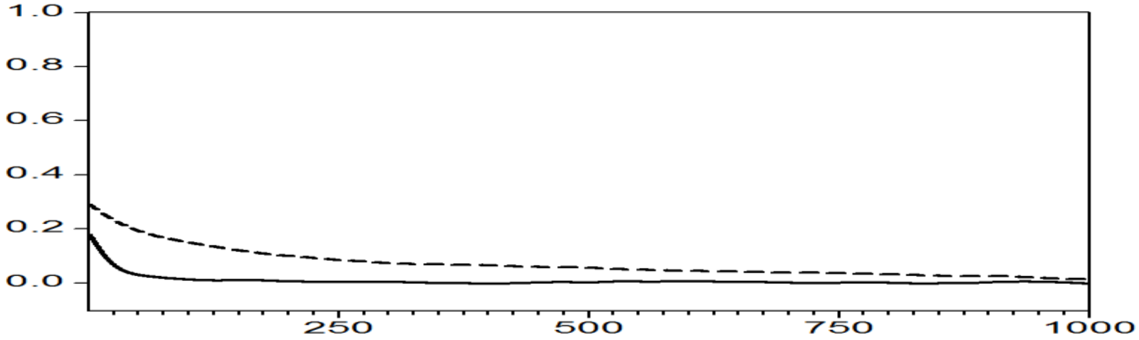
$[\,p_{11} = 0.9\,;\;\; p_{22} = 0.96\,; \mu_1 = 0.4\,;\; \mu_2 = 0;\; \phi = 0.3\,;\; \theta = 0.6\,]$



Autocorrelations of AR Coefficient



Autocorrelations of MA Coefficient



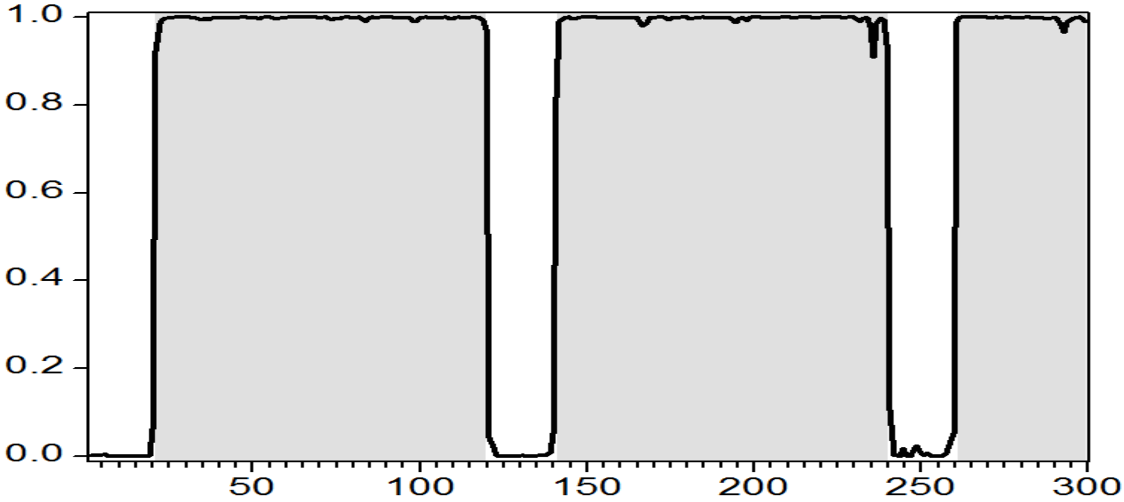
Cumulative Averages of Transition Probability for Regime 1

<u>**Note**</u>

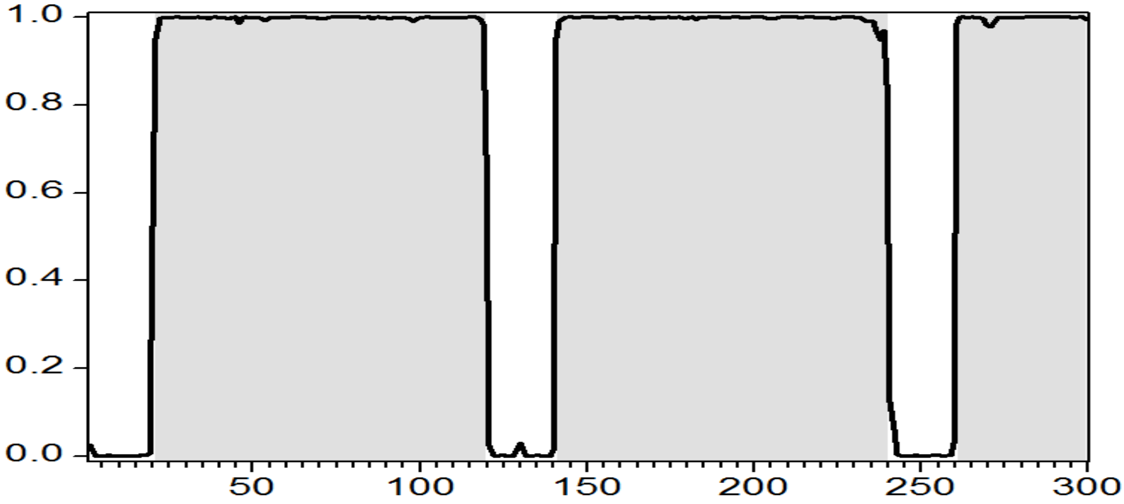
1. Bold lines are associated with the proposed algorithm and dotted lines are based on Billio et. al (1999)'s single-move algorithm.
2. Burn-in / Total iterations = 1,000 / 150,000

Figure 1.2.A. Posterior Probabilities of Regime 2: Simulation Study for Case #2

$[\, p_{11} = 0.95\, ;\;\; p_{22} = 0.99\, ; \mu_1 = 0.4\, ;\; \mu_2 = 0;\; \phi = 0.3\, ;\; \theta = 0.6\,]$



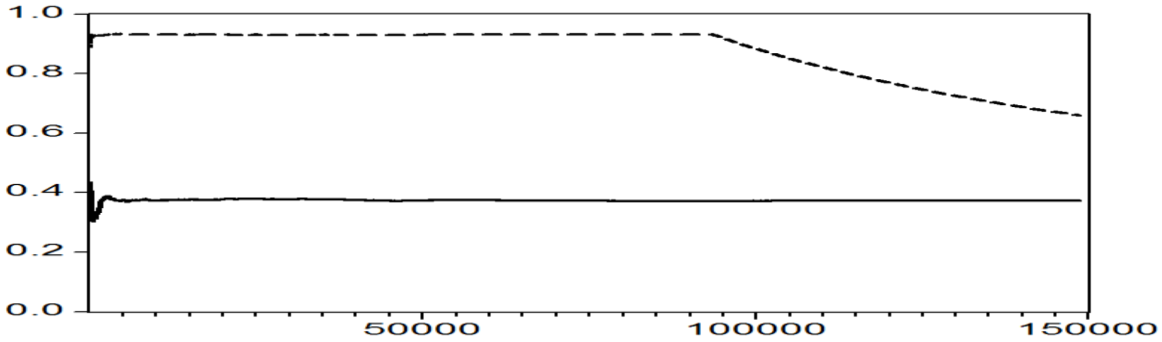
Proposed Multi-move Algorithm
(Burn-in / Total iterations: 5,000 / 15,000)



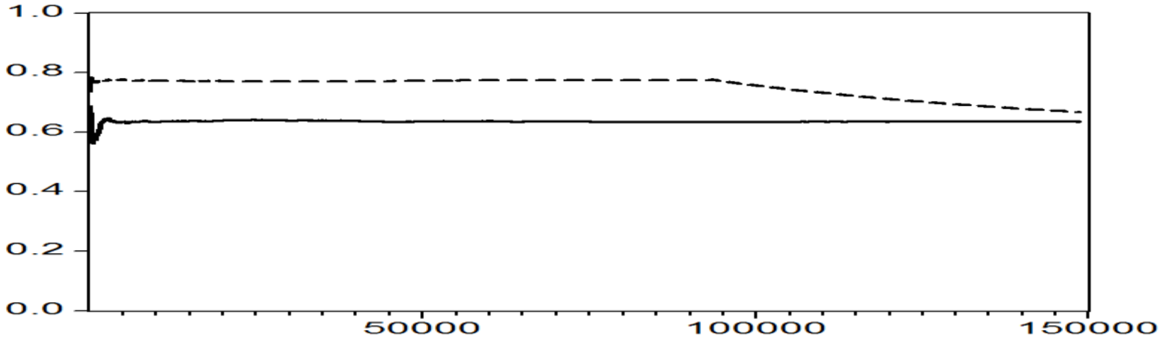
Single-move Algorithm by Billio et. al (1999)
(Burn-in / Total iterations: 140,000 / 150,000)

<u>Note</u> 1. The shaded areas represent the true periods of Regime 2. ($S_t = 2$).

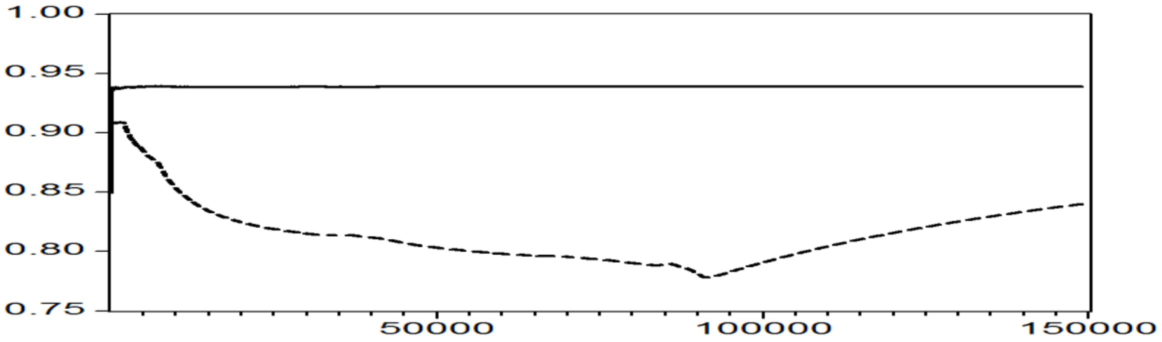
Figure 1.2.B. Cumulative Averages of MCMC Samples: Simulation Study for Case #2 [$p_{11} = 0.95$; $p_{22} = 0.99$; $\mu_1 = 0.4$; $\mu_2 = 0$; $\phi = 0.3$; $\theta = 0.6$]



Cumulative Averages of AR Coefficient



Cumulative Averages of MA Coefficient

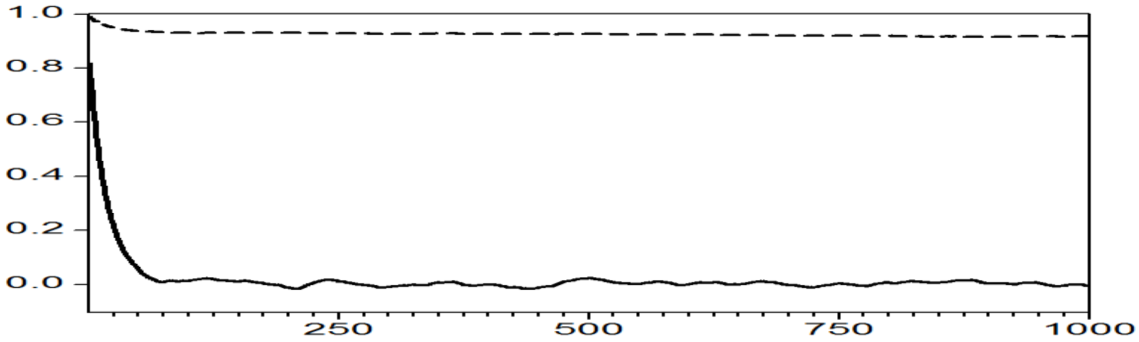


Cumulative Averages of Transition Probability for Regime 1

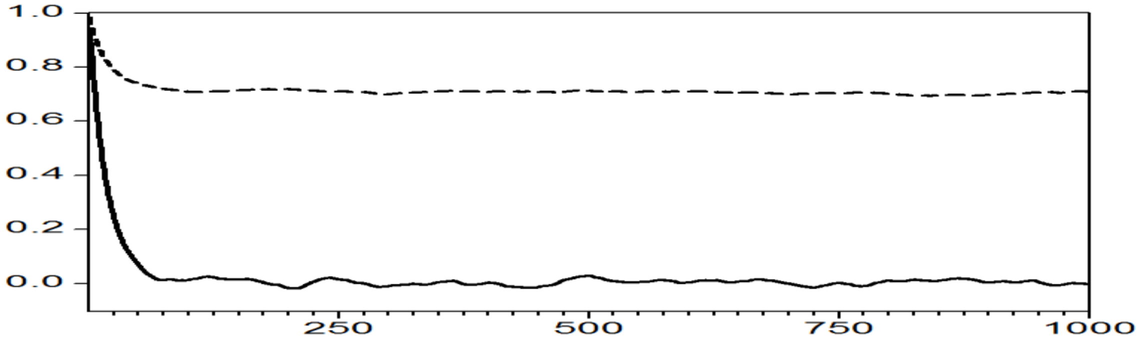
Note

1. Bold lines are associated with the proposed algorithm and dotted lines are based on Billio et. al (1999)'s single-move algorithm.
2. Burn-in / Total iterations = 1,000 / 150,000

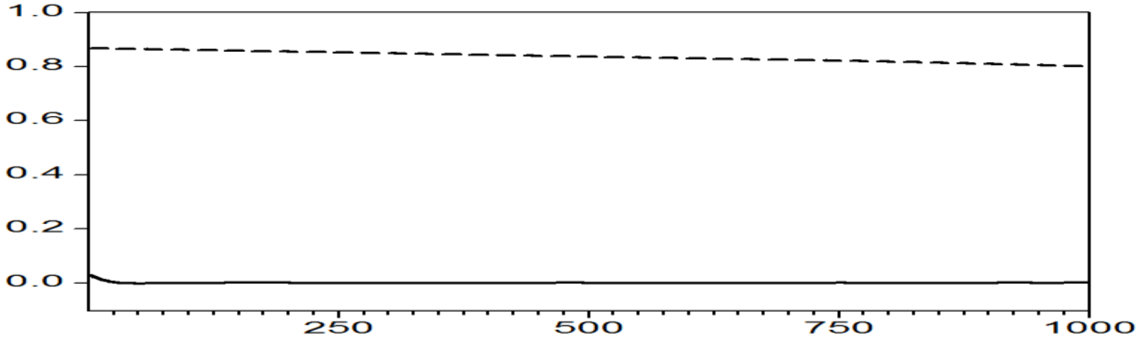
Figure 1.2.C. Autocorrelations of MCMC Samples: Simulation Study for Case #2 [$p_{11} = 0.95$; $p_{22} = 0.99$; $\mu_1 = 0.4$; $\mu_2 = 0$; $\phi = 0.3$; $\theta = 0.6$]



Autocorrelations of AR Coefficient



Autocorrelations of MA Coefficient



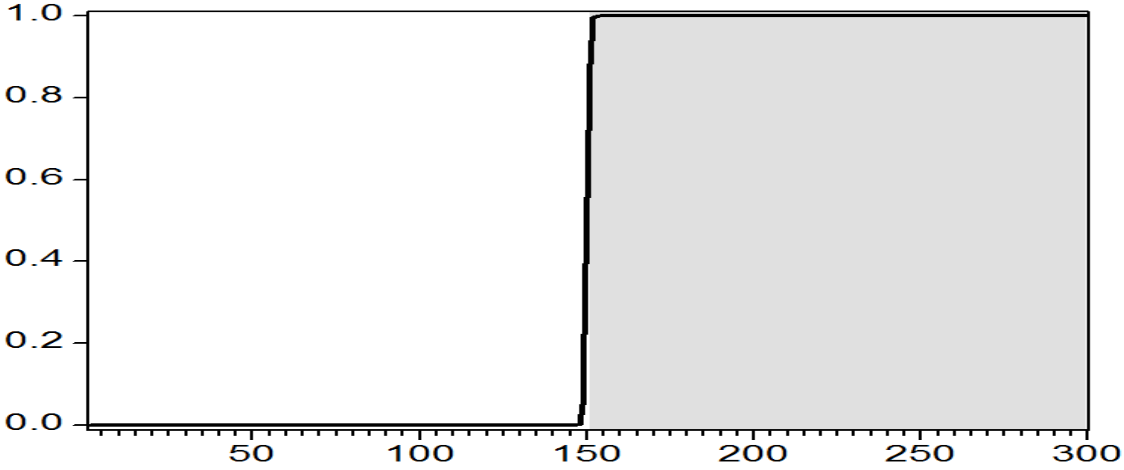
Cumulative Averages of Transition Probability for Regime 1

<u>Note</u>

1. Bold lines are associated with the proposed algorithm and dotted lines are based on Billio et. al (1999)'s single-move algorithm.
2. Burn-in / Total iterations = 1,000 / 150,000

Figure 1.3.A. Posterior Probabilities of Regime 2: Simulation Study for Case #3

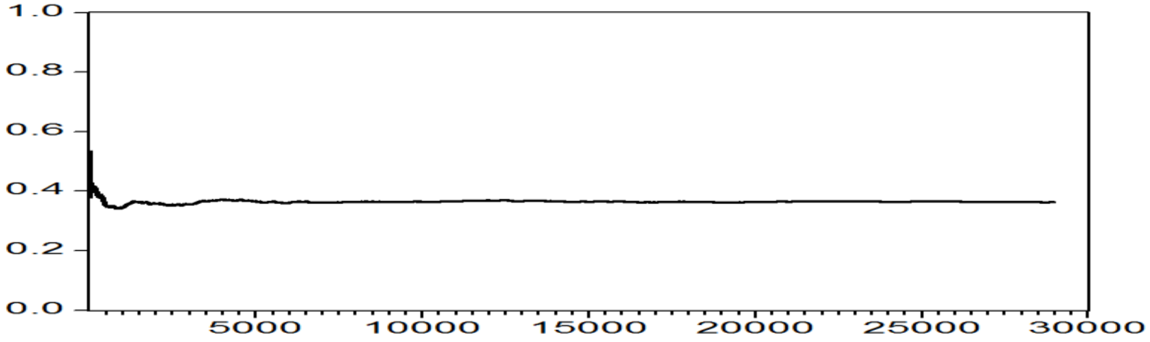
$[\boldsymbol{p_{11} = 0.993}\,;\ \boldsymbol{p_{22} = 1}\,;\boldsymbol{\mu_1 = 0.4}\,;\ \boldsymbol{\mu_2 = 0};\ \boldsymbol{\phi = 0.3}\,;\ \boldsymbol{\theta = 0.6}\,]$



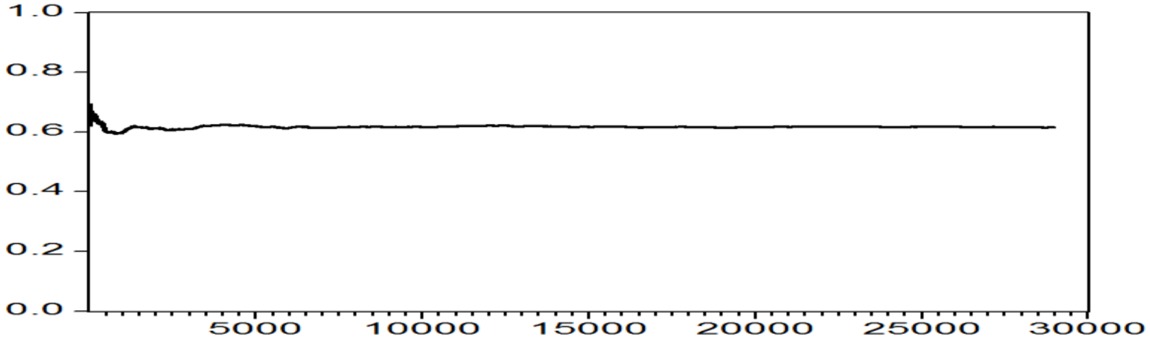
Proposed Multi-move Algorithm
(Burn-in / Total iterations: 5,000 / 15,000)

<u>Note</u> The shaded area represents the true periods of Regime 2. ($S_t = 2$)

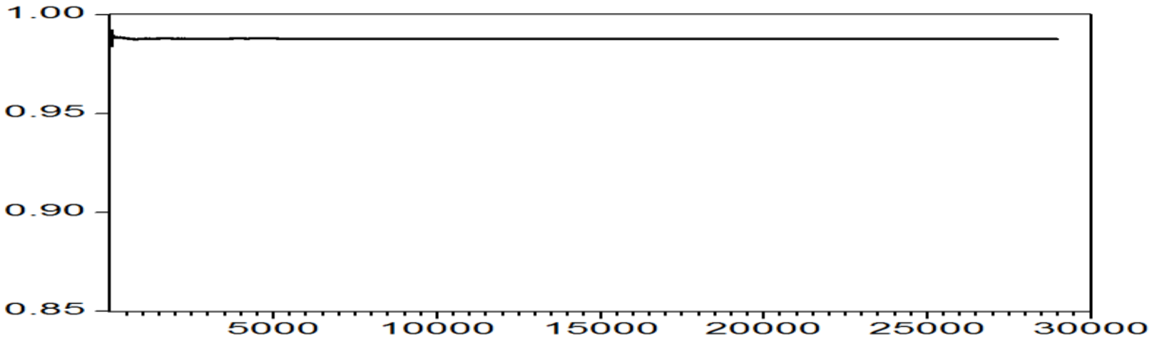
Figure 1.3.B. Cumulative Averages of MCMC Samples: Simulation Study for Case #3
$[\, p_{11} = 0.993\,;\ p_{22} = 1\,; \mu_1 = 0.4\,;\ \mu_2 = 0;\ \phi = 0.3\,;\ \theta = 0.6\,]$



Cumulative Averages of AR Coefficient



Cumulative Averages of MA Coefficient

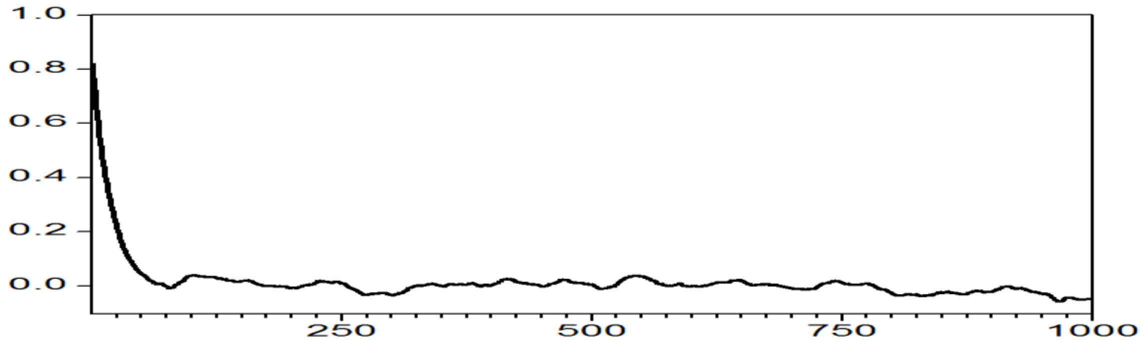


Cumulative Averages of Transition Probability for Regime 1

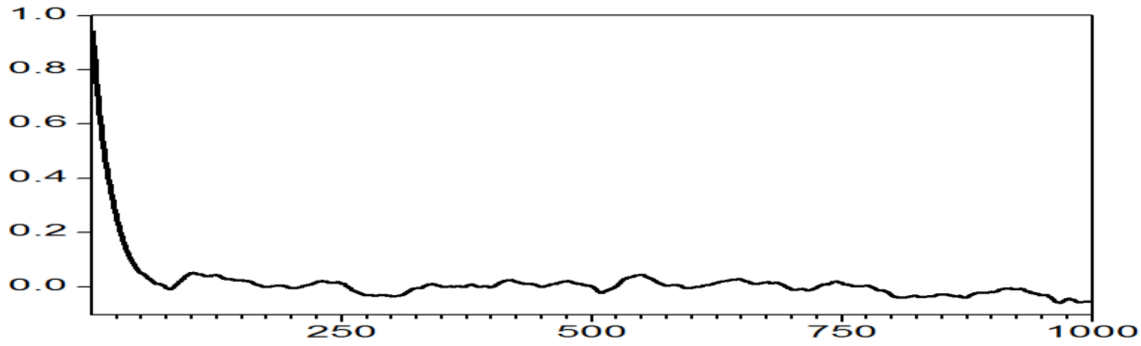
Note Burn-in / Total iterations = 1,000 / 300,000

Figure 1.3.C. Autocorrelations of MCMC Samples: Simulation Study for Case #3

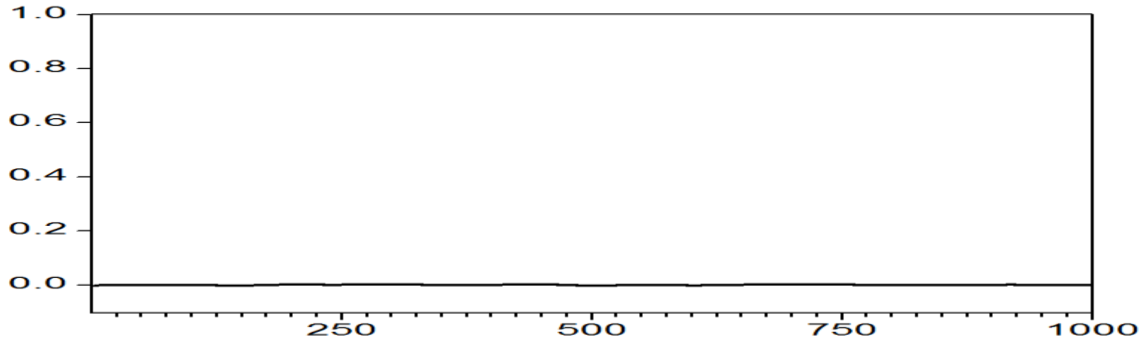
$[\,\boldsymbol{p_{11} = 0.993}\,;\;\; \boldsymbol{p_{22} = 1}\,; \boldsymbol{\mu_1 = 0.4}\,;\; \boldsymbol{\mu_2 = 0};\; \boldsymbol{\phi = 0.3}\,;\; \boldsymbol{\theta = 0.6}\,]$



Autocorrelations of AR Coefficient



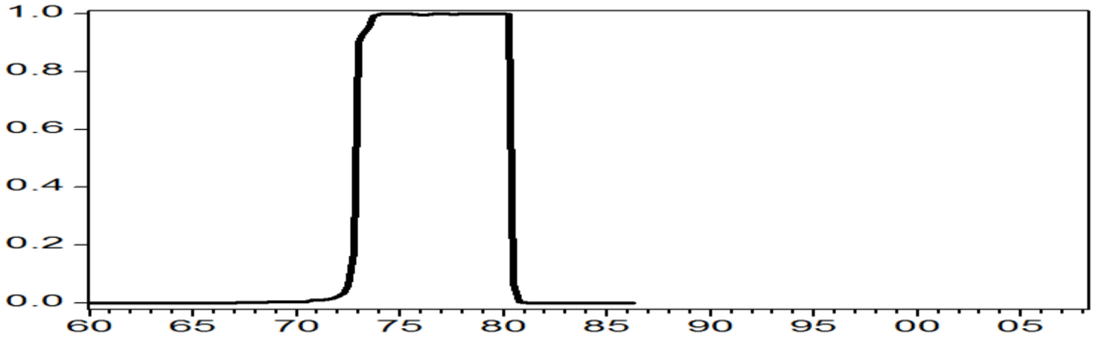
Autocorrelations of MA Coefficient



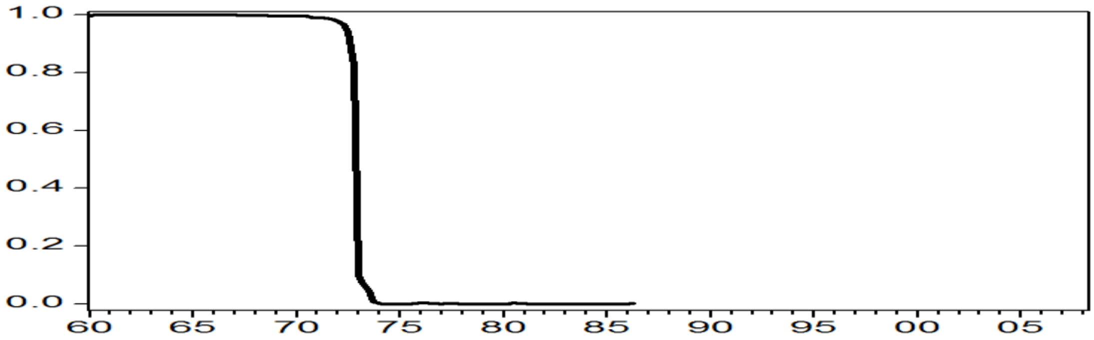
Autocorrelations of Transition Probability for Regime 1

<u>Note</u> 1. Burn-in / Total iterations = 1,000 / 300,000

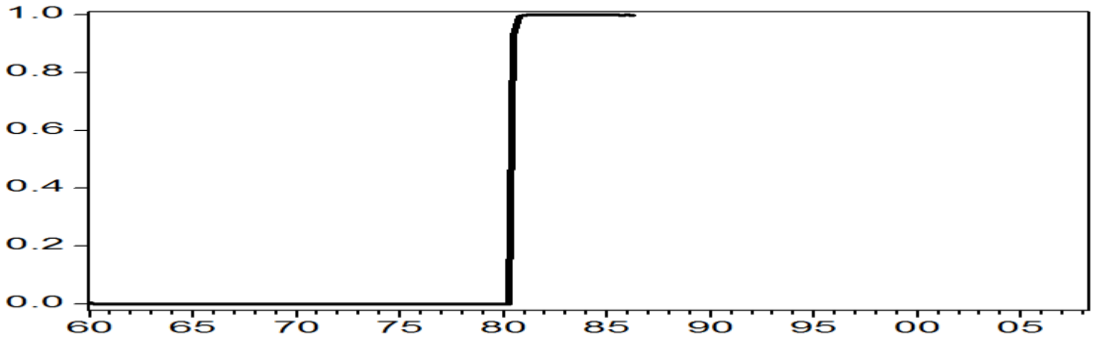
Figure 1.4.A.1. Regime Probabilities: AR (2) Model with Markov-Switching Mean [Garcia and Perron (1996) / 1960:1~1986:2]



Probabilities of a Low Interest Rate Regime



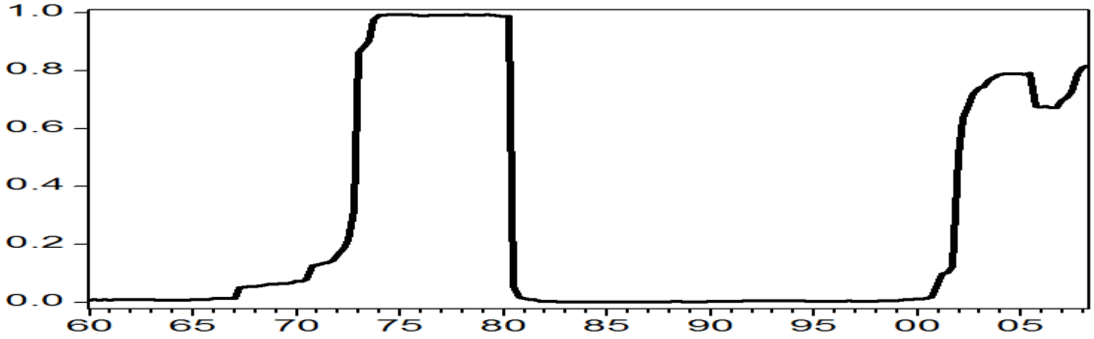
Probability of a Medium Interest Rate Regime



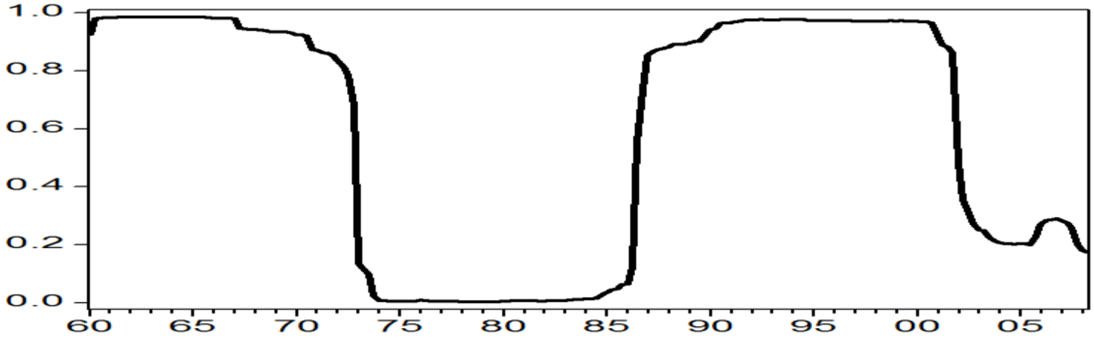
Probability of a High Interest Rate Regime

Note 1. Burn-in / Total iterations = 5,000 / 25,000
2. $S_t = 1$: a low interest regime, $S_t = 2$: a medium interest regime, $S_t = 3$: a high interest regime,

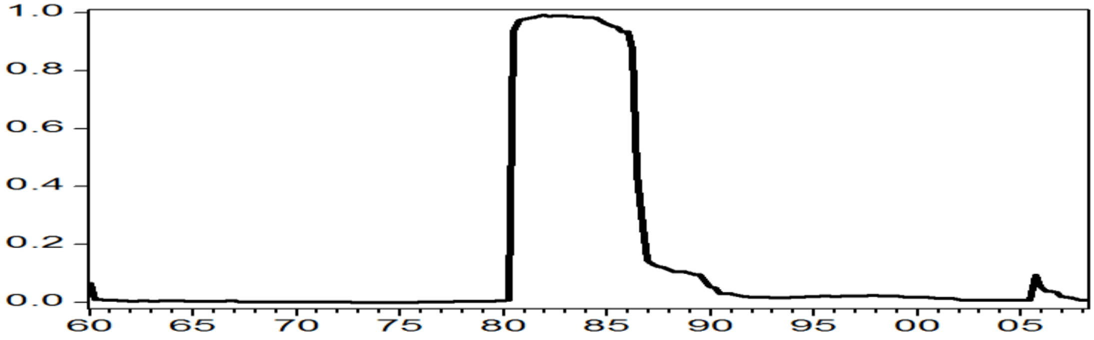
Figure 1.4.A.2. Regime Probabilities: AR (2) Model with Markov-Switching Mean
[Garcia and Perron (1996) / 1960:1~2008:2]



Probability of a Low Interest Rate Regime



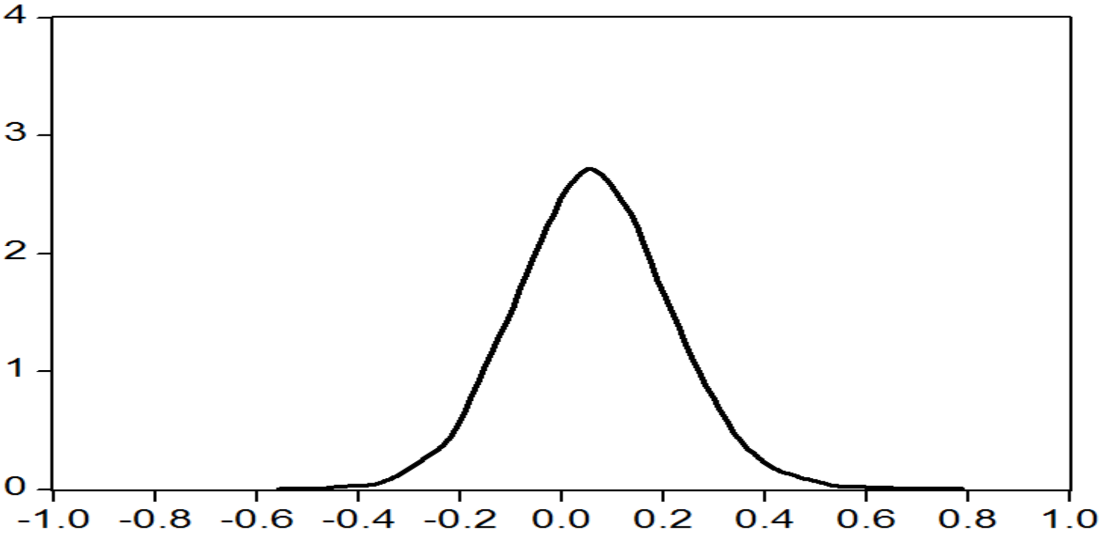
Probability of a Medium Interest Rate Regime



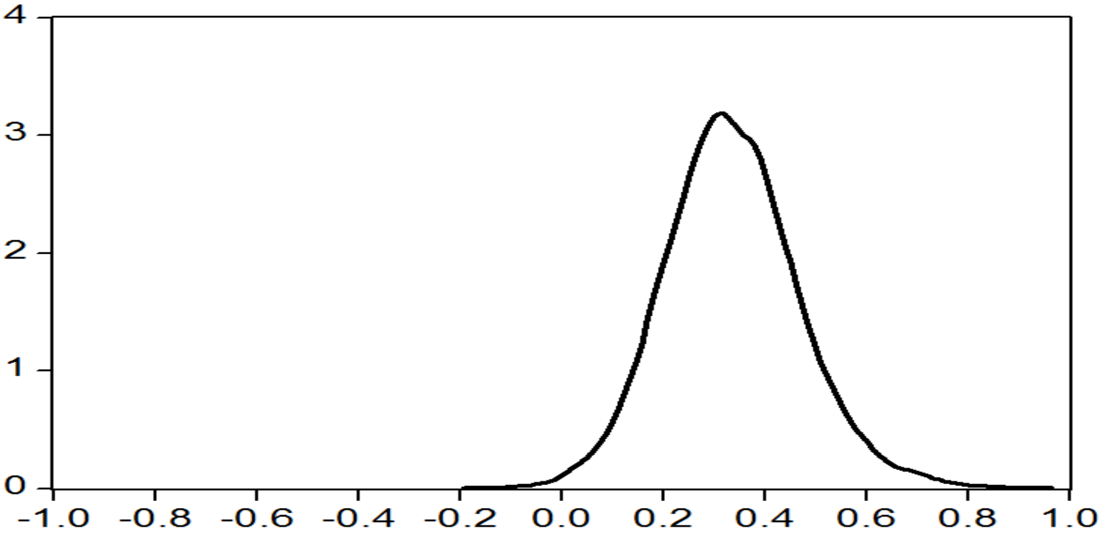
Probability of a High Interest Rate Regime

Note 1. Burn-in / Total iterations = 5,000 / 25,000
2. $S_t = 1$: a low interest regime, $S_t = 2$: a medium interest regime, $S_t = 3$: a high interest regime,

Figure 1.4.B. Posterior Distribution of Sum of AR Coefficients: AR (2) Model with Markov-Switching Mean [Garcia and Perron (1996) / 1960:1~1986:2 vs 1960:1~2008:2]



Sum of AR Coefficients (1960:1~1986:2)

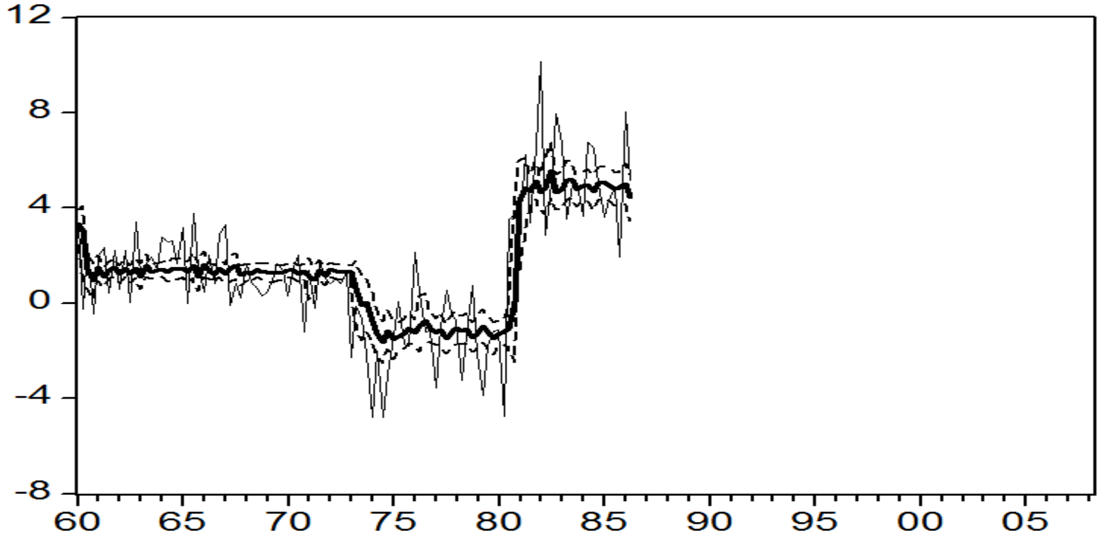


Sum of AR Coefficients (1960:1~2008:2)

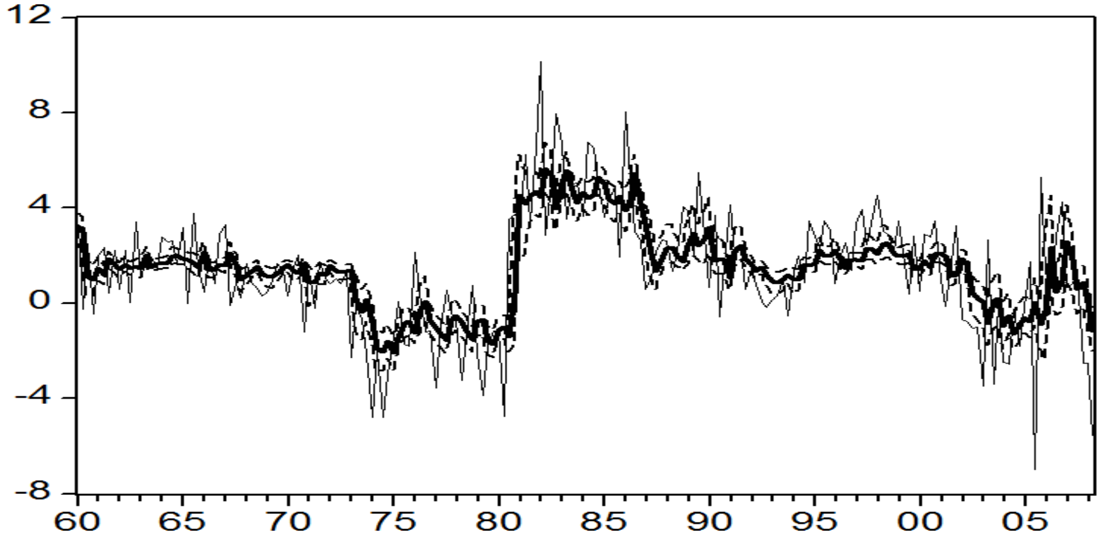
Note 1. Burn-in / Total iterations = 5,000 / 25,000

2. A stationary restriction is imposed using a rejection algorithm, which draws MCMC samples until they satisfy the stationary restriction.

Figure 1.4.C. Ex-Ante Real Interest Rate: AR (2) Model with Markov-Switching Mean [Garcia and Perron (1996) / 1960:1~1986:2 VS 1960:1~2008:2]



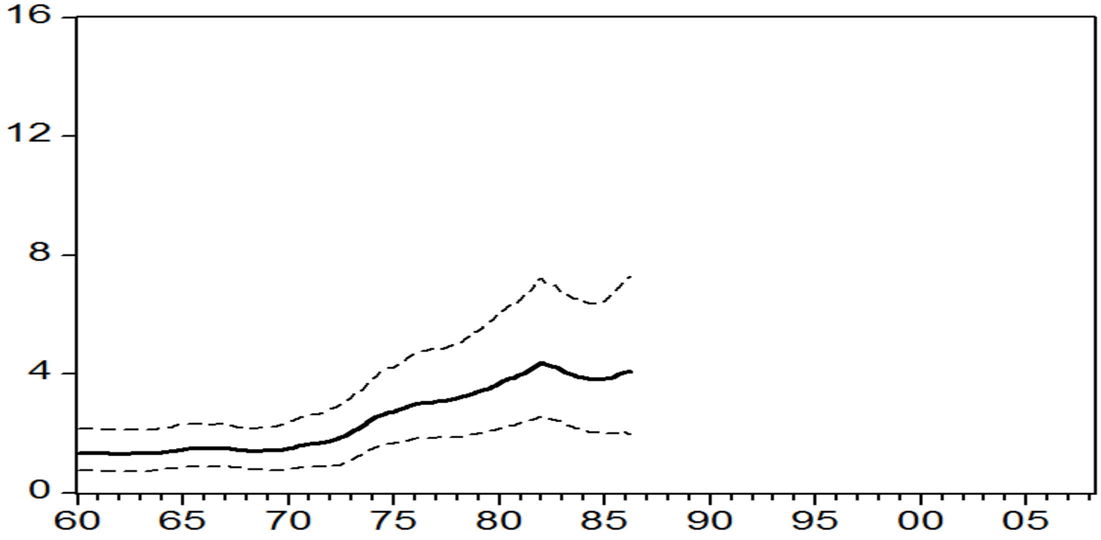
Ex-ante Real Interest Rate (1960:1~1986:2)



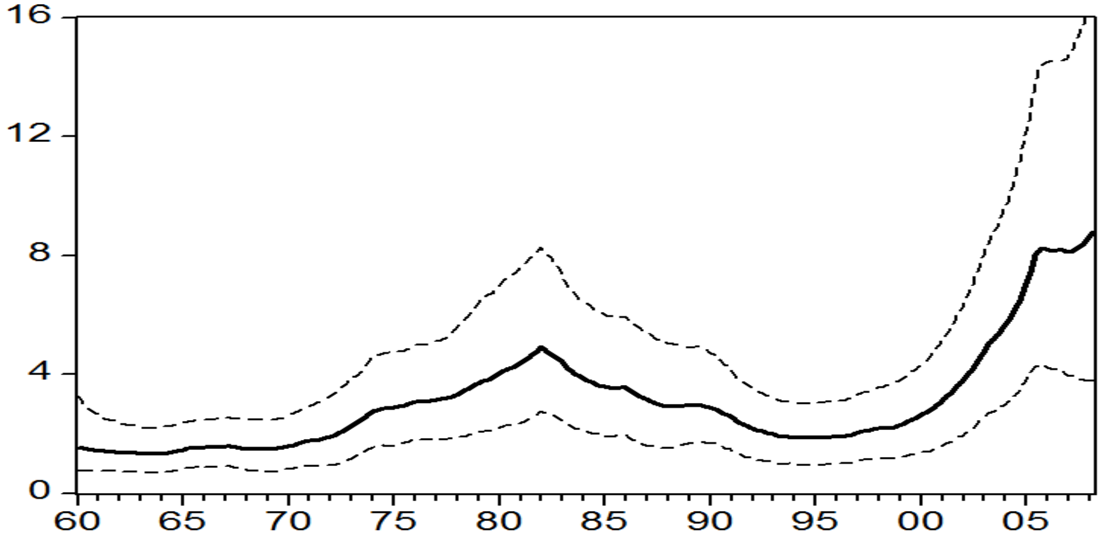
Ex-ante Real Interest Rate (1960:1~2008:2)

Note
1. Burn-in / Total iterations = 5,000 / 25,000
2. The ex-post real interest rate is depicted together with the ex-ante real interest rate in the first figure.
3. The bold and the dotted lines present posterior medians and 90 percent percentiles, respectively.

Figure 1.4.D. Volatility: AR (2) Model with Markov-Switching Mean
[Garcia and Perron (1996) / 1960:1~1986:2 vs 1960:1~2008:2]



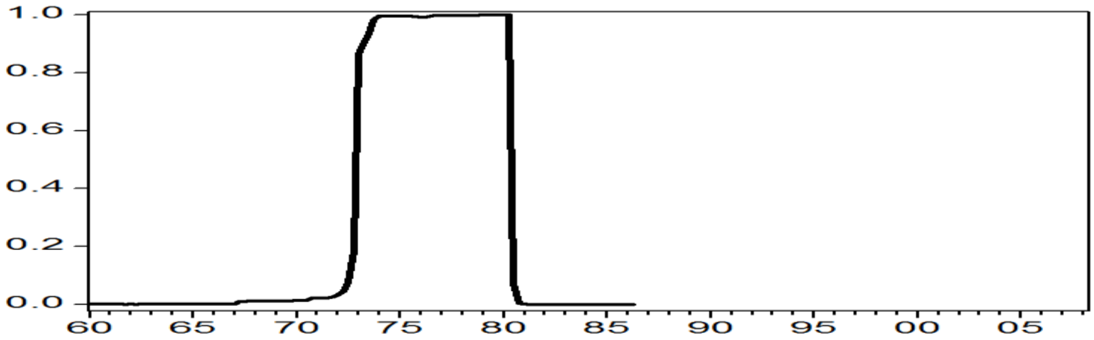
Stochastic Volatility (1960:1~1986:2)



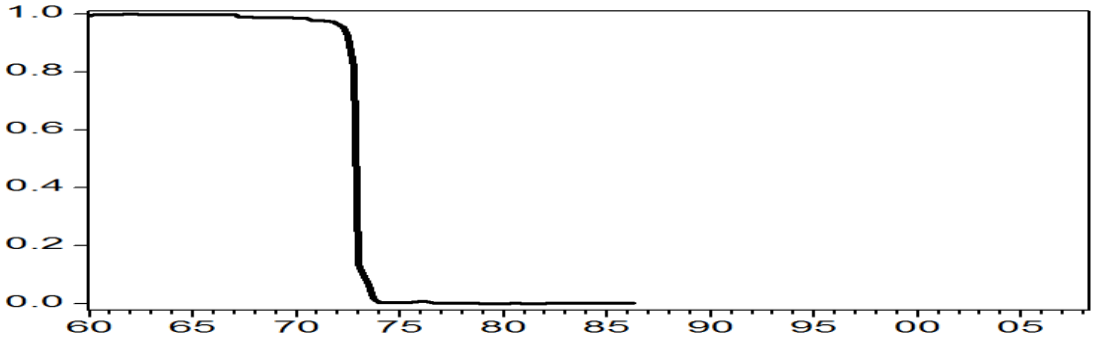
Stochastic Volatility (1960:1~2008:2)

Note 1. Burn-in / Total iterations = 5,000 / 25,000
2. The bold and the dotted lines present posterior medians and 90 percent percentiles, respectively.

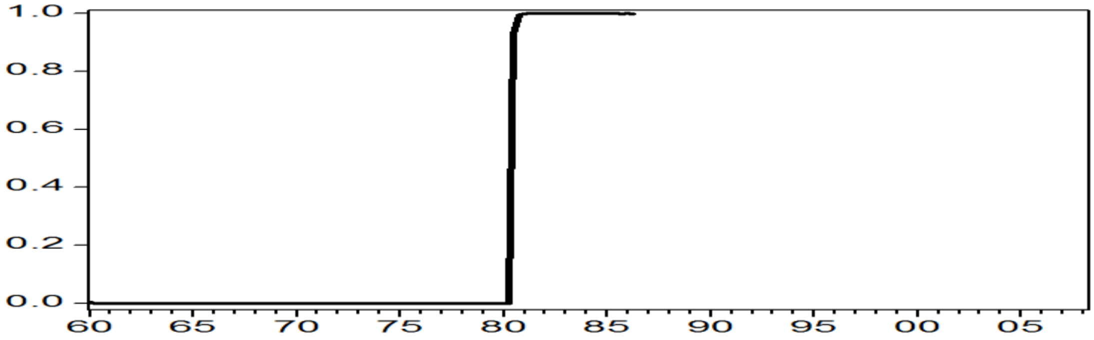
Figure 1.5.A.1. Regime Probabilities: ARMA (2,2) Model with Markov-Switching Mean [Proposed Model / 1960:1~1986:2]



Probability of a Low Interest Rate Regime



Probability of a Medium Interest Rate Regime

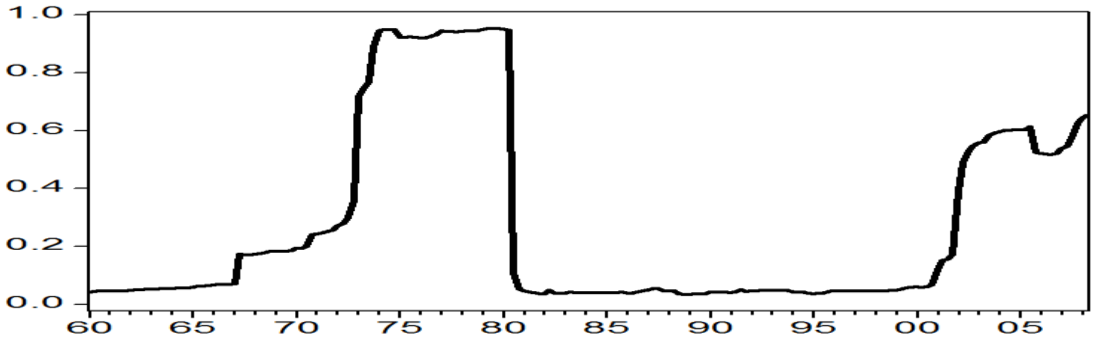


Probability of a High Interest Rate Regime

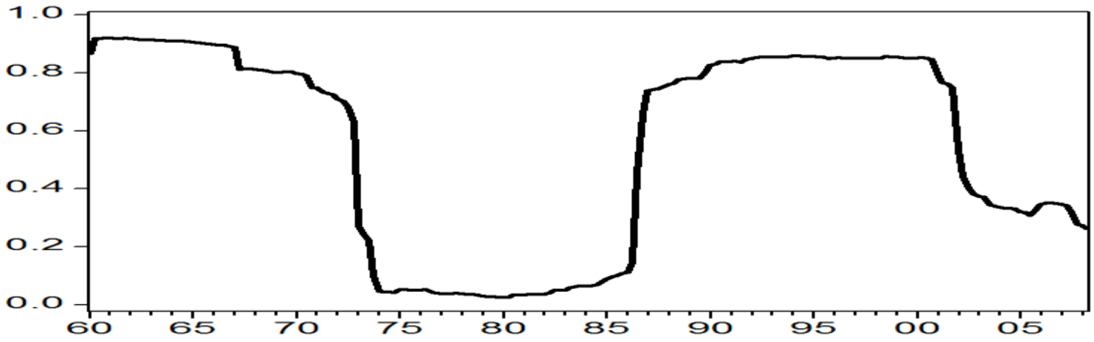
Note 1. Burn-in / Total iterations = 5,000 /25,000

2. $S_t = 1$: a low interest regime, $S_t = 2$: a medium interest regime, $S_t = 3$: a high interest regime,

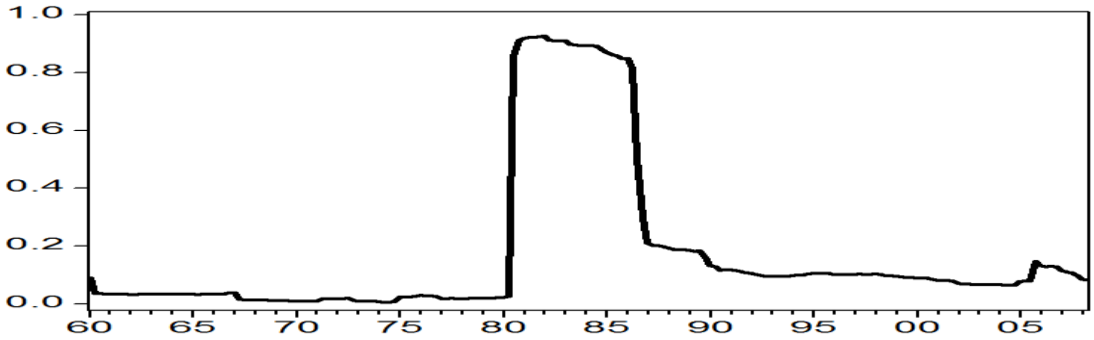
Figure 1.5.A.2. Regime Probabilities: ARMA (2,2) Model with Markov-Switching Mean [Proposed Model / 1960:1~2008:2]



Probability of a Low Interest Rate Regime



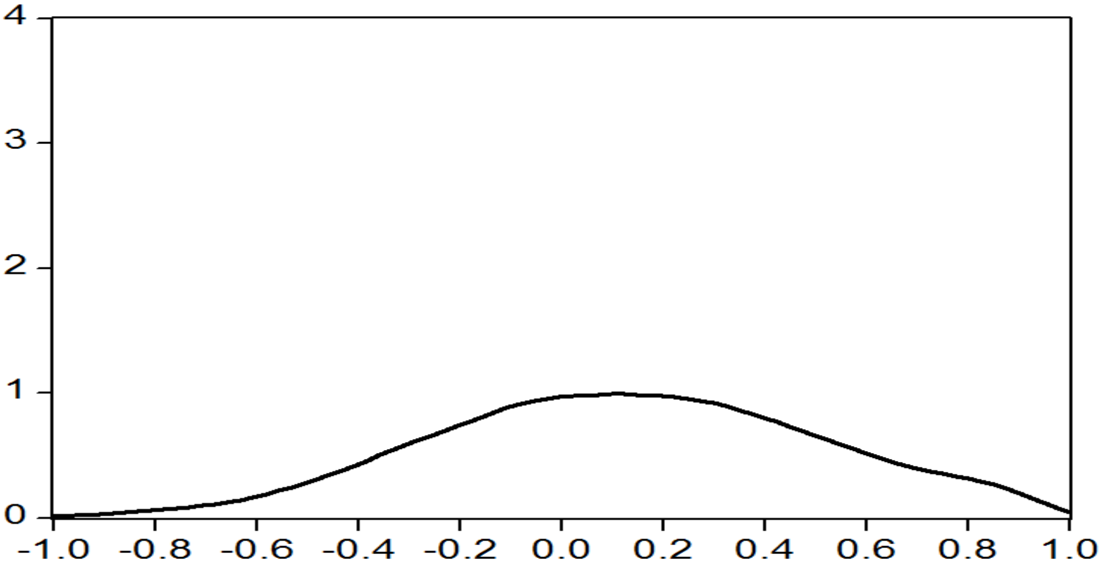
Probability of a Medium Interest Rate Regime



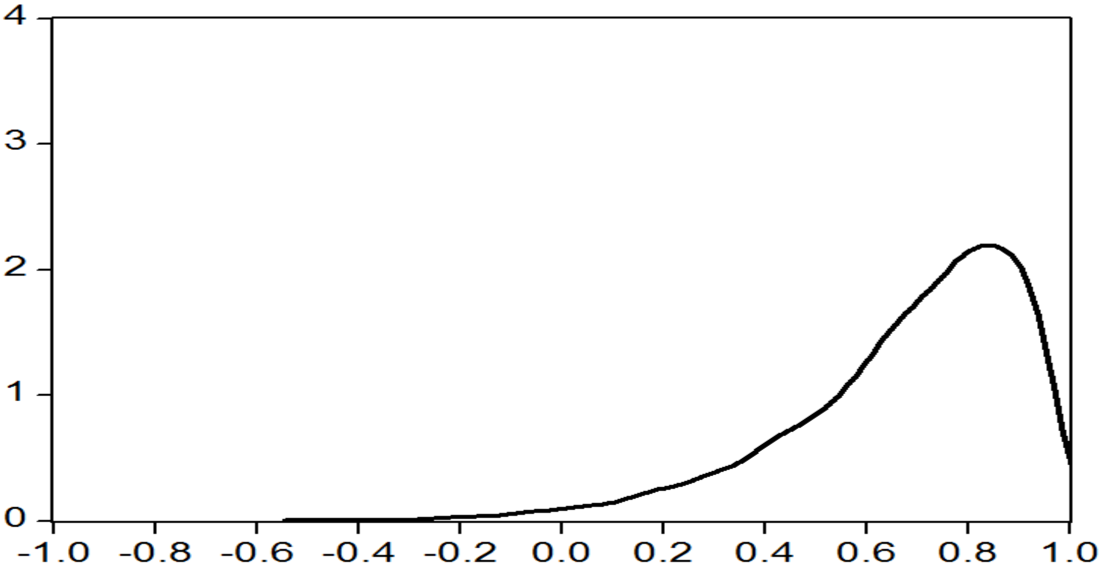
Probability of a High Interest Rate Regime

Note 1. Burn-in / Total iterations = 5,000 / 25,000
2. $S_t = 1$: a low interest regime, $S_t = 2$: a medium interest regime, $S_t = 3$: a high interest regime,

Figure 1.5.B.1. Posterior Distribution of Sum of AR Coefficients: ARMA (2,2) Model with Markov-Switching Mean [Proposed Model / 1960:1~1986:2 vs 1960:1~2008:2]



Sum of AR Coefficients (1960:1~1986:2)

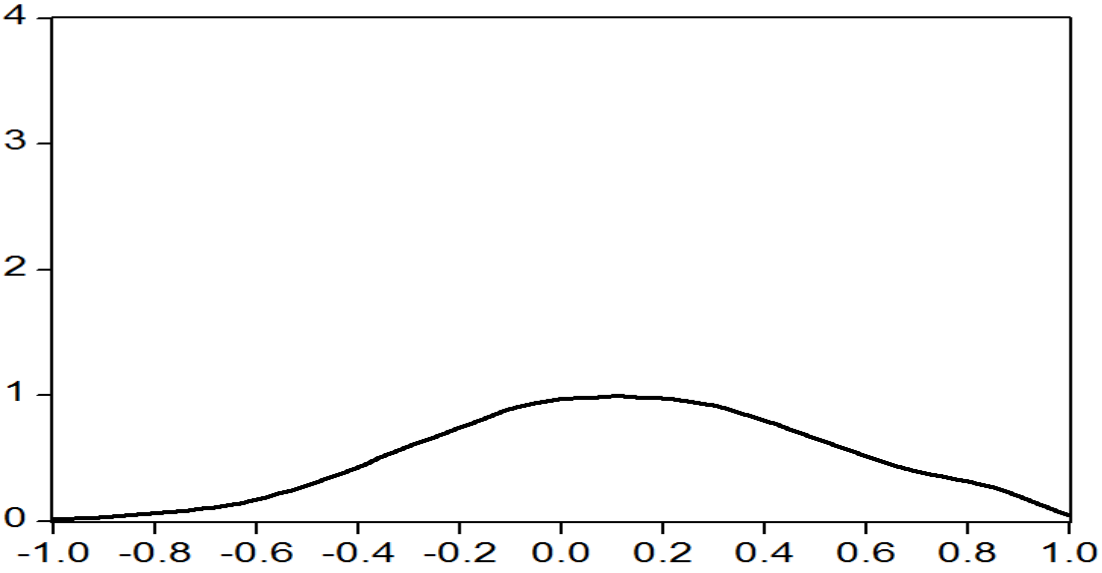


Sum of AR Coefficients (1960:1~2008:2)

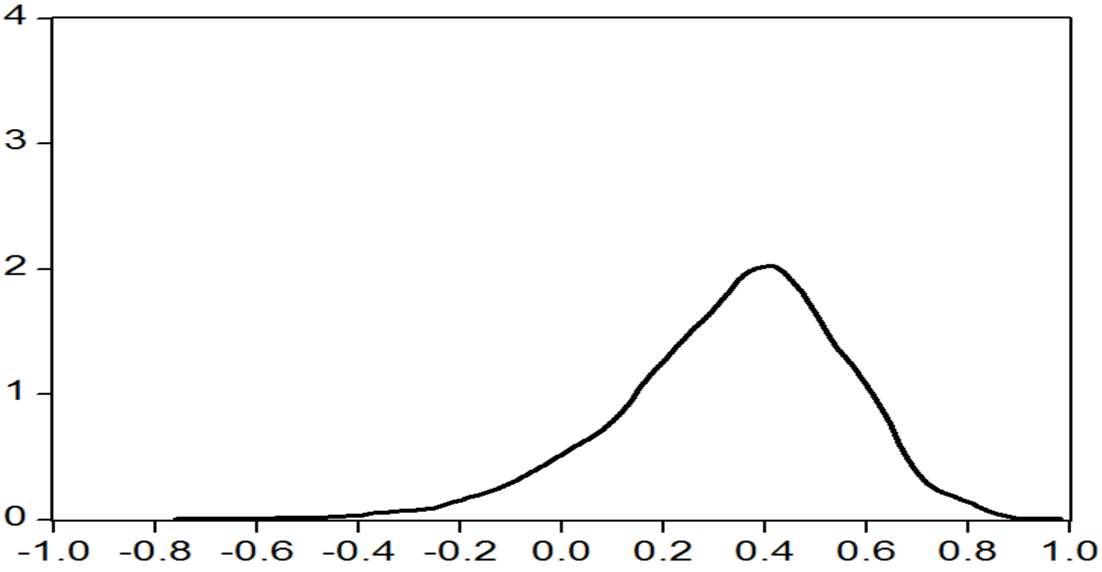
Note 1. Burn-in / Total iterations = 5,000 / 25,000

2. Stationary and invertible restrictions are imposed using a rejection algorithm and Metropolis-Hastings algorithm.

Figure 1.5.B.2. Posterior Distribution of Sum of MA Coefficients: ARMA (2,2) Model with Markov-Switching Mean [Proposed Model / 1960:1~1986:2 vs 1960:1~2008:2]



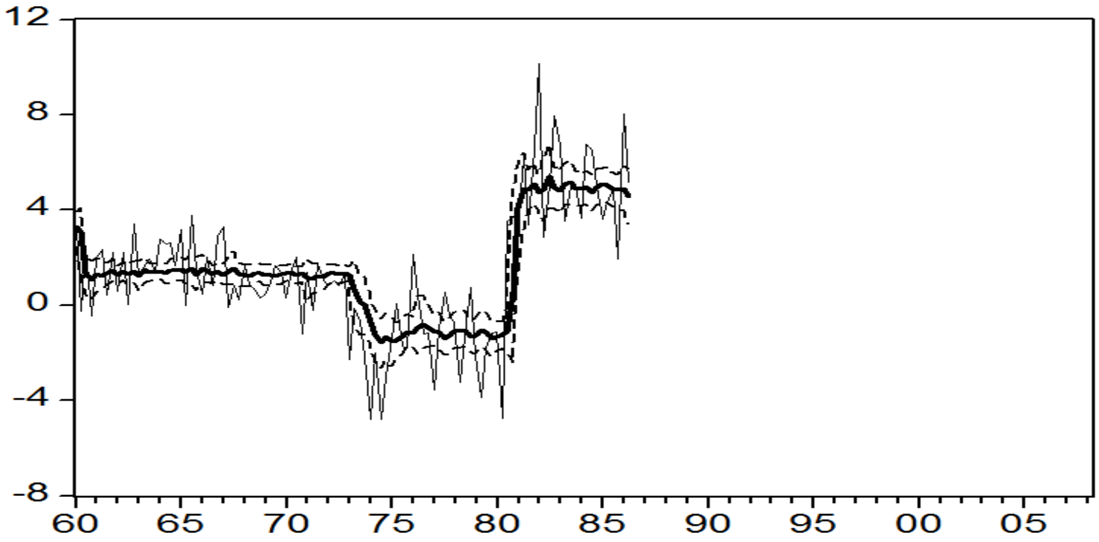
Sum of MA Coefficients (1960:1~1986:2)



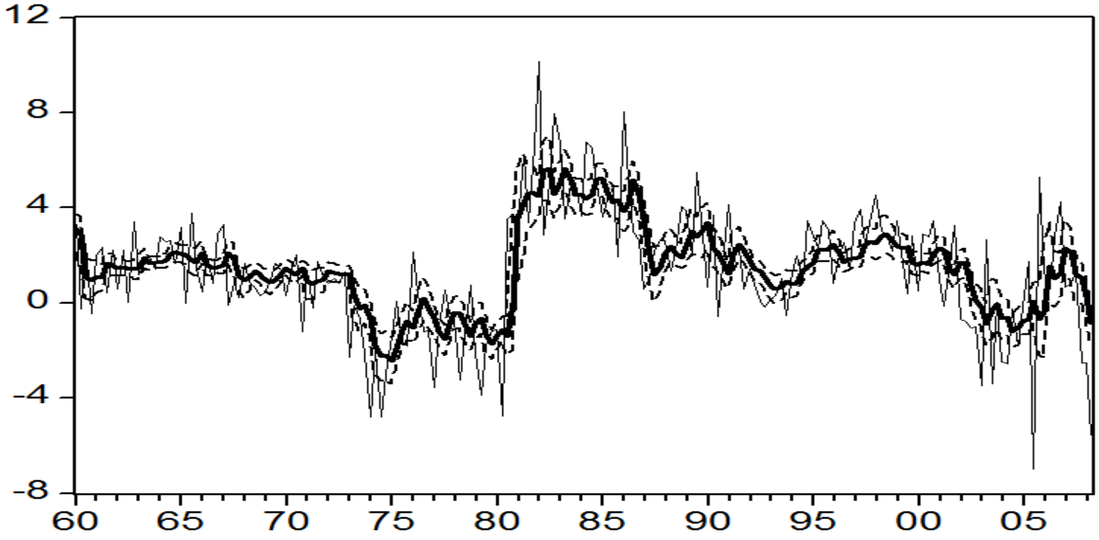
Sum of MA Coefficients (1960:1~2008:2)

Note 1. Burn-in / Total iterations = 5,000 / 25,000
2. Stationary and invertible restrictions are imposed using a rejection algorithm and Metropolis-Hastings algorithm.

Figure 1.5.C. Ex-Ante Real Rate: ARMA (2,2) Model with Markov-Switching Mean [Proposed Model / 1960:1~1986:2 vs 1960:1~2008:2]



Ex-ante Real Interest Rate (1960:1~1986:2)



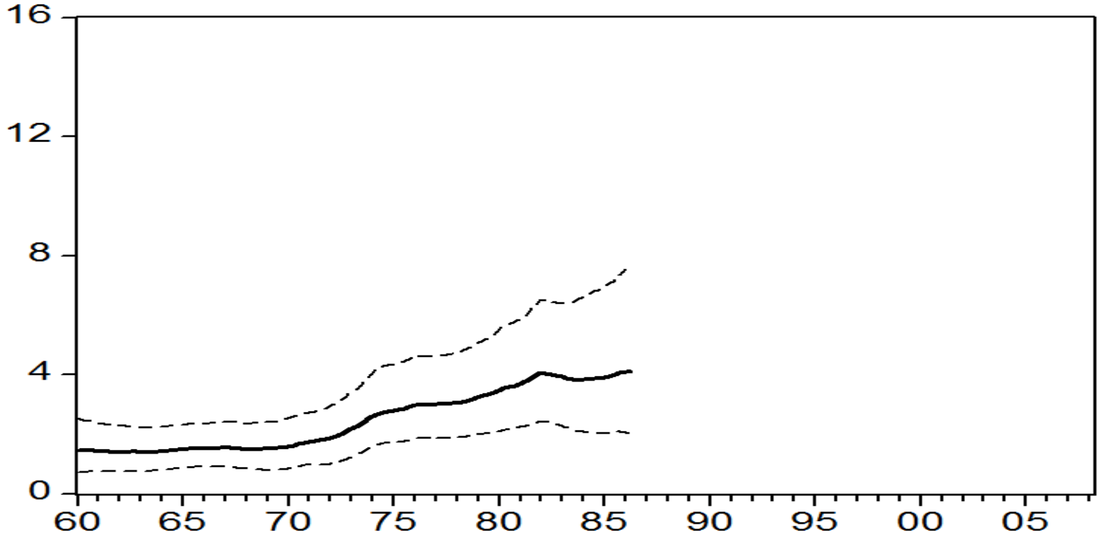
Ex-ante Real Interest Rate (1960:1~2008:2)

<u>Note</u> 1. Burn-in / Total iterations = 5,000 / 25,000

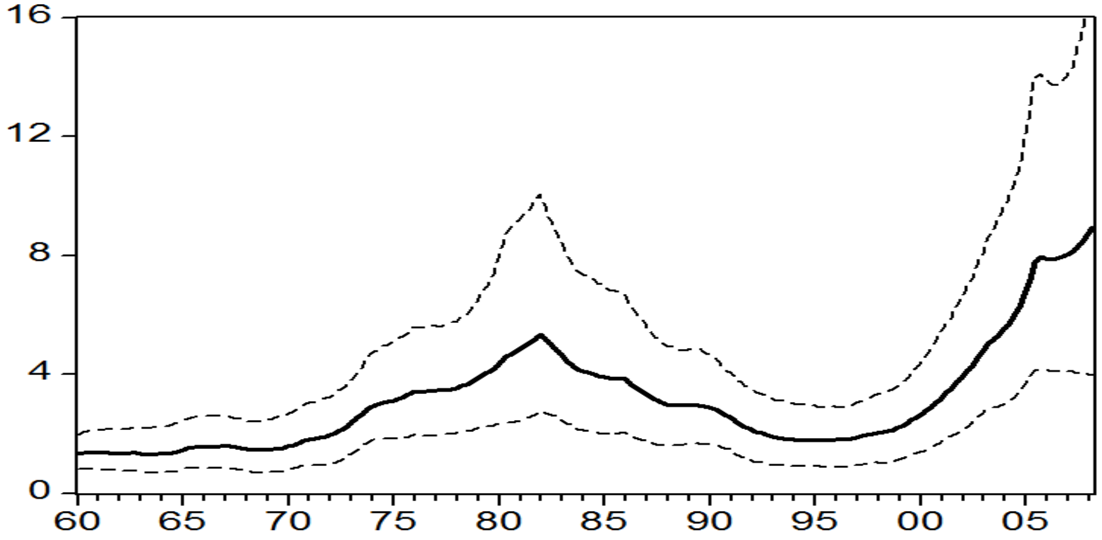
2. The ex-post real interest rate is depicted together with the ex-ante real interest rate in the first figure.

3. The bold and the dotted lines present posterior medians and 90 percent percentiles, respectively.

Figure 1.5.D. Volatility: ARMA (2,2) Model with Markov-Switching Mean [Proposed Model / 1960:1~1986:2 VS 1960:1~2008:2]



Stochastic Volatility (1960:1~1986:2)



Stochastic Volatility (1960:1~2008:2)

Note 1. Burn-in / Total iterations = 5,000 / 25,000
2. The bold and the dotted lines present posterior medians and 90 percent percentiles, respectively.

Chapter 2

The ‘Pile-up Problem’ in Trend-Cycle Decomposition of Real GDP: Classical and Bayesian Perspectives [9]

2.1. Introduction

Since the seminal work of Nelson and Plosser (1982), one of the important issues in empirical macroeconomics has been to investigate the degree of persistence in real economic activities or the relative importance of permanent and transitory shocks. This issue has been investigated in two directions. One strand of research is based on the unit root implication of real GDP and the other is based on a direct measure of the relative sizes of the stochastic trend and cyclical components of real GDP. In both strands of research, researchers provide conflicting evidence on the existence of a unit root or the relative sizes of the stochastic trend and the cyclical components in real GDP.

For example, while Nelson and Plosser (1982) report a unit root for real GDP as well as for most of the macroeconomic variables they considered, Perron (1989) argues that, by allowing for the possibility of a structural break (with known break date) in the trend function of real GDP, the null hypothesis of a unit root can be rejected. This result was criticized by Christiano (1992) and Zivot and Andrews (1992), who argue that the unit root can no longer be rejected once one incorporates uncertainty about the date of a structural break in the trend function. Cheung and Chinn (1997) apply both the unit root test and the stationarity test to post-war real GDP and find that neither test rejects the null hypothesis. They argue that the power of both tests is so low that no unambiguous conclusions can be made. That is, as also suggested by DeJong et al. (1992), the inferences based exclusively on tests for integration may be fragile. [10]

Concerning the second strand of research, in which the relative sizes of the stochastic

---

[9] This chapter is based on a joint work with Chang-Jin Kim.

[10] Furthermore, as surveyed by Murray and Nelson (2002), researchers who employ a long time series that goes back to 1870, Diebold and Senhadji (1996), Cheung and Chinn (1997), Murray and Nelson (2000, 2002), and Newbold, Leyboure, and Wohar (2001) produce mixed conclusions. Their results differ depending on how the period around the Great Depression is treated.

trend and the cyclical components is of central interest, researchers also report conflicting results. Based on estimation of ARMA models for real output growth, Nelson and Plosser (1982) and Campbell and Mankiw (1987) conclude that transitory shocks are relatively unimportant in explaining the dynamics of real output, while permanent shocks must dominate. On the contrary, within an unobserved-components model (hereafter, UC model) framework in which the permanent and transitory shocks are assumed uncorrelated, Clark (1987) reports evidence that a significant portion of real GDP is explained by the cyclical component. This result is then challenged by Morley et al. (2003), who show that the stochastic trend explains most of the variations in real GDP once the assumption of zero correlation between the permanent and the transitory shocks is dropped. They further show that the decomposition of real GDP based on an ARIMA(2,1,2) model (i.e., the Beveridge-Nelson decomposition, 1981) and that based on an unobserved-components model are identical [11].

Recently, by allowing for a structural break in the long-run mean growth rate of real GDP in the mid-1970s within Morley et al's (2003) framework, Perron and Wada (2009) show that all variations in real GDP are ascribed to the cyclical component. In particular, by casting Morley et al.'s (2003) unobserved components model into a reduced-form ARIMA(2,1,2) model, they show that the point estimates of the moving-average coefficients sum to unity, which they interpret as an indication that the first-differences of real GDP are over-differenced (see, for instance, Plosser and Schwert (1977)). For example, if the log of real GDP is a trend stationary process, taking a first difference of it would result in a unit root in the moving-average part of the ARIMA model. As demonstrated by Sargan and Bhargava (1983) within an MA(1) model with a moving-average parameter θ, however, the occurrence of a maximum of the likelihood function at $\theta = 1$ is insubstantial evidence for 'over-differencing'. This is because, in small samples, there exist reasonably high probabili-

---

[11] This might not be true within Bayesian framework. Luo and Startz (2013) investigate the relative size of the stochastic trend and the cyclical components based on structural break UC models. In their work, the cyclical component is dominant regardless of model specifications which is inconsistent with our emperical results based on an reduced-form ARIMA(2,1,2) model. This casts a doubt that the decomposition of real GDP based on an reduced-form ARIMA(2,1,2) model and that based on a corresponding UC model are identical within Bayesian framework. Investigating the difference would be interesting but since our focus is limited to the pile-up problem in ARMA(p,q) models, we leave this as a future research topic.

ties that θ may be estimated to be one even when the true value of θ is less than one. This is known as the 'pile-up problem' in the literature on MA models.

Within the classical framework, the pile-up problem was originally analyzed by Kang (1975) and Davidson (1981), for the cases of simple moving average models. Ansley and Newbold (1980) and Sargan and Bhargava (1983) extend the analysis to the case of general ARMA models and the regression models with MA disturbances, respectively. In particular, based on both theoretical derivations and simulation analysis, Sargan and Bhargava (1983) show that in finite samples the probabilities of the pile-up problem depends on the choice of the regressors. They show that in small samples the probabilities of the pile-up problem are substantially increased with an inclusion of an intercept term or other regressors.

Within the Bayesian framework, however, the nature of the pile-up problem has not been fully investigated. For an MA(1) model without an intercept term, DeJong and Whiteman (1993) show that, while the sampling distributions of the maximum likelihood estimator of θ ($\hat{\theta}_{ML}$) piles up at unity when the true parameter is near unity, the (Bayesian) flat-prior posterior distributions of θ do not pile up regardless of the parameter's proximity to unity. They also show that posterior distributions of peak at the maximum likelihood estimates. These are illustrated by comparing the sampling distribution of $\hat{\theta}_{ML}$ and the posterior distribution of θ, which are obtained from the joint distribution of $\hat{\theta}_{ML}$ and θ constructed based on Monte Carlo simulations. These results are taken by DeJong and Whiteman (1993) as a rationale for favoring the Bayesian approach over the classical approach. [12]

In this chapter, we estimate Perron and Wada's (2009) model by applying the Bayesian approach. [13] Surprisingly, the trend-cycle decomposition of real GDP implied by the Bayesian parameter estimates turn out to be very different from that implied by Perron

---

[12] Smith and Naylor (1987) develop maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution, and they show that the two sets of estimators are very different. They also show that there are practical advantages to the Bayesian approach.

[13] Within the Bayesian framework, DeJong and Whiteman (1991) show that unit AR roots are implausible for a wide range of annual macroeconomic time series considered by Nelson and Plosser (1982). However, Murray and Nelson (2002) argue that, if the effect of the shocks during the Great Depression is controlled for, real shocks persist indefinitely. Murray and Nelson (2000) further argue that, "while more data is preferred to less in a homogeneous time series, the experiments ... show that heterogeneity generally causes severe distortions of test size, " suggesting that empirical evidence based on more homogeneous post-war data may be more reliable. We thus focus our analysis on the post-war data set of Perron and Wada (2009).

and Wada's (2009) maximum likelihood estimates, even with reasonably non-informative priors. That is, most of the variations in real GDP can be explained by the stochastic trend component, consistent with the implications of Nelson and Plosser (1982) and Morley et al. (2003). Unlike the predictions of DeJong and Whiteman (1993), the posterior mode for the sum of moving-average parameters do not peak at its maximum likelihood estimate of one. Instead, the posterior mode is close to the local maximum of the likelihood function in the invertible region, even though there exists a non-negligible probability mass near the non-invertible boundary of one.

In the case of a reasonably flat prior, a conventional wisdom is that Bayesian inference may not be very different from classical inference, as the likelihood dominates the posterior density. This chapter confirms that the ARMA model of real GDP estimated by Perron and Wada (2009) is an example in which this conventional wisdom does not apply. We show such dramatically different results based on the maximum likelihood and the Bayesian approaches stem from the differences in how the nuisance parameters are handled between the two approaches, especially when the parameter estimate of interest is dependent upon the estimates of the nuisance parameters for small samples. For the maximum likelihood approach, as the number of the nuisance parameters increases, we have higher probability that the moving-average root may be estimated to be one even when its true value is less than one, spuriously indicating that the data is over-differenced. However, the Bayesian approach is relatively free from this pile-up problem, as the posterior distribution is not dependent upon the nuisance parameters.

We also apply the Bayesian approach to an ARIMA(2,1,2) model for the log of real GDP, by relaxing the assumption of a known break date for the mean growth rate. A reduction in the variance of the shocks to real GDP, namely the Great Moderation (Kim and Nelson (1999) and McConnell and Perez-Quiros (2000)), is also incorporated. Our results suggest that the posterior mean and mode of $\theta_1 + \theta_2$ are 0.095 and 0.443, respectively, which are further away from unity than in the case of a known structural break date. However, the probability mass at unity almost disappears for the posterior distribution, unlike in the case of known break dates. This suggests that, with the inclusion of the break date uncertainty, we have even less probability of post-war U.S. real GDP being a trend stationary process than in the case of a known break date. Furthermore, the implied cyclical component is noisy and small in magnitude, with most variations in real GDP being explained by the stochastic trend component. That is, even after taking breaks with uncertain break dates, the implications of Nelson and Plosser (1982) and Morley et al. (2003) on trend-cycle decomposition continue to hold within the Bayesian framework, which is relatively free from the pile-up problem.

The chapter is organized as follows. In Section 2.2, we show that results from Bayesian estimation of Perron and Wada's (2009) model are very different from those from maximum likelihood estimation. In Section 2.3, we discuss the nature of the classical pile-up problem, and present a simulation study showing that Perron and Wada's (2009) results may be due to the classical pile-up problem. In Section 2.4, we provide an answer to the question of why the results from the classical and Bayesian approaches are so different. In particular, we provide a discussion of why the Bayesian approach may be relatively free from the pile-up problem. In Section 2.5, we apply the Bayesian approach to an extended ARIMA(2,1,2) model of real GDP, in which we incorporate a structural break in the variance of shocks (Great moderation) and in the long-run mean growth rate with uncertain break dates. Section 2.6 concludes the paper.

2.2. Preliminaries: Classical and Bayesian Perspectives for Trend-Cycle Decomposition of Real GDP [1947:1 1998:2]

Harvey (1985), Clark (1987), and Morley et al. (2003), among others, consider the following unobserved components model of real GDP:

$$\begin{gathered} y_t = x_t + z_t, \\ x_t = \mu_t + x_{t-1} + v_t \qquad (2.1) \\ \phi(L) z_t = \epsilon_t \\ \begin{bmatrix} v_t \\ \epsilon_t \end{bmatrix} \sim i.i.d.N \begin{bmatrix} \sigma_v^2 & \rho\sigma_v\sigma_\epsilon \\ \rho\sigma_\epsilon\sigma_v & \sigma_\epsilon^2 \end{bmatrix}, \end{gathered}$$

where y_t is the log of real GDP; x_t is a stochastic trend component; and z_t is a cyclical component with all the roots of $\phi(L) = 0$ lying outside the complex unit circle.

Literature suggests that different assumptions about the dynamics of the long-run mean growth rate μ_t or a restriction on the correlation coefficient ρ can lead to different trend-cycle decompositions. For example, with a zero restriction on the ρ parameter and a random walk specification for μ_t, Clark (1987) estimates the cyclical component (z_t) to be highly persistent and shows that a significant portion of real GDP is explained by this component. By assuming that μ_t is constant and allowing for a possibility that ρ may be non-zero, Morley et. al (2003) estimates the cyclical component to be noisy and considerably smaller than that in Clark (1987). On the contrary, by modeling μ_t as a constant interrupted by a permanent change occurring in 1973:1, Perron and Wada (2009) estimate the variance of the permanent shocks σ_v^2 to be zero, suggesting that real GDP is a trend stationary process.

As Morley et al. (2003) present, one potential difficulty in estimating the above unobserved components model is that it is identified only when z_t is autoregressive of order higher than one. Furthermore, when they estimate the model with an AR(2) dynamics for z_t, they show that the confidence intervals for the ρ parameter are so large that various trend-cycle decompositions are possible depending on which value of the ρ parameter is chosen within the confidence interval. As they suggest, one way to overcome these difficulties is to estimate a reduced-form ARIMA model for real GDP and employ the Beveridge-Nelson (1981) decomposition procedure. For example, if we assume that $\phi(L) = 1 - \phi_1 L - \phi_2 L^2$ and μ_t is a constant with a permanent shift in 1973:1, a reduced-form ARIMA model considered by Perron and Wada (2009) is given by:

<u>Perron and Wada's (2009) Model</u>

$$\begin{gathered}
\Delta y_t = \mu_0 + \mu_1 D_t + \Delta y_t^*, \\
\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, \qquad (2.2) \\
D_t = 0 \;\; for \;\; t \leq 1973:I; \;\; and \; D_t = 1, \;\; otherwise. \\
e_t \sim i.i.d.N(0, \sigma_e^2),
\end{gathered}$$

where σ_e^2 and the moving-average parameters θ_1 and θ_2 are functions of ϕ_1, ϕ_2, σ_v^2, σ_ϵ^2, and ρ.

It is easy to show that a unit root in the moving average part of the above ARIMA model is equivalent to the case of $\sigma_v^2 = 0$ in the UC model of (2.1). In this case Δy_t is over-differenced, and y_t is a trend stationary process. Perron and Wada (2009) estimate the above ARIMA model as well, and report that the maximum likelihood estimates of the moving-average parameters sum to unity, which is consistent with their estimate of $\sigma_v^2 = 0$ for the UC model in (2.1). We replicate Perron and Wada's (2009) results by employing the same model in (2.2) and data set (quarterly real GDP, 1947:1 to 1998:2) as used by them. [14] Table 2.1.A reports the results, from which we note that a local maximum exists within the invertibility region of the moving average parameters as well as the global maximum at $\theta_1 + \theta_2 = 1$.

In this section, we consider Bayesian inference of the model in (2.2) and the results are compared to those based on classical inference by Perron and Wada (2009). In the case of a flat prior, a conventional wisdom is that Bayesian inference may not be very different from classical inference, as the likelihood dominates the posterior density. In Table 2.1.B, the posterior moments of the parameters are presented. Surprisingly, estimation results based on the Bayesian approach with reasonably non-informative priors are very different from those based on the maximum likelihood method. The posterior mean and the posterior mode of $\theta_1 + \theta_2$ turn out to be 0.267 and 0.483, respectively, as opposed to its maximum likelihood estimate of unity. An interesting finding is that the parameter values at the posterior modes are very close to those at the local maximum of the log likelihood function. However, from the posterior distribution of $\theta_1 + \theta_2$ depicted in Figure 2.1.A, we cannot rule out the possibility that $\theta_1 + \theta_2 = 1$, as there exists a non-negligible probability mass at unity. [15] The probability is about 5%.

At each iteration of the Markov Chain Monte Carlo (MCMC) algorithm, we apply the Beveridge-Nelson (1981) decomposition procedure to get the cyclical component of real GDP. In Figure 2.1.B, the estimates of the cyclical component from this procedure and that implied by Perron and Wada's (2009) maximum likelihood estimation are compared. The

---

[14] This data set was originally used in Morley et al. (2003).

[15] As we impose the constraint that $|\theta| < 1$ when estimating the model, we cannot actually have a probability mass at unity. Throughout the paper, when $Pr[0.995 < \theta < 1]$ is non-zero for the posterior distribution, we state that there exists a probability mass at unity.

corresponding trend components are also depicted against the log of the real GDP series. While variations in real GDP are explained mostly by the cyclical components within the classical framework, they are explained mostly by the stochastic trend component within the Bayesian framework. The impulse-response functions ($\frac{\partial y_{t+j}}{\partial e_t}$) depicted in Figure 2.1.B further confirm this point. Within the Bayesian framework, the posterior mode of the long-run impulse-response coefficient ($lim_{j\to\infty}\frac{\partial y_{t+j}}{\partial e_t}$) is 1.391 with the 90% highest posterior density (HPD) interval being [0.943, 1.891].

An important question then is: "Why do the classical and the Bayesian approaches produce such strikingly different estimates of the ARMA parameters and trend-cycle decompositions?" We believe that one of the keys to the answer to this question lies in the 'pile-up problem' that the maximum likelihood estimator of the moving-average parameter is subject to. Another key is in DeJong and Whiteman (1993), who demonstrate that the posterior distributions of the moving-average parameter from an MA(1) model without intercept do not pile up at unity even when the true moving-average root is close to unity. In the next two sections, we provide an in-depth analysis of the pile-up problem within both the classical and the Bayesian frameworks. In particular, we are interested in knowing whether or not the results of DeJong and Whiteman (1993) also hold for general ARMA models of the form given in (2.2), in which a structural break is incorporated in the mean. Whether or not, in general, the Bayesian approach suffers less from the pile up problem than the classical approach is another issue we investigate. If this is the case, we could reasonably confer more credibility to the results based on the Bayesian approach.

2.3. The Nature of the Pile-up Problem within the Classical Framework: The Effect of Incorporating a Structural Break in Mean

Many authors investigate the finite sample properties of the maximum likelihood estimator of the moving average parameter in an MA(1) model, especially when the moving average parameter is close to unity. Following the initial work of Kang (1975), several authors including Sargan and Bhargava (1983), Anderson and Takemura (1986) and Tanaka and Satchell (1989) show that the process can be estimated to be noninvertible with a unit root even when the true process is invertible, with a considerably high probability in a finite sample. This is referred to as the pile-up problem. [16]

In order to get an intuition about why the pile-up problem occurs, consider the following MA(1) model: [17]

$$y_t = e_t - \theta e_{t-1}, \qquad e_t \sim i.i.d\ N(0, \sigma^2), \tag{2.3}$$

the first-order autocorrelation (ρ_1) of which is given by:

$$\rho_1 = -\frac{\theta}{1+\theta^2}. \tag{2.4}$$

From equation (2.4), it can be shown that two parameter sets, i.e., (θ, σ^2) and $(\frac{1}{\theta}, \sigma^2)$, induce an identical auto-covariance structure and thus an identical log likelihood value, which suggests that the above model is not identified. This identification problem can be handled by restricting the parameter space to $|\theta| \leq 1$, including an 'invertibility region' and unity. Then, we have the restriction that $|\rho_1| \leq 0.5$. However, in case the sample autocorrelation turns out to be greater than 0.5. Then, "the moment estimator of θ obtained by inverting (4) can be defined by stipulating that the estimate is set to 1 ... the estimator takes the value 1 with positive probability" (Davidson, 1981, p. 926). For maximum likelihood estimation of θ, Davidson (1981) further explains that the distribution function of the estimator of θ must possess discontinuities or 'steps' at unity, suggesting that the estimator takes the value of 1 with positive probability. [18]

---

[16] Asymptotic properties of $\hat{\theta}_{ML}$ are derived in Davis and Dunsmuir (1996), and Davis et al. (1995) for the case where θ is close or equal to 1. They show that the conventional central limit theorem does not work in such a case.

[17] Shephard and Harvey (1990) and Shephard (1993) investigate the above pile-up problem within unobserved components models that consist of a random walk and white noise processes. As the reduced-form for this is an IMA(1,1) model in equation (3), the pile-up problem within the unobserved components model is equivalent to the probability of estimating a zero variance for the shocks to the random walk component.

[18] More rigorously, the profile log likelihood ($\Lambda(\theta)$), obtained by concentrating out σ^2, satisfies the property that $\Lambda(\theta) = \Lambda(\theta^{-1})$. Stock (1994) shows that $\partial\Lambda/\partial\theta|_{\theta=1} = 0$ and therefore, Λ will have a local maximum at $\theta = 1$ if $\partial^2\Lambda/\partial\theta^2|_{\theta=1} < 0$. Early literature on this issue, including Kang (1975) and Sargan and Bhargava (1983), derive this probability to be non-zero in a small sample. For more details and more comprehensive survey on the pile-up problem, readers are referred to Stock (1994).

Sargan and Bhargava (1983) further investigate the nature of the pile-up problem within regression models with first-order moving average errors. They show that the probabilities of the pile-up problem are substantially increased in the regression cases and can be quite high even for small values of the moving average parameter for the error term. In particular, they show that when the regressors are trending, the probability of the pile-up problem is "very" high.

In this section, in view of Perron and Wada's (2009) model in (2.2), we show by simulation study that the probability of the pile-up problem can also be "very" high when there is a structural break in the mean of an MA process or an ARMA process. For this purpose, we consider the following four data generating processes:

Model #1: MA(1) without Intercept

$$y_t = e_t - \theta e_{t-1}, \quad e_t \sim i.i.d.N(0, \sigma^2) \tag{2.5}$$

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

$$[\theta = 0.8, \quad \sigma^2 = 1]$$

Model #2: MA(1) with Intercept

$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d.N(0, \sigma^2) \tag{2.6}$$

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

$$[\theta = 0.8, \quad \sigma^2 = 1, \quad \mu = 1]$$

Model #3: MA(1) with a Structural Break in Intercept

$$y_t = \mu + \mu_1 S_t + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\ N(0, \sigma^2), \tag{2.7}$$

$$S_t = 0, \quad for \quad t \leq \frac{T}{2}; \quad S_t = 1, \quad otherwise,$$

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

$$[\theta = 0.8, \quad \sigma^2 = 1, \quad \mu = 1, \quad \mu_1 = -0.3]$$

Model #4: ARMA(1,1) with a Structural Break in Intercept

$$y_t = \mu_0 + \mu_1 S_t + u_t,$$

$$u_t = \phi u_{t-1} + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\ N(0, \sigma^2), \tag{2.8}$$

$$S_t = 0, \quad for \ \ t \leq \frac{T}{2}; \quad S_t = 1, \quad otherwise,$$

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

$$[\theta = 0.8, \quad \sigma^2 = 1, \quad \mu = 1, \quad \mu_1 = -0.3 \ \ \phi = 0.3]$$

For each of the above 4 models, we generate 5,000 sets of data and apply the maximum likelihood estimation procedure to the generated data sets, in order to get the sampling distributions of $\hat{\theta}_{ML}$ and to calculate the probabilities of the pile-up problem. We consider three different sample sizes (T = 50, 100, and 200), and we note that the sample size of 200 is close to the actual sample size (T = 205) for the data employed by Perron and Wada (2009). We assign $\theta = 0.8$ throughout our simulation study. In general, if data are generated with $\theta > 0.8$, for example, $\hat{\theta}_{ML}$ would be subject to more severe pile-up problem, and vice versa. Maximization of the log likelihood function is performed using the Gauss optimization package, using the true values of the parameters as initial values. For the numerical optimization, we impose the constraint that $|\theta| < 1$. Thus, following DeJong and Whiteman (1993), we report $Pr[0.995 < \hat{\theta}_{ML} < 1]$ as the probability of the pile-up problem. The sampling distributions of $\hat{\theta}_{ML}$ are shown in Figure 2 and the results are tabulated in Table 2.2.

For an MA(1) model without an intercept term, $\hat{\theta}_{ML}$ is subject to the pile-up problem when $T = 50$, and the problem almost disappears when T is increased to 100 or 200. With the inclusion of an intercept term, even though the probability of the pile-up problem increases substantially for T = 50 or 100, the problem almost disappears when T = 200. With a structural break in the intercept term for an MA(1) model, the probability of the pile-up problem is almost 1 when $T = 50$, and it decreases to as low as 4.4% when $T = 200$. For an ARMA(1,1) model with a structural break in the intercept term, however, the probability of the pile-up problem remains as high as 23.6% even when T is increased to 200. [19]

As the model gets more complicated with additional nuisance parameters (i.e. the parameters other than θ), the pile-up problem gets worse. We can easily conjecture that for an ARMA(2,2) model with a structural break in the intercept term, the moving average parameters (or the moving average roots) would be subject to more severe pile-up problems than for an ARMA(1,1) model considered in our simulation study. This suggests that we cannot rule out the possibility that the maximum likelihood estimation of Perron and Wada's (2009) model in equation (2.2) may be subject to the pile-up problem.

In order to consider the implication of the pile-up problem for the maximum likelihood estimation of Perron and Wada's (2009) model in (2.2), we conduct an additional Monte Carlo experiment. For this purpose, we generate 5,000 sets of data according to the data generating process in (2.2), by assuming that the posterior modes reported in Table 1.B are the true parameter values. The sample size is set to be the same as that ($T = 205$) employed by Perron and Wada (2009). We then apply the maximum likelihood estimation procedure to the generated data sets. The sampling distribution of the estimator for the sum of the moving average parameters is shown in Figure 2.2. Clearly, the estimator piles up at unity, and the probability of the pile-up problem is calculated to be as high as 0.4! [20]

2.4. The Nature of the Pile-up Problem within the Bayesian Framework: Why Is It So Different from That within the Classical Approach?

Concerning a solution to the classical pile-up problem, Gospodinov (2002) proposes a bootstrap method for obtaining median unbiased estimators and confidence intervals for the moving average parameter in an MA(1) model. In an unobserved components model that consists of a random walk component and a stationary component, Stock and Wat-

---

[19] Even though we do not report the results here, our simulation study also show that the pile-up problem for ARMA models gets worse as we assign the value of the autoregressive parameter (ϕ) closer to that of the moving average parameter θ when generating data. In such cases, we conjecture that there exist higher probabilities for the cancellation of the estimated MA and AR roots, and this tends to make the pile-up problem worse.

[20] We also conducted the same Monte Carlo experiment by generating data using the parameters values at the local maximum of the likelihood function reported in Table 2.1.A. The results were almost the same.

son (1998) develop asymptotically median unbiased estimators and confidence intervals for the variance of the permanent shocks, by inverting quantile functions of regression-based parameter stability test statistics which are computed under the constant-parameter null.

However, the issue of the pile-up problem does not seem to have been investigated rigorously within the Bayesian framework. The only Bayesian paper on the pile-up problem that we know of is DeJong and Whiteman (1993), who show that the posterior distributions of θ do not pile up at unity regardless of the proximity of θ to unity.

In what follows, we replicate their results and investigate whether or not their argument can be extended to a general ARMA model with a structural break in the intercept term, such as the one employed by Perron and Wada (2009).

2.4.1. The Sampling Distribution of $\hat{\theta}_{ML}$ and the Posterior Distribution of θ: MA(1) Model without Intercept

For an MA(1) model without an intercept term in (2.5) and also given below,

$$y_t = e_t - \theta e_{t-1},\ e_t \sim i.i.d.N(0, \sigma^2), \quad t = 1, 2, ..., T$$

we follow DeJong and Whiteman's (1993) procedure to obtain the joint frequency distribution of θ and $\hat{\theta}_{ML}$, the maximum likelihood estimator. For each value of θ in the set $\theta \in \{0.0, 0.05, 0.1, ..., 0.95, 1.0\}$, we generate the sampling distribution (histogram) of maximum likelihood estimator ($\hat{\theta}_{ML}$) based on 5,000 sets of generated data. When generating data, we set $\sigma^2 = 1$ and $T = 50$. [21] When these sampling distributions (histograms) are lined up side by side, they form a surface representing the joint frequency distribution of θ and $\hat{\theta}_{ML}$. [22]

Figure 2.4.A show four angles of the three dimensional joint frequency distribution. A slice of the resulting three-dimensional figure at a specific value of θ is the sampling distribution of $\hat{\theta}_{ML}$. A slice of the same figure at a specific value $\hat{\theta}_{ML} = \hat{\theta}$ is the posterior

---

[21] When estimating θ, we assume that the true value of σ^2 is known, following DeJong and Whiteman (1993). This does not affect the results.

[22] This approach is originally due to Sims and Uhlig (1991). They apply this procedure to investigate the differences between the posterior and the sampling distributions of the autoregressive parameter in an AR(1) model.

distribution of θ for given flat prior and a set of data that results in a maximum likelihood estimate of $\hat{\theta}$. The two resulting distributions are compared side by side in Figure 2.4.B. The results obtained by DeJong and Whiteman (1993), as replicated in this section, can be summarized by the following:

> **Finding #1:** The sampling distributions of $\hat{\theta}_{ML}$ piles up at unity with higher probability as θ approaches unity.
>
> **Finding #2:** The posterior distributions of θ do not pile up at unity.
>
> **Finding #3:** The posterior distributions always peak at $\theta = \hat{\theta}$, the maximum likelihood estimate, conditional on data.

As suggested by DeJong and Whiteman (1993), the implication of the above results is just the opposite of that in Sargan and Bhargava (1983). Sargan and Bhargava (1983) argue that "the occurrence of a maximum at $\theta = 1$ in the likelihood function is an insubstantial evidence for 'over-differencing' since the likelihood function can have a local maximum at $\theta = 1$ with reasonably high probabilities when the true value of θ is less than one." However, according to DeJong and Whiteman (1993) and the above replication of their results, when the maximum likelihood estimate of θ turns out to be one conditional on a particular data set, the most likely values for θ are those near one.

If Finding #3 from the above simulation study were to hold for general ARMA models, then the maximum likelihood estimate of $\theta_1 + \theta_2$ being one for Perron and Wada's (2009) model would be an apparent evidence of 'over-differencing.' However, the dramatic difference between the Bayesian and the classical inference of Perron and Wada's(2009) model reported in Section 2.2 is inconsistent with Finding #3. Judging from the perspective of Finding #3, this would be a puzzle. Otherwise, one might have to admit that, for general ARMA models, the posterior distributions for the sum of the moving-average parameters may not always have a peak at the maximum likelihood estimate.

2.4.2. The Profile Likelihood and the Posterior Distribution of θ: MA(1) Model with Intercept

In this section, we investigate whether or not the implication of DeJong and Whiteman's (1993) results replicated in the previous section continues to hold for more complicated models beyond an MA(1) model without intercept. For this purpose, consider the following MA(1) model given in equation (2.6) of section 2.3:

$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d.N(0, \sigma^2),$$
$$[\theta = 0.8, \sigma^2 = 1, \mu = 1, T = 50].$$

We generate many arbitrary data sets according to the above data generating process, in which the true value of θ is 0.8. Then, for each data set generated, we apply the Bayesian estimation procedure to obtain the posterior distribution of θ, by assuming σ^2 is known. [23] Contrary to Finding #3, we find that the posterior distribution of θ do not always have peaks at the maximum likelihood estimate. Instead, we have three categories of shapes for the posterior distributions, which are shown in Figure 2.5 and described below: [24]

> **Type #1:** $\hat{\theta}_{ML}$ is within the invertible region, and the peak of the posterior distribution is at around $\hat{\theta}_{ML}$;
> **Type #2:** $\hat{\theta}_{ML} = 1$ and the peak of the posterior distribution is at $\theta = 1$;
> **Type #3:** $\hat{\theta}_{ML} = 1$ but the peak of the posterior distribution is at the invertible region.

The first two types are consistent with Finding #3 of the simulation study performed for an MA(1) model without intercept, even though the Bayesian and classical econometricians may have different interpretations for the second type. From the classical point of view, given that the true value of θ is 0.8, Type #2 suggests that there are cases in which the posterior distribution piles up at unity, contrary to Finding #2 in Section 2.4.1. From the Bayesian point of view, however, one might argue that the properties of the particular data set generating this type of posterior distribution cannot be distinguished from those of data for which the true value of θ is one. But from the classical point of view, if one treats the

---

[23] When we do the same exercise for an MA(1) model without an intercept (by setting $\mu = 0$), the flat-prior posterior distributions of θ always have peaks at the maximum likelihood estimates, especially when σ^2 is assumed known.

[24] The relative occurrence of each of the three types is investigated in Section 2.4.3.

posterior mode as an estimator, the occurrence of the second type in repeated samples is referred to as the pile-up problem.

It is the third type that is not consistent with the Finding #3 of the simulation study in Section 2.4.1. For Type #3, the posterior distribution peaks at around the true value of 0.8, while the maximum likelihood estimate of θ is one. Its existence suggests that the Bayesian method may be less subject to the pile-up problem than the maximum likelihood method, in the repeated sampling context. It also suggests that the implication of DeJong and Whiteman (1993) applies only to simple MA(1) models without intercept. For more complicated models than this simple one, the most likely values of θ may not always be the ones near the maximum likelihood estimate (0.8, in this particular case), resulting in divergence between the inferences based on the Bayesian and the maximum likelihood methods. Furthermore, we think that the existence of the third type may explain the differences between the Bayesian and classical inferences of Perron and Wada's (2009) model reported in Section 2.2. The existence of the third type definitely warrants more investigation and begs an explanation.

As the model gets more complicated, we have more nuisance parameters. In this case, a major difference between the two methods lies in the way these nuisance parameters are handled. Following Smith and Naylor (1987) and Berger et al. (1999), we can explain the reason for a potential divergence between the Bayesian and the classical inferences by comparing the profile likelihood and the flat-prior posterior density (or integrated likelihood) defined below:

$$\underline{Profile\ Likelihood:}\ \ \hat{L}(\theta) = sup_{\mu}\ L(\theta, \mu), \tag{2.9}$$

$$\underline{Posterior\ Density\ (Integrated\ Likelihood):}\ \ L(\theta) = \int L(\theta, \mu) d\mu, \tag{2.10}$$

where $L(\theta, \mu)$ is the likelihood function.

The posterior density of θ is not dependent upon the nuisance parameters, as it is obtained by integrating the likelihood function with respect to the nuisance parameters. However, this is not always the case for the profile likelihood, as it is obtained by maximizing with respect to the nuisance parameters. Pierce (1971) proves that, in a regression model with ARMA(1,1) disturbances, the maximum likelihood estimator of θ and the regression coefficients (in our case, μ) are correlated in finite sample, even though they are asymptotically independent and jointly normal. Thus, the likelihood function may not be quadratic, making the shape of the profile likelihood different from that of the flat-prior posterior distribution in small sample. This suggests that the posterior mode at the peak of the posterior density may be different from the maximum likelihood estimate at the peak of the profile likelihood. [25]

In order to illustrate how the profile likelihood and the posterior density can be different in small sample (say, $T = 50$), we pick a representative sample from which each of the above three types of the posterior distributions is obtained. For each data set, we draw a three dimensional likelihood surface a 2-dimensional likelihood contour as a function of μ and θ, by fixing σ^2 at its true value. [26] The four angles of the three dimensional likelihood function are drawn in Figure 2.6.A.1, 2.6.B.1, or 2.6.C.1. The likelihood contour is depicted in Figure 2.6.A.2, 2.6.B.2, or 2.6.C.2, along with the corresponding profile likelihood and flat-prior posterior distribution for θ. [27]

For Type #1 and Type #2 in Figures 2.6.A.2 and 2.6.B.2, in which the posterior distributions peak at $\hat{\theta}_{ML}$, the shapes of the posterior density and the likelihood are very similar. Thus, the Bayesian and the classical inferences may not be very different from each other. For Type #3 in Figure 2.6.C.2, in which the posterior distribution does not peak at the maximum likelihood estimate, the shapes of the profile likelihood and the posterior density are very different from each other. Such a difference makes inferences based on the Bayesian and the maximum likelihood method different. This is the case in which the correlation between $\hat{\theta}_{ML}$ and $\hat{\mu}_{ML}$ is non-negligible in finite sample.

Note that, as shown in Type #3 of Figure 2.6.C.2, there may exist a local maximum in the invertible region and it may be close to the posterior mode. From Tables 2.1.A and

---

[25] But such finite sample discrepancy between the posterior mode and the maximum likelihood estimate disappears as the sample size increases, because the maximum likelihood estimators of μ and the nuisance parameters are asymptotically independent of each other.

[26] This does not affect the results, as the maximum likelihood estimator for σ^2 is independent of that for the rest of the parameters in the model.

[27] The three posterior distributions shown in Figure 5 are the same as those in Figures 6.A.2, 6.B.2, and 6.C.2 in order.

2.1.B, in which the classical and the Bayesian inferences are compared for Perron and Wada's (2009) ARMA model, the parameter values at the local maximum of the likelihood function are very close to those at the posterior modes of the parameters. Besides, the posterior distribution of $\theta_1 + \theta_2$ does not peak at the maximum likelihood estimate of one. It peaks at the invertible region. These suggest that the results reported in Tables 2.1.A and 2.1.B for Perron and Wada's (2009) model may be an example of Type #3.

2.4.3. Sampling Distributions for the Posterior Mode of θ: Monte Carlo Experiment

For Bayesians, there is only one realization of the data set, so contemplating the probability of the pile-up problem in repeated sampling may be conceptually irrelevant. However, in order for us to be able to evaluate and directly compare the probabilities of the pile-up problem for both the Bayesian and classical approaches, we have no choice but to treat the posterior mode of θ as a Bayesian estimator, which is treated as a random variable in repeated samples.

In this section, we evaluate the relative frequencies of Type #2 occurring in repeated samples, or the probabilities of the pile-up effect for the Bayesian approach, for various models and sample sizes. These are compared to the relative frequencies of Type #2 or Type #3 occurring, or the probabilities of the pile-up effect for the maximum likelihood approach, which are invested in Section 2.3. For this purpose, we perform a Monte Carlo experiment, by generating 5,000 sets of data according to each data generating process in equations (2.5)-(2.8) of Section 2.3. For each of the models and sample sizes ($T = 50$, 100, 200), we get the sampling distribution for the posterior mode of θ. [28] Throughout the MCMC iterations, we impose the constraint that $|\theta| < 1$. Thus, as in Section 2.3, we report $Pr[0.995 < \hat{\theta}_{mode} < 1]$ as the probability of the pile-up problem, where $\hat{\theta}_{mode}$ is the posterior mode of θ.

The results are depicted in Figure 2.7 and summarized in Table 2.3, along with the probabilities of the pile-up problem. As in the case of the maximum likelihood approach

---

[28] The model is estimated based on the MCMC algorithm proposed by Chib and Greenberg (1994). The priors we employ for $(\mu, \mu_1, \phi, \theta)$ are $N(0, 10^2)$, and the prior for σ^2 is diffuse.

reported in Table 2.2, the probabilities of the pile-up problem increase as the model gets more complicated; they decrease as the sample size increases. Without an exception, however, for all the models and sample sizes considered, the probabilities of the pile-up problem are 'considerably' smaller than in the case of the maximum likelihood approach. For example, when the sample size as big as 200, the Bayesian approach does not suffer from the pile-up problem, even for an ARMA(1,1) model with a structural break in intercept, the most complicated model under consideration. Note that, under the same situation, the probability of the pile-up problem remain as high as 0.24 for the maximum likelihood approach.

Based on the results in this section, along with those in Section 2.3, we can conclude that the Bayesian approach suffers considerably less from the pile-up problem than the maximum likelihood approach. This allows us to confer more credibility to the Bayesian inference of Perron and Wada's (2009) model reported in Table 2.1.B, Figures 2.1.A and 2.1.B.

2.5. Empirical Results: Let's Take Uncertain Breaks

2.5.1. The Perron and Wada's (2009) Sample [1947:1 1998:2]

We continue to employ Perron and Wada's (2009) sample, in order to examine the effect of relaxing the assumption of a known break date for the mean growth rate in Perron and Wada's (2009) model. We believe that incorporating uncertainty in the break date is equally as important as incorporating uncertainty in the rest of the parameters of the model. A reduction in the variance of the shocks to real GDP, namely the Great Moderation (Kim and Nelson (1999) and McConnell and Perez-Quiros (2000)), is also incorporated. The model we estimate using the Bayesian approach is given by:

$$
\begin{gathered}
\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*, \\
\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_2 e_{t-1} - \theta_2 e_{t-2}, \qquad (2.11) \\
e_t | D_t \sim i.i.dN(0, (1 - D_t)\sigma_0^2 + D_t \sigma_1^2), \\
Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \qquad Pr[S_t = 1 | S_{t-1} = 1] = 1,
\end{gathered}
$$

$$Pr[D_t = 0|D_{t-1} = 0] = q_{00}, \quad Pr[D_t = 1|D_{t-1} = 1] = 1,$$

where Δy_t is the real GDP growth rate, Δy_t^* is the demeaned growth rate. We incorporate uncertainty in the break dates for mean and variance by restricting the transition probabilities of S_t and D_t to allow for an absorbing state. [29] The estimates of the expected durations of regime 0 (i.e., $1/(1-p_{00})$ and $1/(1-q_{00})$) are the estimates of the break dates. For Bayesian estimation of the above model, we employ a multi-move Markov-Chain Monte Carlo (MCMC) algorithm recently proposed by Kim and Kim (2013).

Figure 2.8.A depicts the cumulative posterior probabilities of a structural break in the mean and the variance of real GDP growth. The nature of the structural break in the variance is such that the structural break is very sharp, leaving little uncertainty in the break date. However, it is interesting to observe that the structural break in the mean growth rate is not very sharp, leaving considerably high uncertainty in the break date.

Table 2.4.A reports the posterior moments of the model parameters, along with their 90 percent HPD (highest posterior density) interval. The posterior mean and mode of $\theta_1 + \theta_2$ are 0.095 and 0.443, respectively. The posterior distribution of $\theta_1 + \theta_2$ in Figure 2.8.B graphically illustrates these. It has little probability mass near unity. Table 4.B reports that the posterior moments of $\theta_1+\theta_2$ are robust with respect to various priors employed. The long-run impulse-response analysis in Figure 2.8.B shows that a shock to real output generates highly persistent fluctuations in real GDP. With uncertainty in the break dates incorporated, the posterior mode for the long-run impulse-response coefficient (1.515) reported in Table 2.4.A is even larger than that (1.391) for the Perron and Wada model reported in Table 2.1.B.

Figure 2.8.C depicts the posterior modes of the trend and the cyclical components of real GDP. The trend component explains most of the variations in real output and the resulting cyclical component is small in magnitude and noisy. That is, even after taking breaks with uncertain break dates, the implications of Nelson and Plosser (1982) and Morley et al. (2003) on trend-cycle decomposition continue to hold within the Bayesian framework, which is relatively free from the pile-up problem. All these results imply that it is highly unlikely that the log of real GDP may be a trend stationary process, contrary to the implication of

---

[29] This approach has been suggested by Chib (1998).

the classical inference for Perron and Wada's (2009) model.

2.5.2. Extended Sample with Stochastic Volatility [1947:1 2007:4]

In Sections 2.2 and 2.5.1, we focused our analysis to Perron and Wada's (2009) sample in order to directly compare our Bayesian results to their maximum likelihood estimation results. In this section, we employ an extended sample that covers observations up to 2007:4. For this extended sample, we incorporate a more flexible specification of the volatility process. Instead of assuming a one-time structural break, we employ a random walk stochastic volatility model in an attempt to better approximate the time-varying nature of the volatility process. [30] However, we continue to assume that there exists one structural break with an unknown break point in the long-run mean growth rate. The model is given below:

$$\begin{gathered}
\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*, \\
\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_2 e_{t-1} - \theta_2 e_{t-2}, \qquad (2.12) \\
e_t \sim N(0, \sigma_t^2), \\
ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \epsilon_t, \quad \epsilon_t \sim i.i.dN(0, \sigma_\epsilon^2), \\
Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \quad Pr[S_t = 1 | S_{t-1} = 1] = 1.
\end{gathered}$$

The empirical results are reported in Table 2.5 and in Figures 2.9.A, 2.9.B, and 2.9.C. It is worth mentioning that the case for the difference stationarity of real GDP is much more strengthened than in previous sections. The posterior mean and mode of $\theta_1 + \theta_2$ are -0.232 and -0.342, respectively, with the 90% HPD (Highest Posterior Density) interval covering the range of $[-0.786, 0.453]$. Figure 2.5.B reports that the posterior moments of $\theta_1 + \theta_2$ are robust with respect to various priors unless very strong priors are imposed on the sum of MA coefficients. An interesting observation is that the posterior distribution of the long-run impulse-response coefficient depicted in Figure 2.9.B is almost the same as that in Figure

---

[30] We would like to note that inferences reported in this section remain almost unchanged even when we incorporate a one-time structural break in the volatility process as in Section 2.5.1.

2.8.B. That is, 1% increase in the shock to real GDP increases real GDP by much more than 1% in the long-run, regardless of the sample and regardless of how the variance term is treated. [31]

2.6. Summary and Conclusion

A conventional belief is that a non-informative prior leads to the Bayesian posterior mode being very close to the maximum likelihood estimate, since the maximum likelihood estimate is not influenced by priors. If this is the case, the most likely values for the parameter of interest would be those near the maximum likelihood estimate, from the Bayesian perspective. We show that this common belief does not hold for general ARMA models, especially when there is a structural break in the mean. There are cases in which we may have the posterior mode of the moving-average parameter inside the invertible region, even when the maximum of the likelihood function occurs at unity. In the repeated sampling context, this suggests that the Bayesian approach is relatively freer from the pile-up problem than the maximum likelihood approach, which is confirmed by our simulation study.

Based on maximum likelihood estimation of an ARMA model of real GDP growth with a structural break in mean, Perron and Wada (2009) show that real GDP may be a trend stationary process. On the contrary, our results based on Bayesian estimation of the same model implies that most of the variations in real GDP can be explained by the stochastic trend component, consistent with the implications of Nelson and Plosser (1982) and Morley et al. (2003).

We argue that Perron and Wada's (2009) results may be due to the pile-up problem, to which the maximum likelihood method is subject in finite samples. Based on a Monte Carlo experiment, which is performed by taking the posterior modes of the parameters for Perron and Wada's (2009) model as true values, we show that the probability of the pile-up problem for the maximum likelihood approach is as high as 0.4. We conclude that, even

---

[31] Note that this implication remain unchanged even for the Bayesian inference in Table 1.B of Perron and Wada's (2009) model, in which the structural break point is assumed known.

after taking a break in the mean growth rate of real GDP in the mid-1970s as in Perron and Wada (2009), the implications of Nelson and Plosser (1982) and Morley et al. (2003) on trend-cycle decomposition continue to hold within the Bayesian framework, which is relatively free from the pile-up problem. This conclusion is unchanged even if we incorporate uncertainty in the dates of structural breaks in the mean growth rate and in the volatility process. When the sample is extended to include recent observations (1947:1-2007:4) up to just before the financial crisis, the evidence for the difference stationarity of real GDP is much more strengthened. These results are robust with respect to various priors employed.

Table 2.1.A. Maximum Likelihood Estimation of ARIMA Model for Real GDP: Perron and Wada's (2009) Model [Sample: 1947:1~1998:2]

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

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

$$S_t = 0\ for\ \ t \leq 1973{:}1,\ \ S_t = 1\ for\ t > 1973{:}1.$$

Parameters	Global Maximum		Local Maximum	
	Estimates	S.E.	Estimates	S.E.
μ_0	0.951	0.021	0.979	0.116
μ_1	-0.287	0.038	-0.332	0.166
$\phi_1 + \phi_2$	0.921	0.020	0.630	0.104
ϕ_2	-0.601	0.109	-0.731	0.150
$\theta_1 + \theta_2$	0.999	0.003	0.546	0.142
θ_2	-0.283	0.137	-0.542	0.211
σ^2	0.876	0.086	0.922	0.091
Long-run Impulse-Response	0.000	0.042	1.228	0.312
Log Likelihood	-278.930		-282.710	

Note: 1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009), namely the (log) quarterly US real GDP series seasonally adjusted for the period 1947:1–1998:2.
2. S.E. refers to the standard errors of the estimates.
3. S.E. of the long-run impulse response is reported using delta method.

Table 2.1.B. Bayesian Estimation of ARIMA Model for Real GDP: Perron and Wada's (2009) Model [Sample: 1947:1~1998:2]

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

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

$$S_t = 0 \ for \ \ t \leq 1973{:}1, \ \ S_t = 1 \ for \ t > 1973{:}1.$$

Parameters	Prior		Posterior				
	Mean	SD	Mean	Mode	Median	SD	90 % HPDI
μ_0	1.2	2	0.997	0.958	0.991	0.130	[0.780, 1.210]
μ_1	-0.5	2	-0.365	-0.326	-0.350	0.177	[-0.649, -0.040]
$\phi_1 + \phi_2$	0.5	2	0.489	0.596	0.540	0.269	[0.068, 0.940]
ϕ_2	0	2	-0.411	-0.534	-0.466	0.278	[-0.842, 0.077]
$\theta_1 + \theta_2$	0.5	2	0.267	0.483	0.345	0.422	[-0.397, 0.999]
θ_2	0	2	-0.330	-0.382	-0.343	0.192	[-0.661, 0.024]
σ^2	1	2	0.960	0.925	0.954	0.097	[0.800, 1.127]
Long-run Impulse-Response			1.365	1.391	1.387	0.315	[0.943, 1.891]

Note:
1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009), namely the (log) quarterly US real GDP series seasonally adjusted for the period 1947:1–1998:2.
2. Burn-in / Total iterations = 5,000 / 135,000
3. S.D. refers to the standard deviations of the posterior distributions.
4. A highest posterior density interval (HPDI) is an interval, the narrowest one possible with a chosen probability.
5. The Bayesian algorithm by Chib and Greenberg (1993) is used for estimation.
6. The acceptance probabilities of the Metropolis-Hastings algorithm in MCMC are all above 0.85.

Table 2.2. Sampling Distributions of Maximum Likelihood estimators for θ and the Probabilities of the Pile-up Problem: Monte Carlo Experiment

$$y_t = \mu + \mu_1 S_t + u_t, \quad u_t = \phi\, u_{t-1} + e_t - \theta e_{t-1}, \ e_t \sim i.i.d\ N\,(0, \sigma^2)$$

$$S_t = 0, \ for \ \ t \leq \frac{T}{2}; \quad S_t = 1, \ otherwise,$$

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

$$[\theta = 0.8; \ \sigma^2 = 1; \mu = \ 1; \ \mu_1 = -0.3, \phi = 0.3 \,]$$

	$\Pr\left[\hat{\theta}_{ML} \leq k \,\middle|\, \theta = 0.8\right]$					Prob. of Pile-up
			$\underline{k}$			
	0.6	0.7	**0.8**	0.9	1	
			MA(1) without Intercept			
$\boldsymbol{T = 50}$	0.025	0.124	**0.425**	0.755	1	**0.119**
$\boldsymbol{T = 100}$	0.001	0.051	**0.442**	0.906	1	**0.016**
$\boldsymbol{T = 200}$	0.000	0.011	**0.471**	0.979	1	**0.000**
			MA(1) with Intercept			
$\boldsymbol{T = 50}$	0.016	0.070	**0.238**	0.403	1	**0.582**
$\boldsymbol{T = 100}$	0.001	0.038	**0.329**	0.752	1	**0.169**
$\boldsymbol{T = 200}$	0.000	0.008	**0.374**	0.948	1	**0.008**
			MA(1) with a Structural Break in Intercept			
$\boldsymbol{T = 50}$	0.008	0.037	**0.104**	0.133	1	**0.867**
$\boldsymbol{T = 100}$	0.002	0.028	**0.215**	0.520	1	**0.447**
$\boldsymbol{T = 200}$	0.000	0.007	**0.301**	0.891	1	**0.044**
			ARMA(1,1) with a Structural Break in Intercept			
$\boldsymbol{T = 50}$	0.027	0.042	**0.058**	0.060	1	**0.939**
$\boldsymbol{T = 100}$	0.025	0.067	**0.165**	0.263	1	**0.730**
$\boldsymbol{T = 200}$	0.010	0.058	**0.281**	0.665	1	**0.236**

Note: 1. Probability of pile-up is approximated as Pr[0.995 ≤ $\hat{\hat{\theta}}$ ≤ 1] as in DeJong and Whiteman (1993).

Table 2.3. Sampling Distributions of Bayesian Posterior Modes for θ and the Probabilities of the Pile-up Problem: Monte Carlo Experiment

$$y_t = \mu + \mu_1 S_t + u_t, \quad u_t = \phi\, u_{t-1} + e_t - \theta e_{t-1}, \; e_t \sim i.i.d\; N\,(0, \sigma^2)$$

$$S_t = 0, \; for \;\; t \leq \frac{T}{2}; \quad S_t = 1, \; otherwise,$$

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

$$[\theta = 0.8; \; \sigma^2 = 1; \mu = \; 1; \; \mu_1 = -0.3, \phi = 0.3 \,]$$

	$\Pr\left[\hat{\theta}_{Mode} \leq k \;\middle|\; \theta = 0.8\right]$					Prob. of Pile-up
	k					
	0.6	0.7	**0.8**	0.9	1	
			MA(1) without Intercept			
$\boldsymbol{T = 50}$	0.035	0.143	**0.446**	0.777	1	**0.083**
$\boldsymbol{T = 100}$	0.002	0.062	**0.453**	0.902	1	**0.010**
$\boldsymbol{T = 200}$	0.000	0.013	**0.467**	0.975	1	**0.000**
			MA(1) with Intercept			
$\boldsymbol{T = 50}$	0.044	0.157	**0.407**	0.631	1	**0.110**
$\boldsymbol{T = 100}$	0.004	0.072	**0.455**	0.860	1	**0.013**
$\boldsymbol{T = 200}$	0.000	0.016	**0.462**	0.973	1	**0.003**
			MA(1) with a Structural Break in Intercept			
$\boldsymbol{T = 50}$	0.055	0.172	**0.365**	0.533	1	**0.197**
$\boldsymbol{T = 100}$	0.007	0.080	**0.435**	0.808	1	**0.037**
$\boldsymbol{T = 200}$	0.000	0.018	**0.456**	0.957	1	**0.003**
			ARMA(1,1) with a Structural Break in Intercept			
$\boldsymbol{T = 50}$	0.183	0.257	**0.341**	0.441	1	**0.241**
$\boldsymbol{T = 100}$	0.102	0.239	**0.422**	0.653	1	**0.137**
$\boldsymbol{T = 200}$	0.030	0.129	**0.458**	0.855	1	**0.006**

Note: 1. Probability of pile-up is approximated as Pr[0.995 ≤ $\hat{\hat{\theta}}$ ≤ 1] as in DeJong and Whiteman (1993).

Table 2.4.A. Bayesian Estimation of ARIMA Model for Real GDP: Unknown Break Points in Mean and Variance [Sample: 1947:1~1998:2]

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

$$e_t \,|D_t \sim i.i.d\; N\,(\,0, (1 - D_t)\,\sigma_0^2 + D_t\,\sigma_1^2),$$

$$\Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1 | S_{t-1} = 1] = 1,$$

$$\Pr[D_t = 0 | D_{t-1} = 0] = q_{00}, \quad \Pr[D_t = 1 | D_{t-1} = 1] = 1.$$

Parameters	Prior		Posterior				
	Mean	SD	Mean	Mode	Median	SD	90 % HPDI
p_{00}	0.99	0.01	0.987	0.994	0.989	0.009	[0.974, 0.999]
q_{00}	0.99	0.01	0.991	0.994	0.993	0.005	[0.984, 0.999]
μ_0	1.2	2	1.093	0.980	1.033	0.277	[0.697, 1.511]
μ_1	-0.5	2	-0.387	-0.258	-0.334	0.272	[-0.743, -0.000]
$\phi_1 + \phi_2$	0.5	2	0.433	0.593	0.460	0.257	[0.011, 0.832]
ϕ_2	0	2	-0.261	-0.385	-0.296	0.312	[-0.745, 0.245]
$\theta_1 + \theta_2$	0.5	2	0.095	0.443	0.094	0.405	[-0.555, 0.999]
θ_2	0	2	-0.285	-0.359	-0.300	0.195	[-0.594, 0.038]
σ_0^2	1	2	1.279	1.260	1.267	0.153	[1.031, 1.525]
σ_1^2	1	2	0.245	0.226	0.239	0.048	[0.170, 0.317]
Long-run Impulse-Response			1.587	1.515	1.571	0.292	[1.208, 2.044]

Note: 1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009), namely the (log) quarterly US real GDP series seasonally adjusted for the period 1947:1–1998:2.
2. Burn-in / Total iterations = 5,000 / 135,000
3. S.D. refers to the standard deviations of the posterior distributions.
4. A highest posterior density interval (HPDI) is an interval, the narrowest one possible with a chosen probability.
5. The Bayesian algorithms by Kim and Kim (2013), Chib and Greenberg (1993) are used for estimation.
6. The acceptance probabilities of the Metropolis-Hastings algorithms in MCMC are all above 0.5.

Table 2.4.B. Sensitivity of the Posterior Moments of $\boldsymbol{\theta_1 + \theta_2}$ to its Prior Moments: Unknown Break Points in Mean and Variance [Sample: 1947:1~1998:2]

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

$$e_t \,|D_t \sim i.i.d\; N\,(\,0, (1 - D_t)\,\sigma_0^2 + D_t\,\sigma_1^2),$$

$$\Pr[S_t = 0|S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1|S_{t-1} = 1] = 1,$$

$$\Pr[D_t = 0|D_{t-1} = 0] = q_{00}, \quad \Pr[D_t = 1|D_{t-1} = 1] = 1.$$

$$Prior\ Distirbution: \theta_1 + \theta_2 \sim N(\mu_{\theta_1+\theta_2}, \Omega_{\theta_1+\theta_2}^2\,)$$

Prior		Posterior				
Mean ($\mu_{\theta_1+\theta_2}$)	SD ($\Omega_{\theta_1+\theta_2}$)	Mean	Mode	Median	SD	90 % HPDI
0.5	2	0.095	0.443	0.094	0.405	[-0.555, 0.999]
1	2	0.135	0.404	0.147	0.405	[-0.515, 0.999]
0.5	1	0.132	0.424	0.133	0.384	[-0.506, 0.747]
1	1	0.210	0.423	0.242	0.380	[-0.403, 0.999]
0.5	0.5	0.261	0.423	0.295	0.327	[-0.309, 0.779]
1	0.5	0.447	0.475	0.467	0.298	[0.012, 0.999]

Note:
1. Burn-in / Total iterations = 5,000 / 135,000
2. S.D. refers to the standard deviations of the posterior distributions.
3. The priors for all the other parameters are same as in Table 4.A.
4. A highest posterior density interval (HPDI) is an interval, the narrowest one possible with a chosen probability.

Table 2.5.A. Bayesian Estimation of ARIMA Model for Real GDP: Stochastic Volatility and Unknown Break Point in Mean [Sample: 1947:1~2007:4]

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

$$e_t \sim N(0, \sigma_t^2),$$

$$\Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1 | S_{t-1} = 1] = 1,$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t, \ \ \varepsilon_t \sim i.i.d. N(0, \sigma_\varepsilon^2).$$

Parameters	Prior		Posterior				
	Mean	SD	Mean	Mode	Median	SD	90 % HPDI
p_{00}	0.99	0.01	0.988	0.994	0.990	0.009	[0.975, 0.999]
μ_0	1.2	2	1.076	0.967	1.026	0.225	[0.771, 1.379]
μ_1	-0.5	2	-0.292	-0.213	-0.250	0.218	[-0.546, -0.000]
$\phi_1 + \phi_2$	0.5	2	0.224	0.193	0.224	0.258	[-0.171, 0.649]
ϕ_2	0	2	-0.059	-0.018	-0.047	0.293	[-0.541, 0.410]
$\theta_1 + \theta_2$	0.5	2	-0.232	-0.342	-0.278	0.371	[-0.786, 0.453]
θ_2	0	2	-0.230	-0.230	-0.233	0.191	[-0.537, 0.081]
σ_ε^2	0.1	0.1	0.059	0.050	0.055	0.022	[0.027, 0.091]
Long-run Impulse-Response			1.607	1.564	1.584	0.227	[1.273, 1.946]

Note: 1. The raw data are (log) seasonally adjusted quarterly U.S. real GDP for the sample period of 1947:1 to 2007:4 and were taken from the St. Louis Fed (FRED) database.
2. Burn-in / Total iterations = 5,000 / 135,000
3. S.D. refers to the standard deviations of the posterior distributions.
4. A highest posterior density interval (HPDI) is an interval, the narrowest one possible with a chosen probability.
5. The Bayesian algorithms by Kim and Kim (2013), Chib and Greenberg (1993), and Kim et al. (1998) are used for estimation.
6. The initial value of σ_t^2 is treated as a parameter to be estimated.
7. The acceptance probabilities of the Metropolis-Hastings algorithms in MCMC are all above 0.85.

Table 2.5.B. Sensitivity of the Posterior Moments of $\boldsymbol{\theta_1 + \theta_2}$ to its Prior Moments: Stochastic Volatility and Unknown Break Point in Mean [Sample: 1947:1~2007:4]

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

$$e_t \sim N(0, \sigma_t^2),$$

$$\Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1 | S_{t-1} = 1] = 1,$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t, \;\; \varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2).$$

$$Prior\ Distirbution: \theta_1 + \theta_2 \sim N(\mu_{\theta_1 + \theta_2}, \Omega^2_{\theta_1 + \theta_2})$$

Prior		Posterior				
Mean ($\mu_{\theta_1+\theta_2}$)	SD ($\Omega_{\theta_1+\theta_2}$)	Mean	Mode	Median	SD	90 % HPDI
0.5	2	-0.232	-0.342	-0.278	0.371	[-0.786, 0.453]
1	2	-0.200	-0.368	-0.256	0.392	[-0.772, 0.499]
0.5	1	-0.160	-0.322	-0.213	0.380	[-0.729, 0.499]
1	1	-0.083	-0.271	-0.138	0.401	[-0.681, 0.581]
0.5	0.5	0.042	-0.125	0.021	0.351	[-0.508, 0.594]
1	0.5	0.295	0.421	0.331	0.348	[-0.291, 0.999]

Note:
1. Burn-in / Total iterations = 5,000 / 135,000
2. S.D. refers to the standard deviations of the posterior distributions.
3. The priors for all the other parameters are same as in Table 5.A.
4. A highest posterior density interval (HPDI) is an interval, the narrowest one possible with a chosen probability.

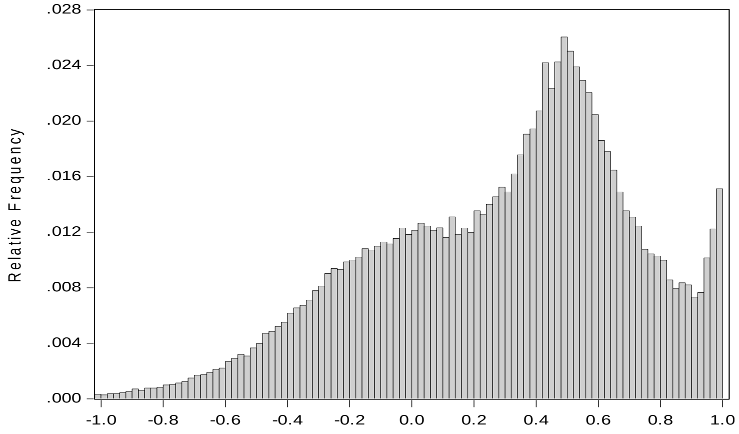
Figure 2.1.A. Posterior Distributions for the Sum of MA Parameters and Long-Run Impulse-Response Coefficient: Perron and Wada's (2009) Model [Sample: 1947:1~1998:2]

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

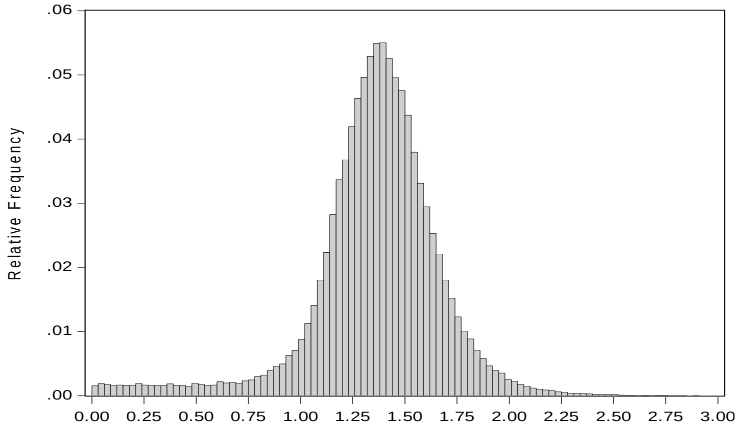
$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

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

$$S_t = 0 \ \ for \ \ t \leq 1973{:}1, \ \ S_t = 1 \ \ for \ \ t > 1973{:}1.$$



$\theta_1 + \theta_2$



Long-run Impulse Response Coefficient

Note:

1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009), namely the (log) quarterly US real GDP series seasonally adjusted for the period 1947:1–1998:2.
2. The model is estimated by the MCMC algorithm by Chib and Greenberg (1993). The total number of Bayesian MCMC iterations is 135,000 and the first 5,000 samples are discarded.

Figure 2.1.B. Comparison of Classical and Bayesian Inferences for Trend-Cycle Decomposition and Impulse-Response Analysis: Perron and Wada's (2009) Model [Sample 1947:1~1998:2]

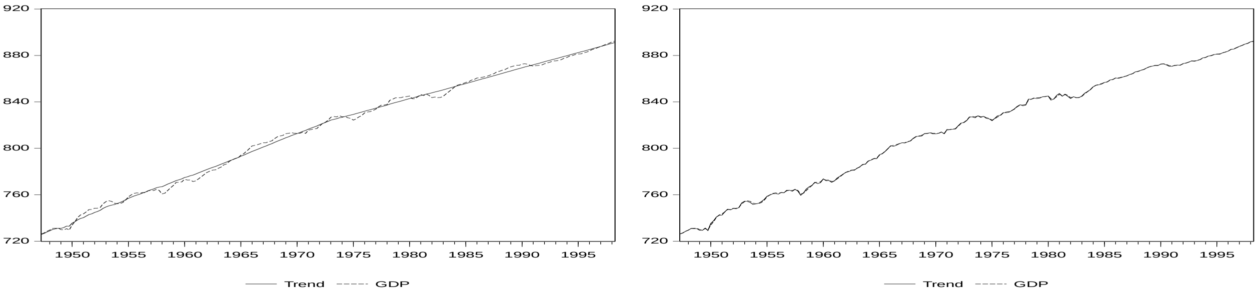
$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

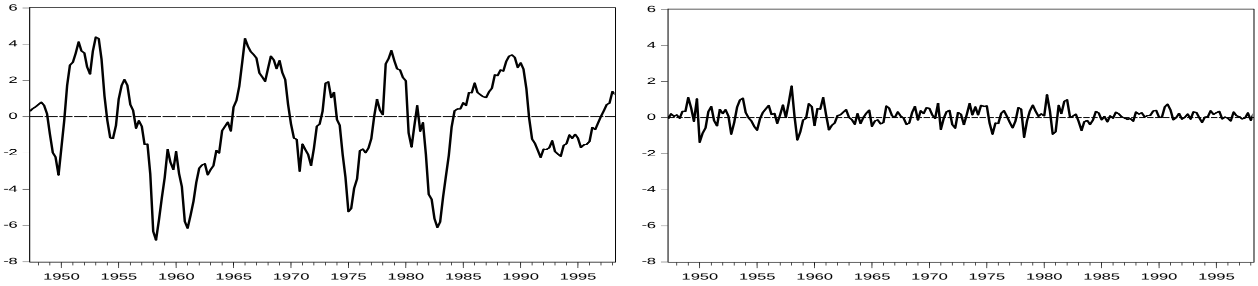
$$e_t \sim i.i.d\ N(0, \sigma^2),$$

$$S_t = 0 \ for \ \ t \leq 1973{:}1, \ \ S_t = 1 \ for \ t > 1973{:}1.$$

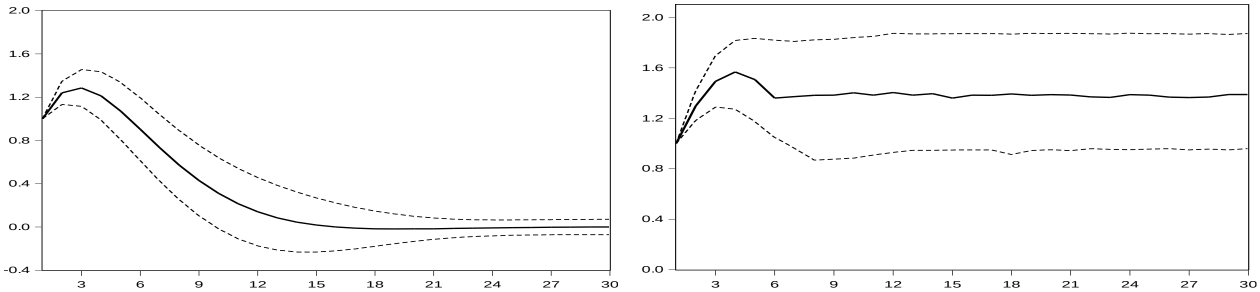
Classical Inference Bayesian Inference



Log of Real GDP vs. Trend Component



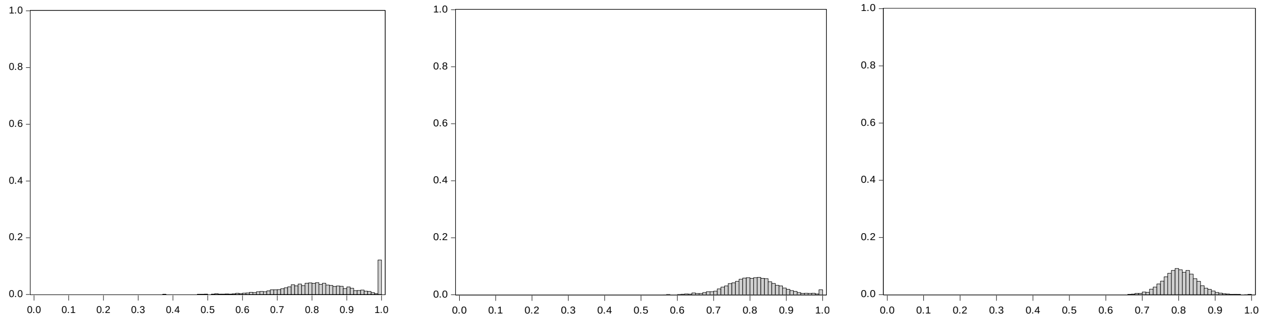
Cyclical Component



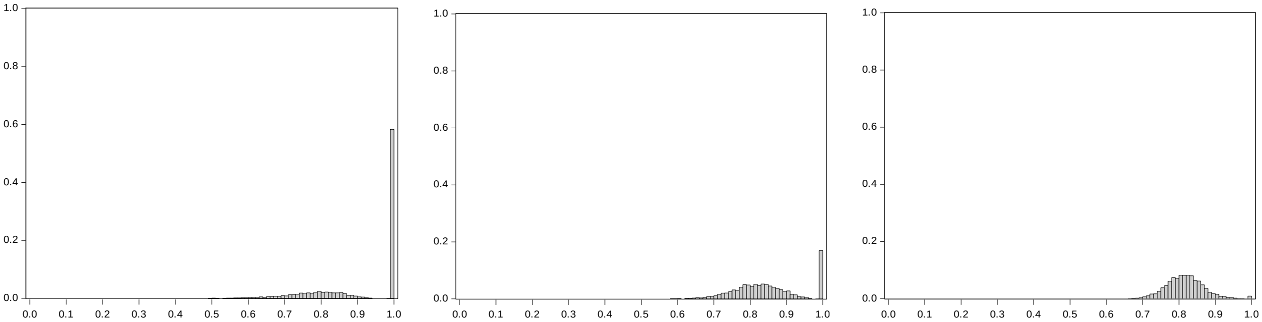
Impulse-Response Analysis

Note:
1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009), namely the (log) quarterly US real GDP series seasonally adjusted for the period 1947:1–1998:2.
2. The model is estimated by the MCMC algorithm by Chib and Greenberg (1993). The total number of Bayesian MCMC iterations is 135,000 and the first 5,000 samples are discarded.
3. The confidence band for the impulse-response function analysis is calculated by the Delta method.

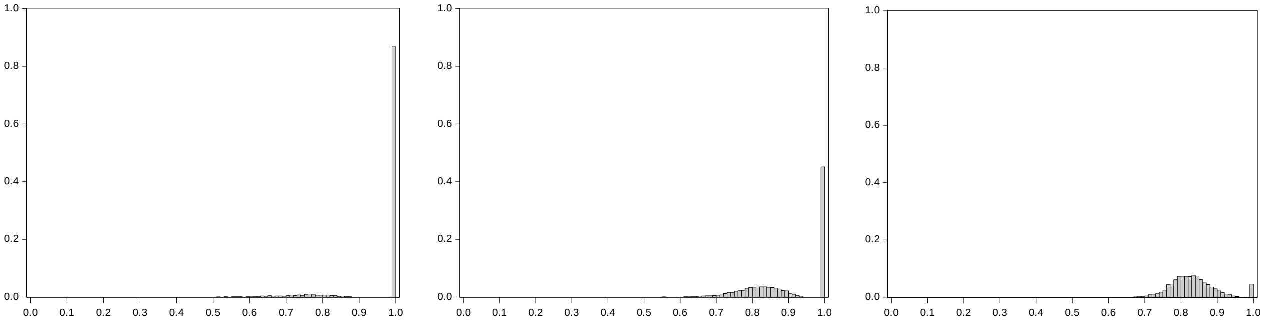
Figure 2.2. Sampling Distributions of Maximum Likelihood estimators for θ: Monte Carlo Experiment



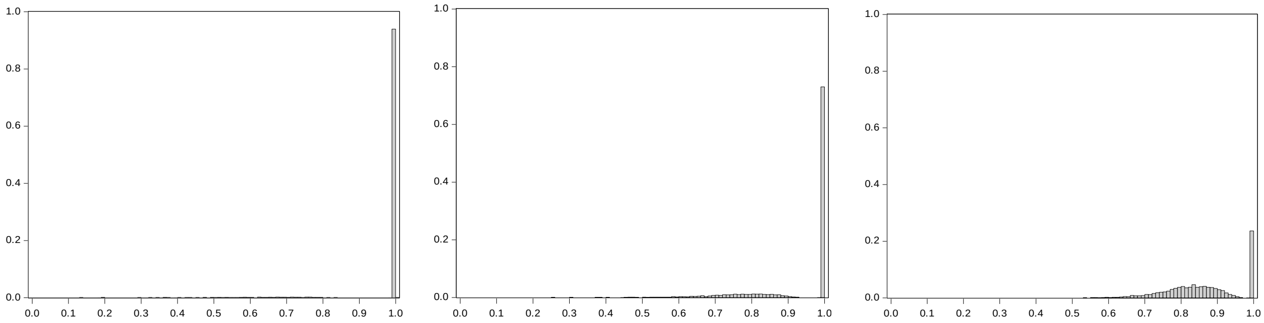
Model #1 (T=50) Model #1 (T=100) Model #1 (T=200)



Model #2 (T=50) Model #2 (T=100) Model #2 (T=200)



Model #3 (T=50) Model #3 (T=100) Model #3 (T=200)



Model #4 (T=50) Model #4 (T=100) Model #4 (T=200)

Note:
1. The number of simulations is 5,000.
2. The vertical axis represents relative frequency.

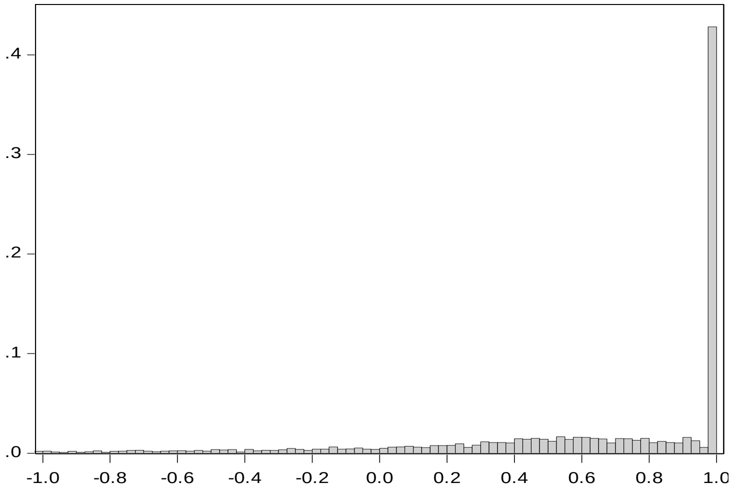
Figure 2.3. Sampling Distribution of the Sum of MA Parameters from Monte Carlo Experiment: Perron and Wada's (2009) Model

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

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

$$S_t = 0\ \ for\ \ t \leq 1973{:}1,\ \ S_t = 1\ \ for\ \ t > 1973{:}1.$$

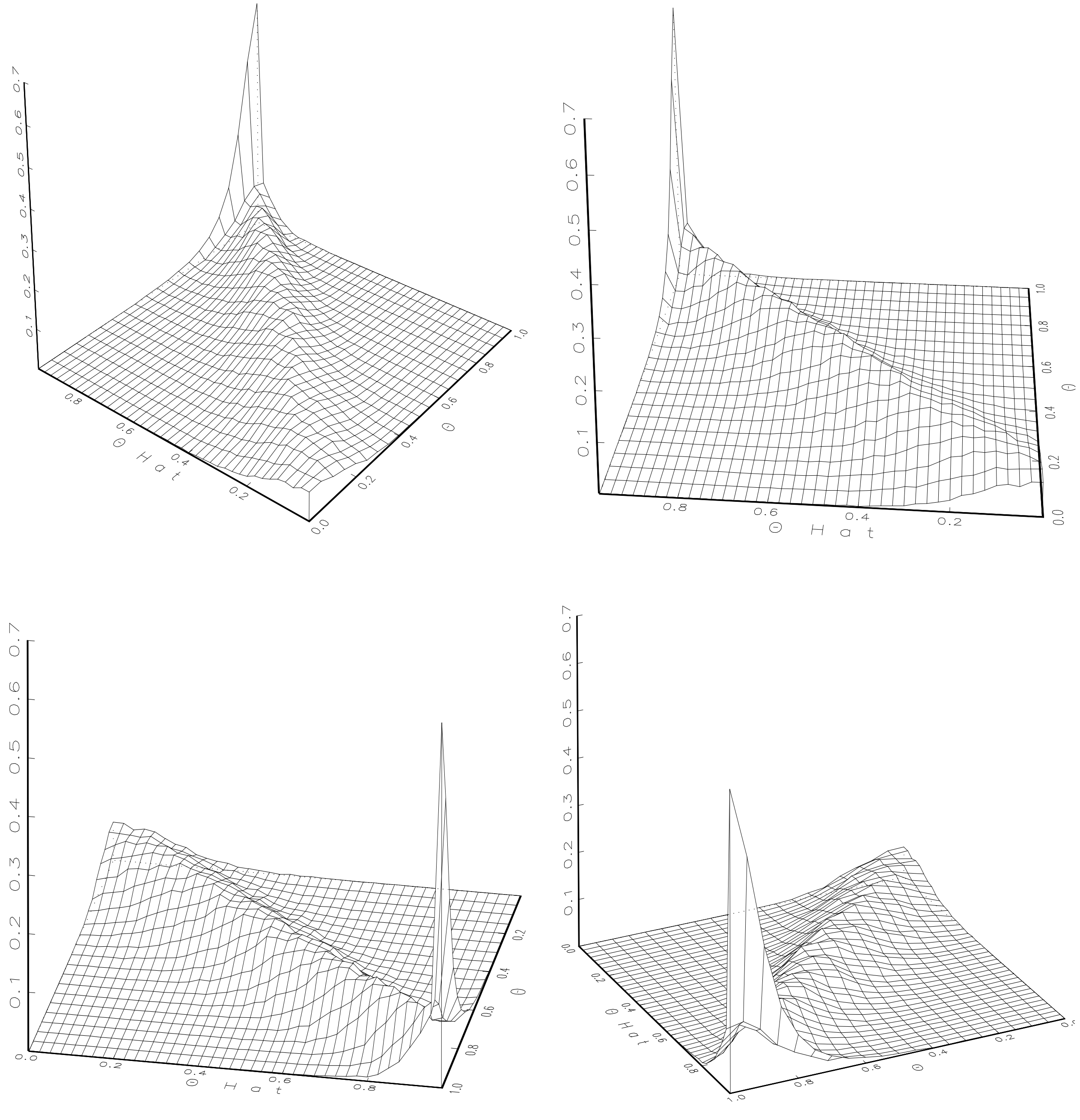


$\theta_1 + \theta_2$

Note:
1. The posterior modes in Table 1.B are used as true parameter values when generating data.
2. The total number of simulations is 5,000.
3. The vertical axis represents relative frequency.

Figure 2.4.A. Joint Frequency Distribution of θ and $\widehat{\theta}_{ML}$ for an MA(1) Model without Intercept: DeJong and Whiteman (1993)

$$y_t = e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\ N\ (0,1), \quad t = 1,2,\ldots,50$$

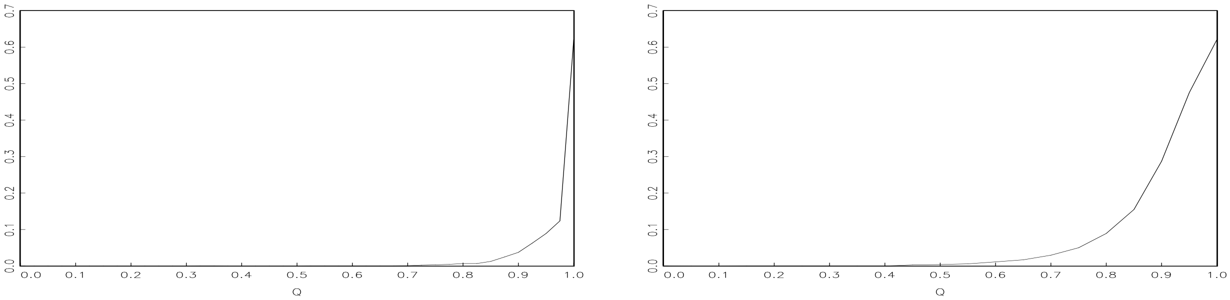


Note: 1. The vertical axis represents relative frequency.

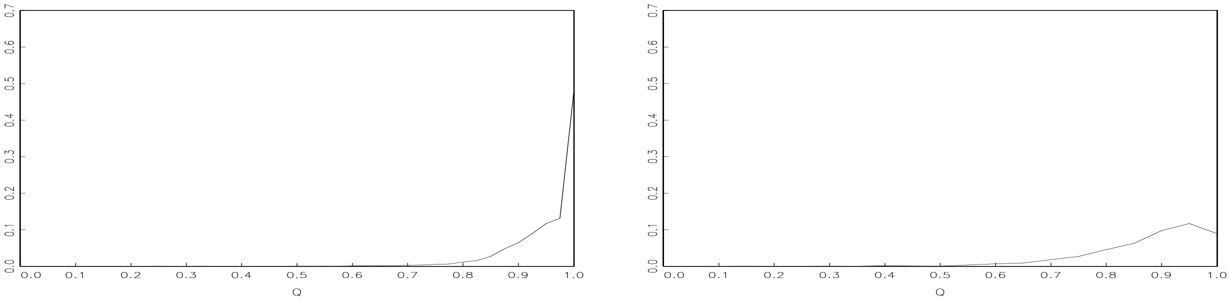
Figure 2.4.B. Sampling Distributions of ML estimator and Bayesian Posterior Distributions

$$y_t = \ e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\ N\ (0,1), \quad t = 1,2,\ldots,50$$

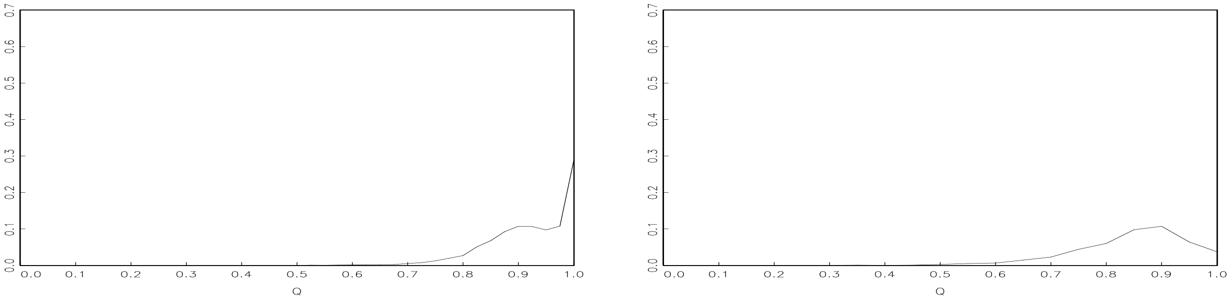
Sampling Distributions for $\hat{\theta}_{ML}$ Posterior Distributions for θ



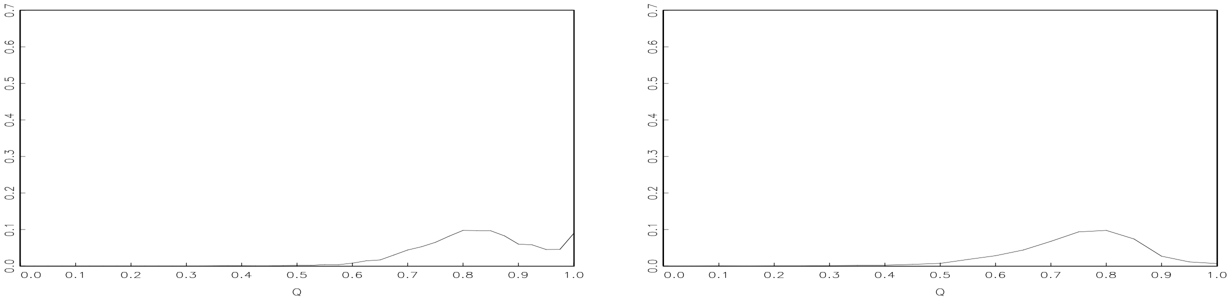
$\hat{\theta}_{ML} | \ \theta = 1$ $\theta \mid \hat{\theta}_{ML} = 1$



$\hat{\theta}_{ML} | \ \theta = 0.95$ $\theta \mid \hat{\theta}_{ML} = 0.95$



$\hat{\theta}_{ML} | \ \theta = 0.9$ $\theta \mid \hat{\theta}_{ML} = 0.9$



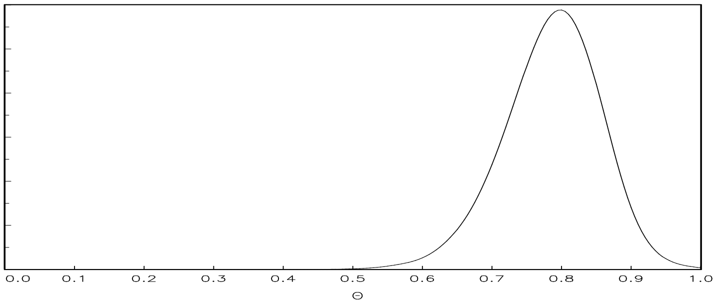
$\hat{\theta}_{ML} | \ \theta = 0.8$ $\theta \mid \hat{\theta}_{ML} = 0.8$

Note: 1. The vertical axis represents relative frequency.

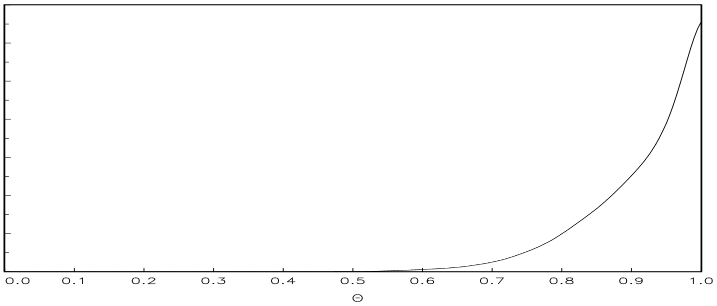
Figure 2.5. Typical Posterior Distributions of θ for Particular Sets of Data Generated

$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\; N\,(0, \sigma^2), \quad t = 1, 2, \ldots, T,$$

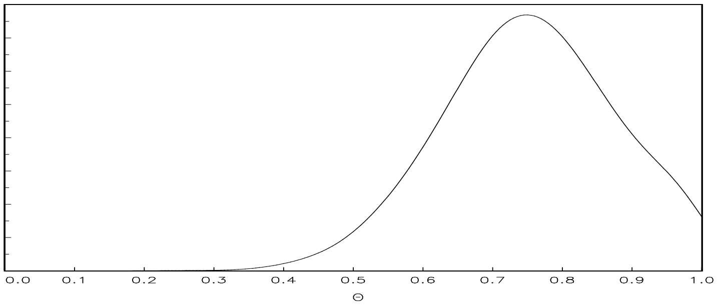
$$[\,\mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50].$$



Type #1 [$\hat{\theta}_{ML} = 0.816$]



Type #2 [$\hat{\theta}_{ML} = 1$]



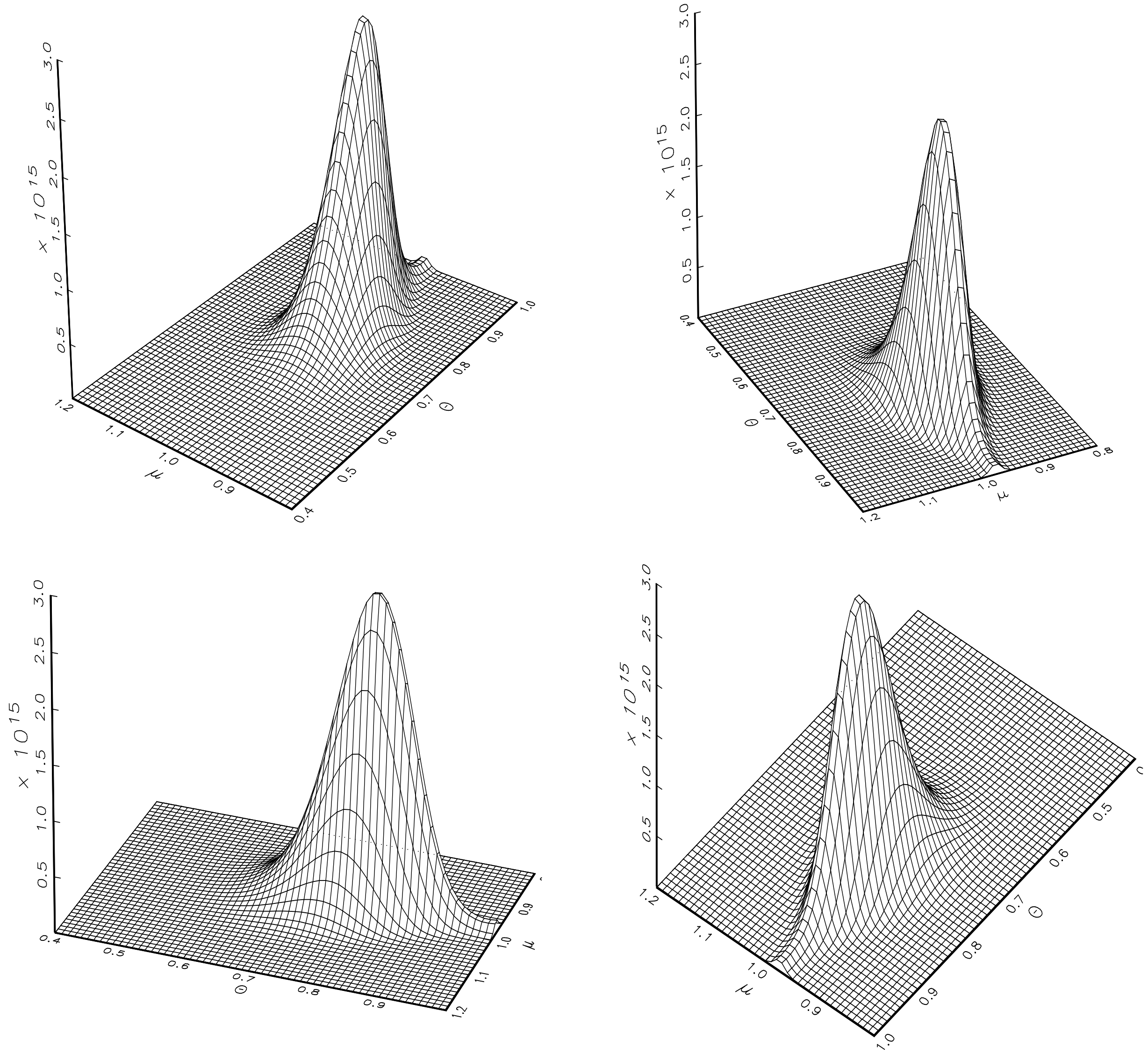
Type #3 [$\hat{\theta}_{ML} = 1$]

Note: 1. The model is estimated by the MCMC algorithm by Chib and Greenberg (1993). The total number of Bayesian MCMC iterations is 50,000 and the first 1,000 samples are discarded. σ^2 is assumed to be known.

Figure 2.6.A.1. Likelihood Surface of a Representative Sample for Type #1: Four Angles

$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\; N\,(0, \sigma^2), \quad t = 1,2,\ldots,T,$$

$$[\,\mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50].$$



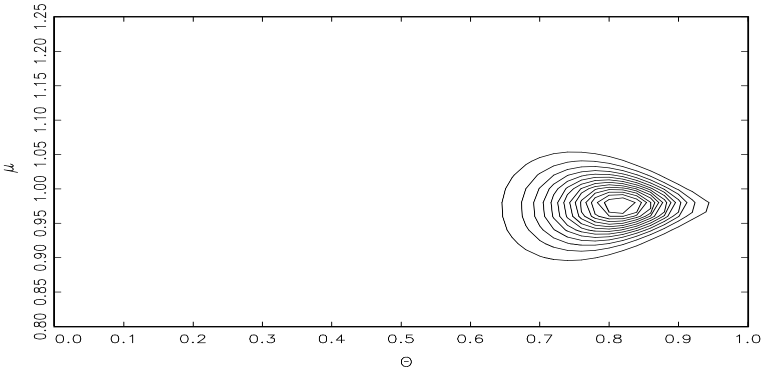
Note: 1. The vertical axis represents log likelihood values.

Figure 2.6.A.2. Comparison of Profile likelihood and Posterior Distribution for MA coefficient: Representative Sample for Type #1

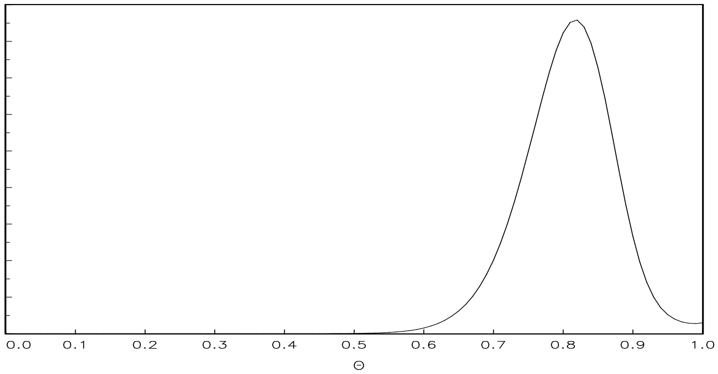
$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\ N\ (0, \sigma^2), \quad t = 1,2,\dots,T,$$

$$[\,\mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50].$$

Likelihood Contour



Profile Likelihood



Posterior Distribution

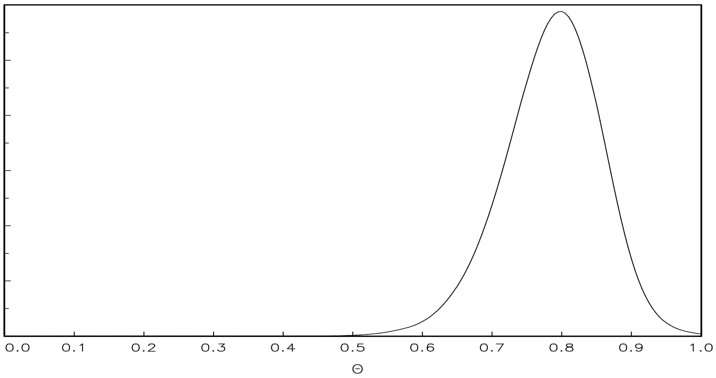
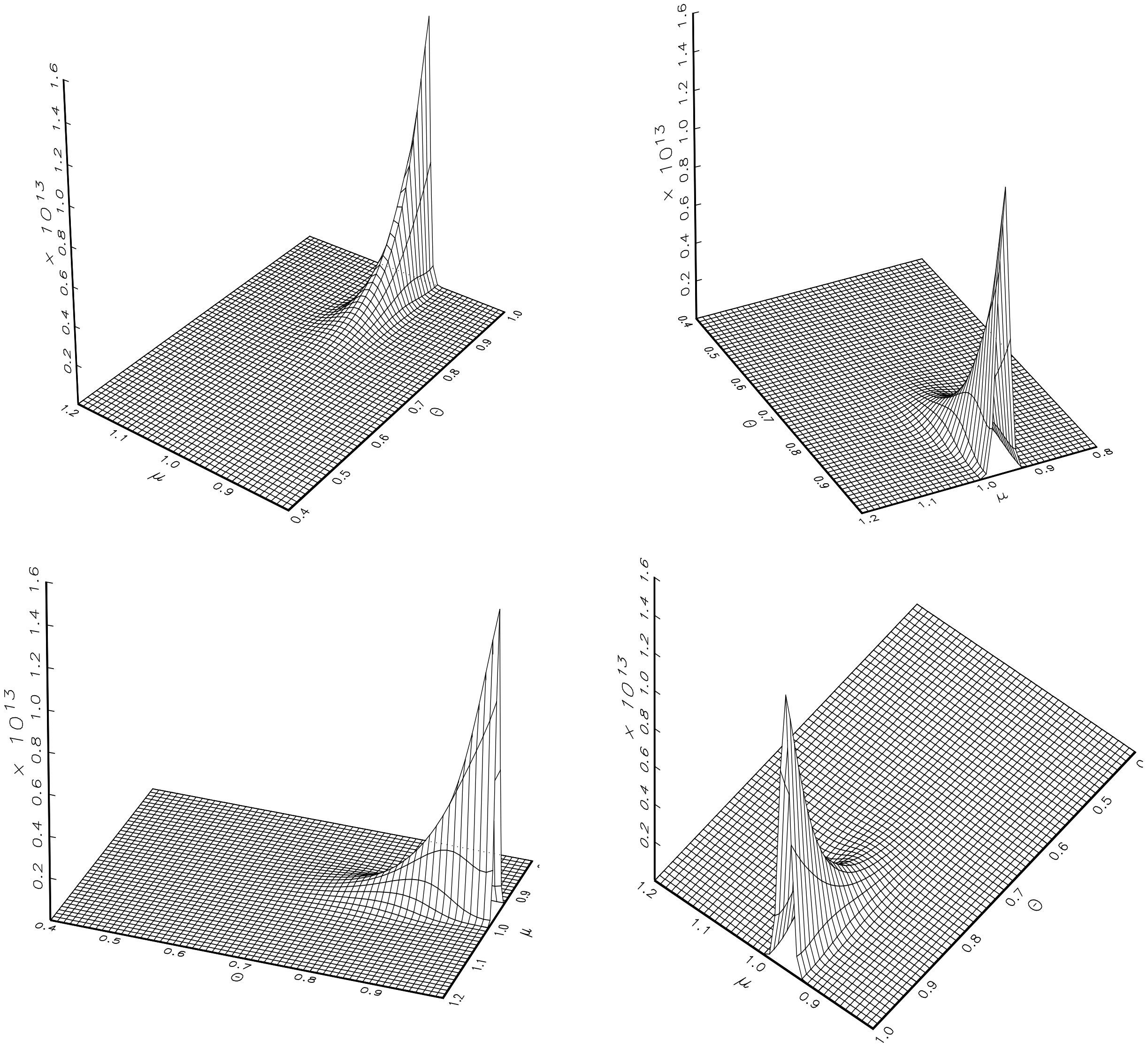


Figure 2.6.B.1. Likelihood Surface of a Representative Sample for Type #2: Four Angles

$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\; N\,(0, \sigma^2), \quad t = 1, 2, \dots, T,$$

$$[\,\mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50].$$



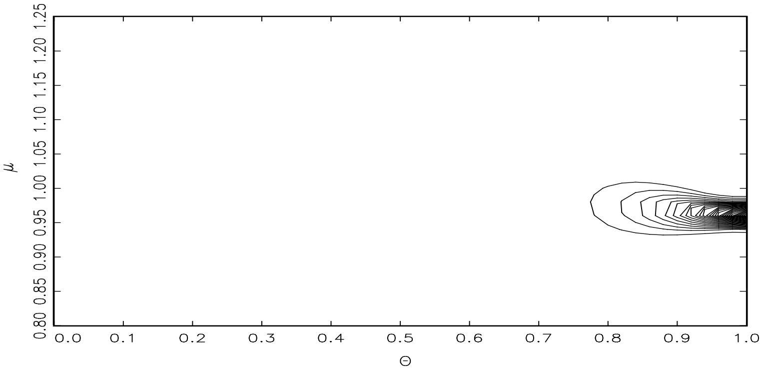
Note: 1. The vertical axis represents log likelihood values.

Figure 2.6.B.2. Comparison of Profile likelihood and Posterior Distribution for MA coefficient: Representative Sample for Type #2

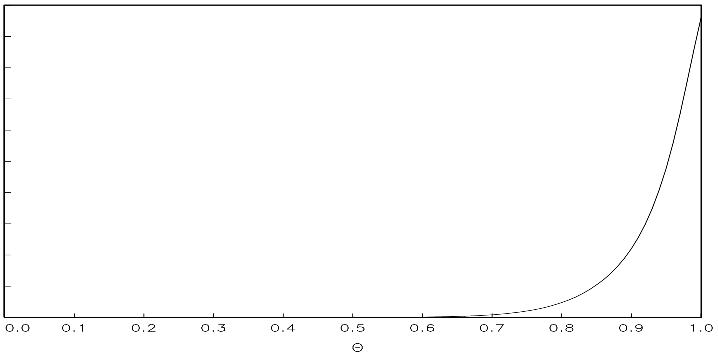
$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\ N(0, \sigma^2), \quad t = 1,2,\dots,T,$$

$$[\mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50].$$

Likelihood
Contour



Profile
Likelihood



Posterior
Distribution

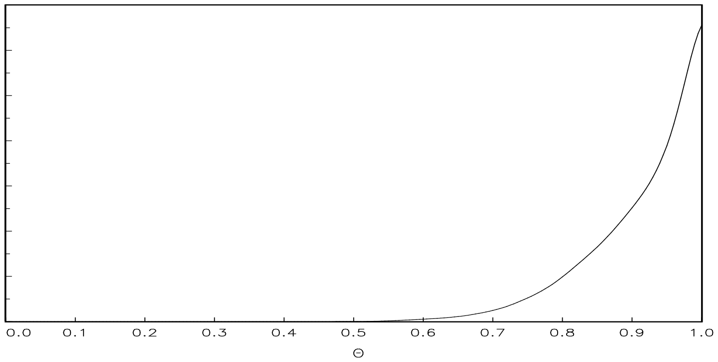
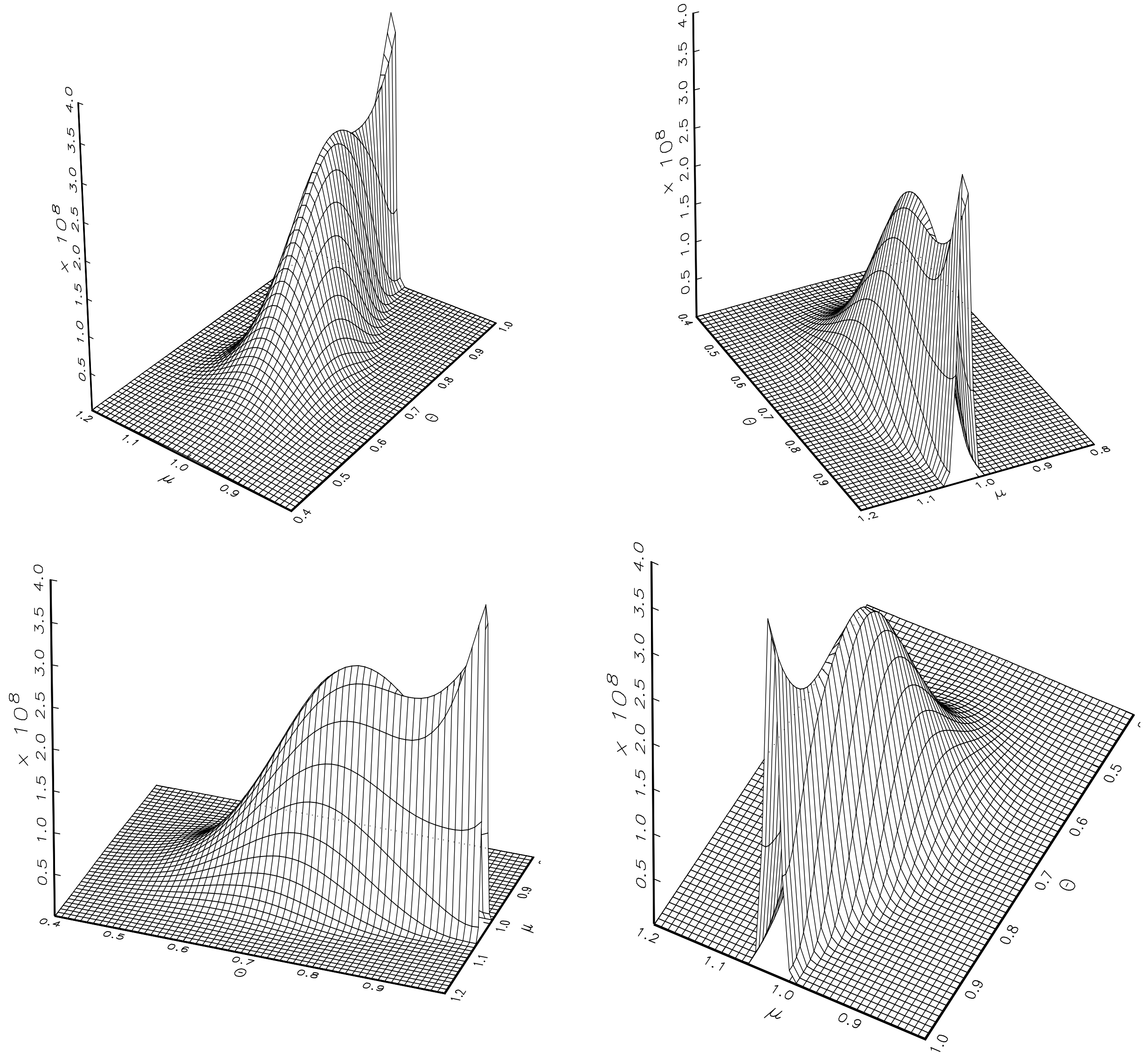


Figure 2.6.C.1. Likelihood Surface of a Representative Sample for Type #3: Four Angles

$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\ N\ (0, \sigma^2), \quad t = 1,2,\dots,T,$$

$$[\mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50].$$



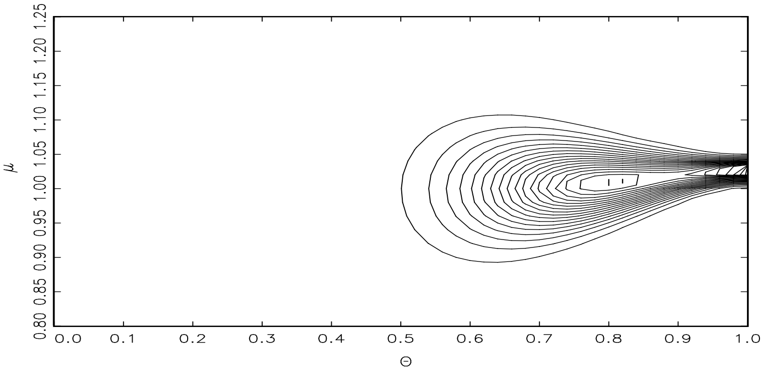
Note: 1. The vertical axis represents log likelihood values.

Figure 2.6.C.2. Comparison of Profile likelihood and Posterior Distribution for MA coefficient: Representative Sample for Type #3

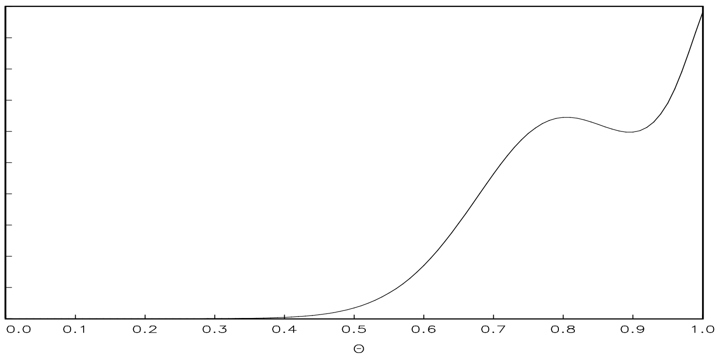
$$y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d\ N\,(0, \sigma^2), \quad t = 1,2,\dots,T,$$

$$[\,\mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50].$$

Likelihood
Contour



Profile
Likelihood



Posterior
Distribution

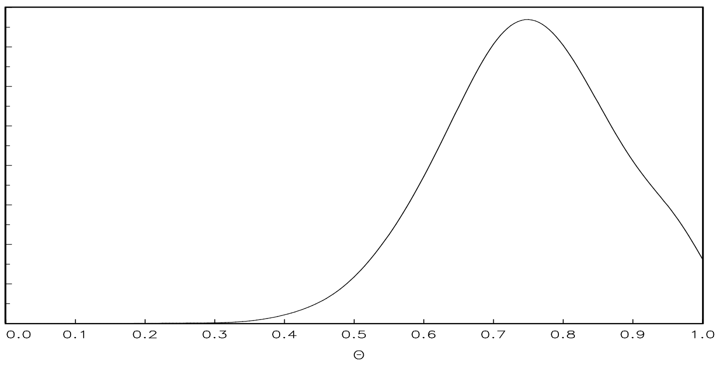
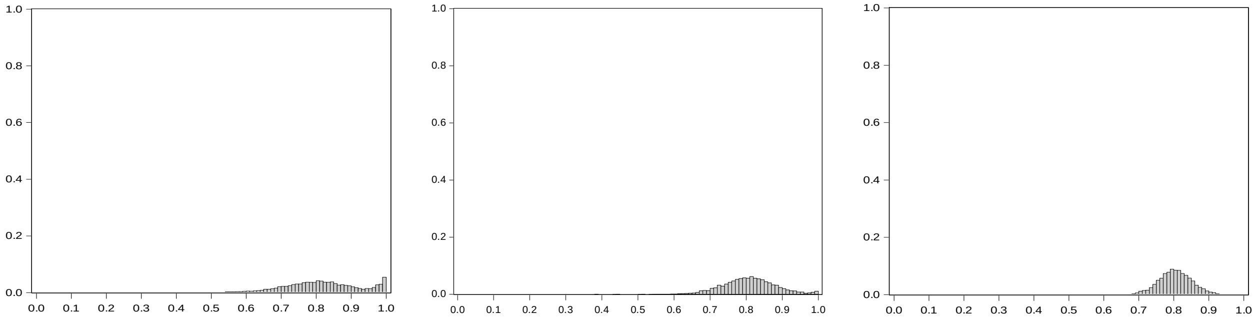
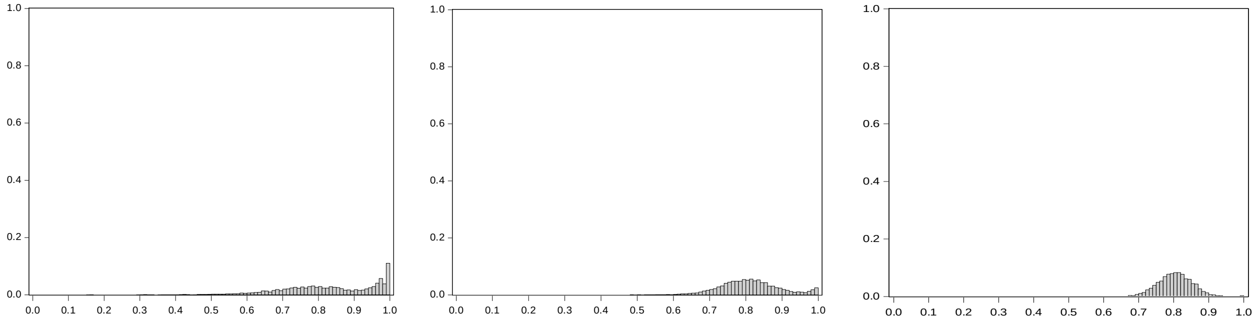


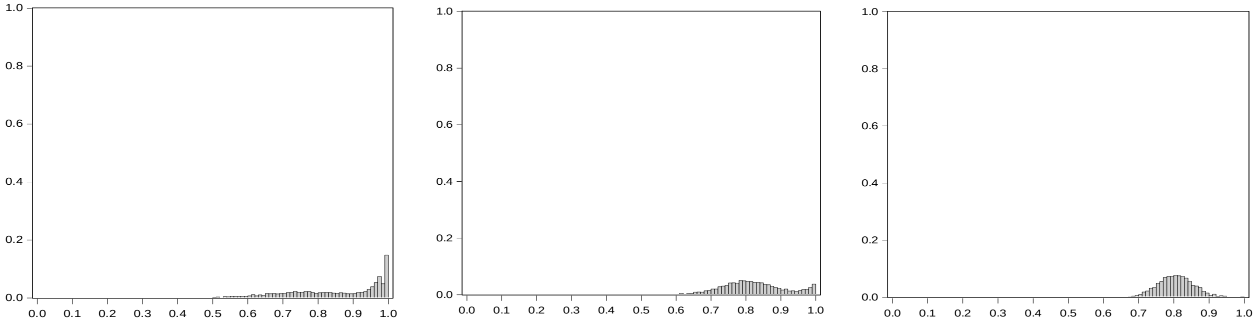
Figure 2.7. Sampling Distributions of Bayesian Posterior Modes for θ: Monte Carlo Experiment



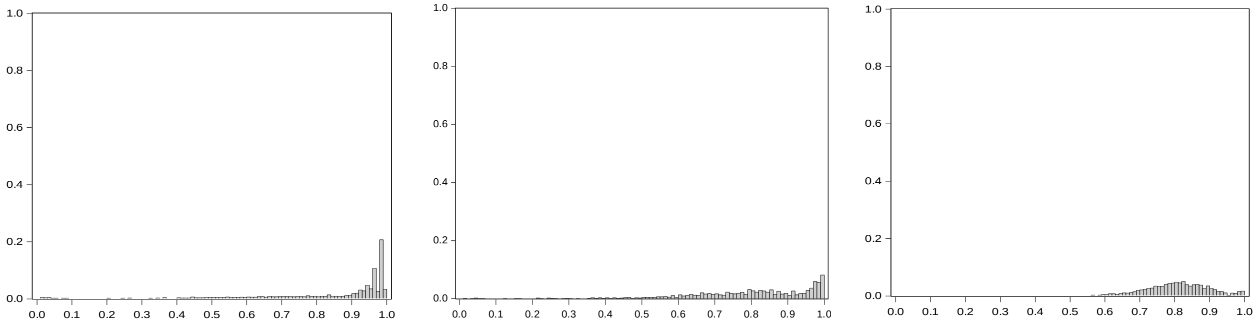
Model #1 (T=50) Model #1 (T=100) Model #1 (T=200)



Model #2 (T=50) Model #2 (T=100) Model #2 (T=200)



Model #3 (T=50) Model #3 (T=100) Model #3 (T=200)



Model #4 (T=50) Model #4 (T=100) Model #4 (T=200)

Note:
1. The total number of simulations is 5,000.
2. The total number of Bayesian MCMC iterations for each simulation is 6,000 and the first 1,000 samples are discarded.

Figure 2.8.A. Bayesian Inference of Structural Breaks in Mean and Variance: ARIMA Model with Unknown Break Points in Mean and Variance [Sample: 1947:1~1998:2]

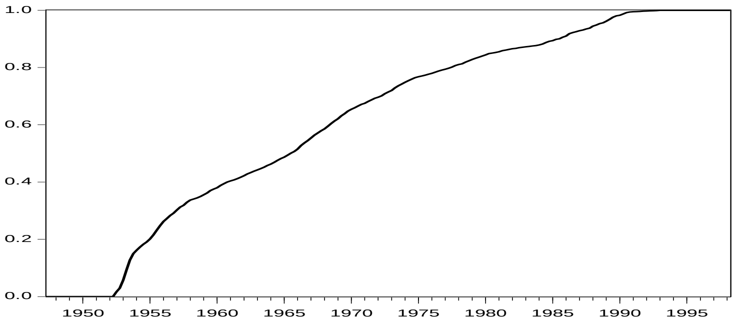
$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

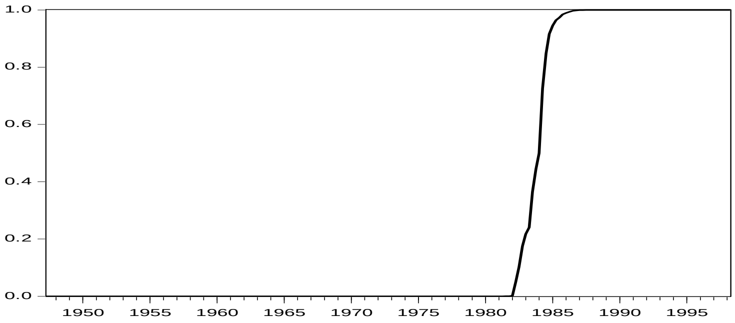
$$e_t \,|D_t \sim i.i.d\; N\,(0, (1 - D_t)\,\sigma_0^2 + D_t\,\sigma_1^2),$$

$$\Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1 | S_{t-1} = 1] = 1,$$

$$\Pr[D_t = 0 | D_{t-1} = 0] = q_{00}, \quad \Pr[D_t = 1 | D_{t-1} = 1] = 1.$$



Cumulative Posterior Probability of Structural Break in Mean



Cumulative Posterior Probability of Structural Break in Variance

Note:

1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009), namely the (log) quarterly US real GDP series seasonally adjusted for the period 1947:1–1998:2.
2. The model is estimated by the MCMC algorithm by Kim and Kim (2013), Chib and Greenberg (1993). The total number of Bayesian MCMC iterations is 135,000 and the first 5,000 samples are discarded.

Figure 2.8.B. Posterior Distributions for the Sum of MA Parameters and Long-Run Impulse-Response Coefficient: ARIMA Model with Unknown Break Points in Mean and Variance [Sample: 1947:1~1998:2]

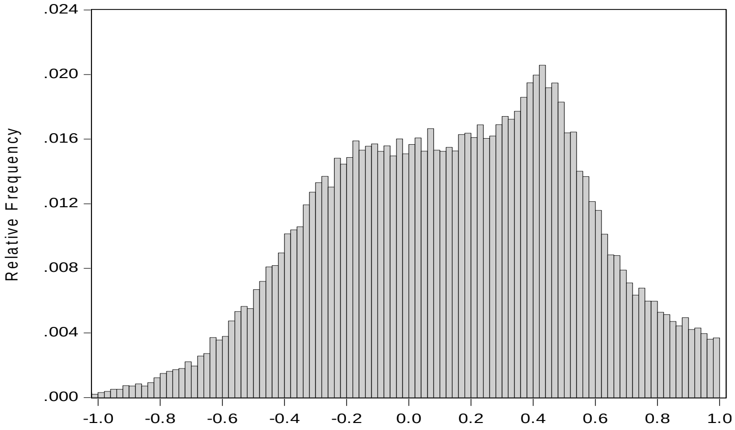
$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

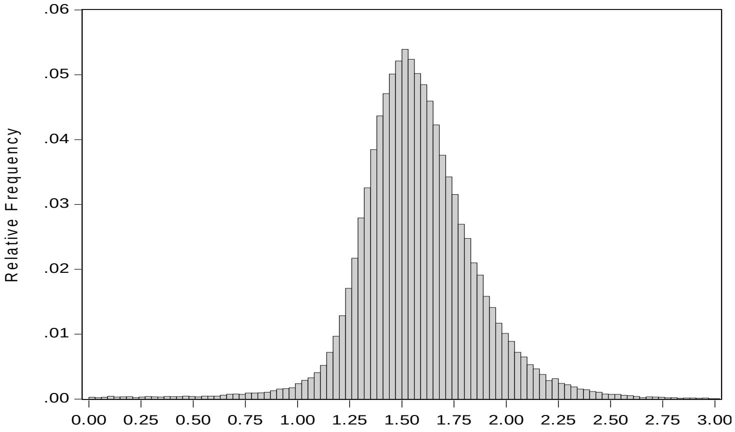
$$e_t \mid D_t \sim i.i.d\ N\left(0, (1 - D_t)\,\sigma_0^2 + D_t\,\sigma_1^2\right),$$

$$\Pr[S_t = 0 \mid S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1 \mid S_{t-1} = 1] = 1,$$

$$\Pr[D_t = 0 \mid D_{t-1} = 0] = q_{00}, \quad \Pr[D_t = 1 \mid D_{t-1} = 1] = 1.$$



$\theta_1 + \theta_2$



Long-Run Impulse Response Coefficient

Note:

1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009), namely the (log) quarterly US real GDP series seasonally adjusted for the period 1947:1–1998:2.
2. The model is estimated by the MCMC algorithm by Kim and Kim (2013), Chib and Greenberg (1993). The total number of Bayesian MCMC iterations is 135,000 and the first 5,000 samples are discarded.

Figure 2.8.C. Trend-Cycle Decomposition and Impulse-Response Analysis: ARIMA Model with Unknown Break Points in Mean and Variance [Sample: 1947:1~1998:2]

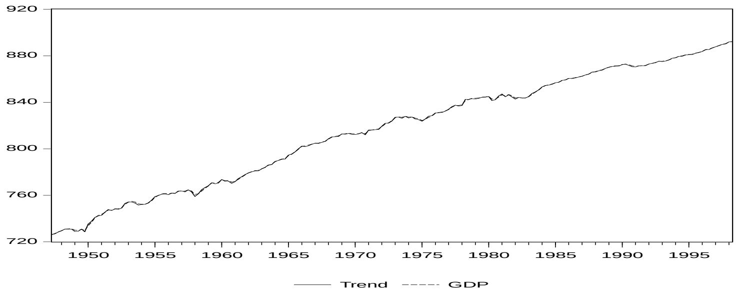
$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

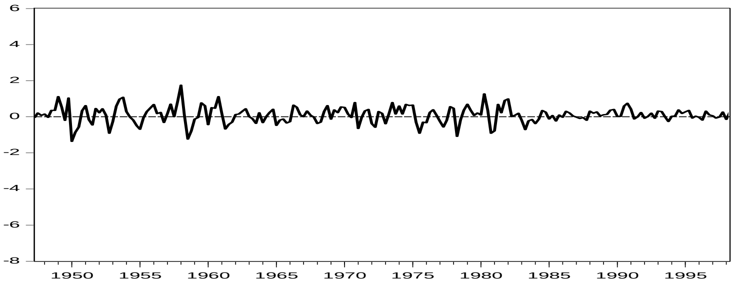
$$e_t \,|D_t \sim i.i.d\; N\,(0, (1 - D_t)\,\sigma_0^2 + D_t\,\sigma_1^2),$$

$$\Pr[S_t = 0|S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1|S_{t-1} = 1] = 1,$$

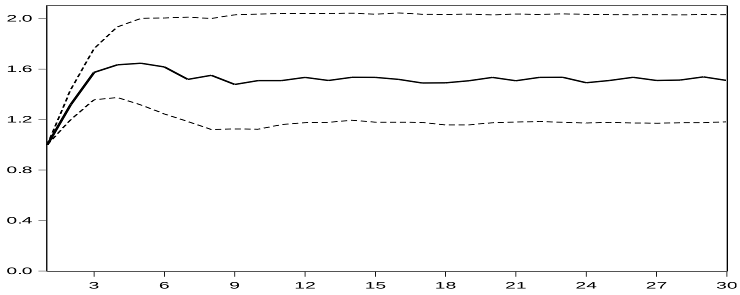
$$\Pr[D_t = 0|D_{t-1} = 0] = q_{00}, \quad \Pr[D_t = 1|D_{t-1} = 1] = 1.$$



Log of Real GDP vs. Trend Component



Cyclical Component



Impulse-Response Analysis

Note:

1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009), namely the (log) quarterly US real GDP series seasonally adjusted for the period 1947:1–1998:2.
2. The model is estimated by the MCMC algorithm by Kim and Kim (2013), Chib and Greenberg (1993). The total number of Bayesian MCMC iterations is 135,000 and the first 5,000 samples are discarded.

Figure 2.9.A. Bayesian Inference of Stochastic Volatility and Structural Break in Mean: ARIMA Model with Stochastic Volatility and Unknown Break Point in Mean [Sample: 1947:1~2007:4]

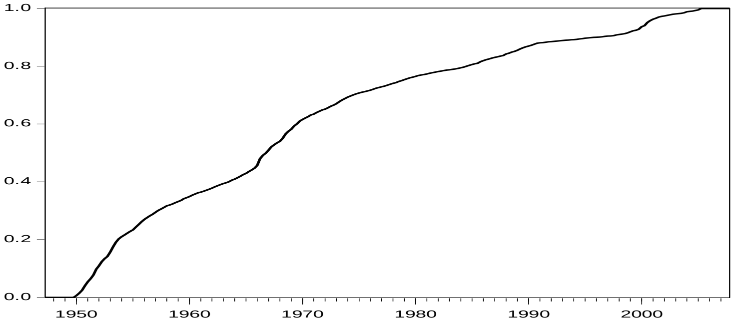
$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

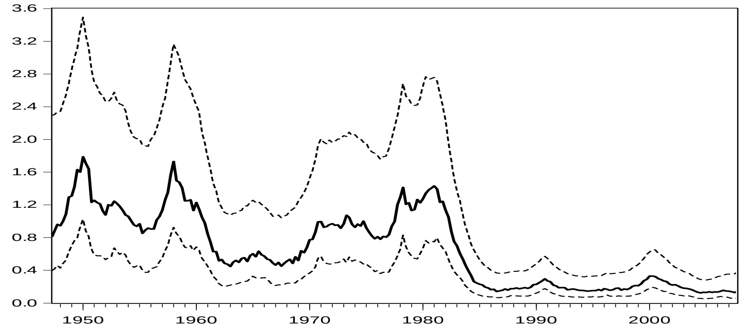
$$e_t \sim N(0, \sigma_t^2),$$

$$\Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1 | S_{t-1} = 1] = 1,$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t, \;\; \varepsilon_t \sim i.i.d. N(0, \sigma_\varepsilon^2).$$



Cumulative Posterior Probability of Structural Break in Mean



90% HPDI and Median of Stochastic Volatility

Note:
1. The raw data are (log) seasonally adjusted quarterly U.S. real GDP for the sample period of 1947:1 to 2007:4 and were taken from the St. Louis Fed (FRED) database.
2. The model is estimated by the MCMC algorithm by Kim and Kim (2013), Chib and Greenberg (1993) and Kim et al. (1998). The total number of Bayesian MCMC iterations is 135,000 and the first 5,000 samples are discarded.
3. A highest posterior density interval (HPDI) is an interval, the narrowest one possible with a chosen probability.
4. The initial value of σ_t^2 is treated as a parameter to be estimated.

Figure 2.9.B. Posterior Distributions for the Sum of MA Parameters and Long-Run Impulse-Response Coefficient: ARIMA Model with Stochastic Volatility and Unknown Break Point in Mean [Sample: 1947:1~2007:4]

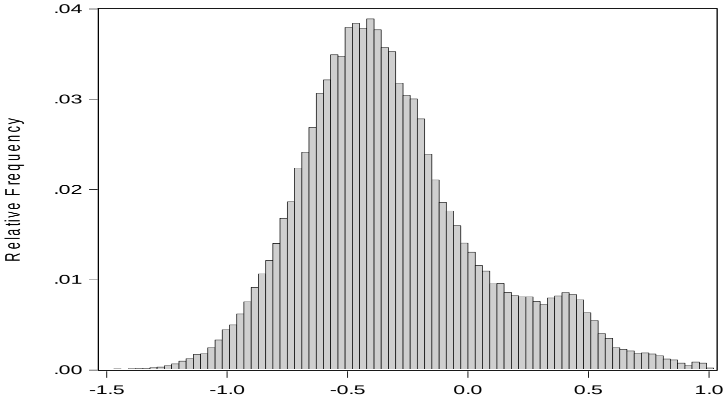
$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

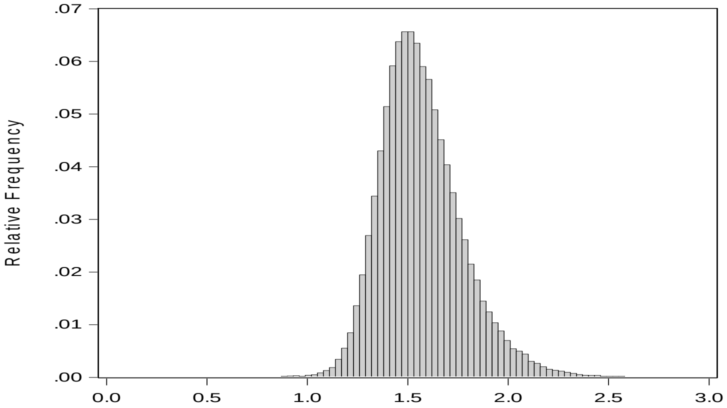
$$e_t \sim N(0, \sigma_t^2),$$

$$\Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1 | S_{t-1} = 1] = 1,$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t, \ \ \varepsilon_t \sim i.i.d.N(0, \sigma_\varepsilon^2).$$



$\theta_1 + \theta_2$



Long-Run Impulse Response Coefficient

Note:

1. The raw data are (log) seasonally adjusted quarterly U.S. real GDP for the sample period of 1947:1 to 2007:4 and were taken from the St. Louis Fed (FRED) database.
2. The model is estimated by the MCMC algorithm by Kim and Kim (2013), Chib and Greenberg (1993) and Kim et al. (1998). The total number of Bayesian MCMC iterations is 135,000 and the first 5,000 samples are discarded.
3. The initial value of σ_t^2 is treated as a parameter to be estimated.

Figure 2.9.C. Trend-Cycle Decomposition and Impulse-Response Analysis: ARIMA Model with Stochastic Volatility and Unknown Break Point in Mean [Sample: 1947:1~2007:4]

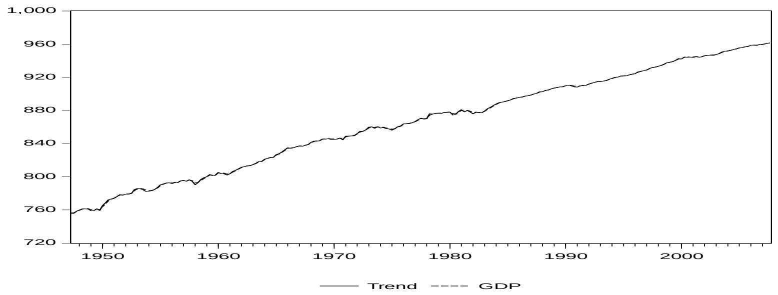
$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

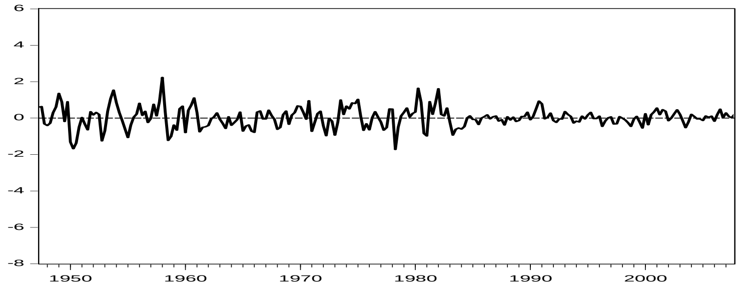
$$e_t \sim N(0, \sigma_t^2),$$

$$\Pr[S_t = 0 | S_{t-1} = 0] = p_{00}, \quad \Pr[S_t = 1 | S_{t-1} = 1] = 1,$$

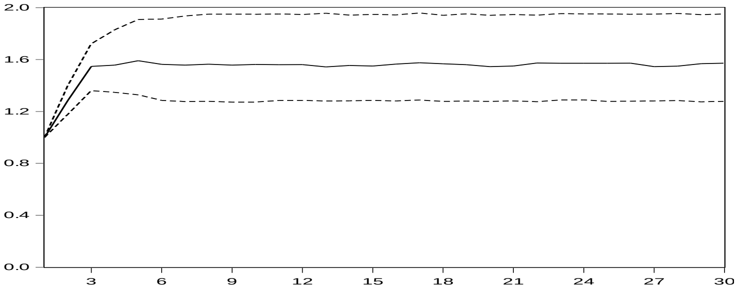
$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t, \; \varepsilon_t \sim i.i.d. N(0, \sigma_\varepsilon^2).$$



Log of Real GDP vs. Trend Component



Cyclical Component



Impulse-Response Analysis

Note:
1. The raw data are (log) seasonally adjusted quarterly U.S. real GDP for the sample period of 1947:1 to 2007:4 and were taken from the St. Louis Fed (FRED) database.
2. The model is estimated by the MCMC algorithm by Kim and Kim (2013), Chib and Greenberg (1993) and Kim et al. (1998). The total number of Bayesian MCMC iterations is 135,000 and the first 5,000 samples are discarded.
3. The initial value of σ_t^2 is treated as a parameter to be estimated.

Chapter 3

Robust Bayesian Model Selection for Nested Models

3.1. Introduction

A critical problem of *Bayes factor* on its sensitivity to prior specifications has been well documented in the literature. When comparing two competing models denoted by M_i and M_j, *Baye factor* for M_i against M_j is defined as below:

$$B_{i,j}(\tilde{Y}) = \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i)d_{\Psi_i}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j)d_{\Psi_j}} \tag{3.1}$$

where $\tilde{Y}$ presents observations and the likelihood and the prior distribution of the parameters, Ψ_k, are denoted by $f_k(\tilde{Y}|\Psi_k)$ and $\pi_k(\Psi_k)$, respectively under $M_k(k = i, j)$. In equation (3.1), a prior distribution can be considered as a weight function for likelihood values in obtaining a marginal likelihood. Consequently, a proper diffuse prior under a particular model automatically penalize the corresponding model in Bayesian model comparison. In the case of improper flat priors, which are defined only up to arbitrary constants, *Bayes Factor* becomes just a multiple of these constants and a likelihood ratio, which can result in a misleading statistical inference.

O'Hagan (1995) develop *Partial Bayes factor* to resolve the prior sensitivity problem. The underlying idea of the alternative approach is to update prior distributions with a proper portion of $\tilde{Y}$ to get rid of the sensitivity of a prior assumption. More specifically, a subsample of size l, denoted by $\tilde{Y}(l)$, is selected and employed to obtain the updated prior distribution $\pi_k(\Psi_k|\tilde{Y}(l))$, which is a posterior distribution with $\tilde{Y}(l)$. The selected subsample called a training sample mitigates the effects of the initially assumed prior on *Bayes factor*. *Partial Bayes factor* is computed with the updated prior, $\pi_k(\Psi_k|\tilde{Y}(l))$ and the remaining data $\tilde{Y}(-l)$. *Partial Bayes factor* is given by:

$$
\begin{aligned}
B_{i,j}(\tilde{Y}(-l)|\tilde{Y}(l)) &= \frac{\int f_i(\tilde{Y}(-l)|\Psi_i)\pi_i(\Psi_i|\tilde{Y}(l))d_{\Psi_i}}{\int f_j(\tilde{Y}(-l)|\Psi_j)\pi_j(\Psi_j|\tilde{Y}(l))d_{\Psi_j}} \\
&= \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i)d_{\Psi_i}}{\int f_i(\tilde{Y}(l)|\Psi_i)\pi_i(\Psi_i)d_{\Psi_i}} \; \frac{\int f_j(\tilde{Y}(l)|\Psi_j)\pi_j(\Psi_j)d_{\Psi_j}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j)d_{\Psi_j}}
\end{aligned}
\tag{3.2}
$$

A practical difficulty of *Partial Bayes factor* is selection of a particular training sample. Berger and Pericchi (1996) propose using all possible training samples with the minimal training sample size and averaging all resulting *Partial Bayes factors*, which is called *Intrinsic Bayes factor*. When the number of possible training samples is too large in practice, they suggest averaging *Partial Bayes factors* from a random collection of all training samples. *Intrinsic Bayes factor* by Berger and Pericchi (1996) can resolve the dependency problem on selection of a particular training sample. In obtaining *Intrinsic Bayes factor*, they consider both arithmetic and geometric means, which are defined as:

$$
\begin{aligned}
B^{AI}_{i,j}(\tilde{Y}) &= \frac{1}{L}\sum B_{i,j}(\tilde{Y}(l)|\tilde{Y}(-l)) \\
B^{GI}_{i,j}(\tilde{Y}) &= \{\prod B_{i,j}(\tilde{Y}(l)|\tilde{Y}(-l))\}^{\frac{1}{L}}
\end{aligned}
\tag{3.3}
$$

where L is the number of the selected training samples, and $B_{i,j}(\tilde{Y}(l)|\tilde{Y}(-l))$ is *Partial Bayes factor* in equation (3.2). Although *Intrinsic Bayes factor* could be a proper solution for the dependence of *Partial Bayes factor* on a specific training sample, it requires huge computations to calculate individual *Partial Bayes factors*.

Among the different proposals to remove the dependence problem, *Fractional Bayes factor* by O'Hagan (1995) has been widely used. He suggests replacing the likelihood $f_k(\tilde{Y}(l)|\Psi_k)$ in *Partial Bayes factor* with the training likelihood $f_k(\tilde{Y}|\Psi_k)^b$ where b is a ratio of the numbers of observations on total data and its training sample. He shows that when the numbers are large enough, the likelihood $f_k(\tilde{Y}(l)|\Psi_k)$ based only on a training sample $\tilde{Y}(l)$approximates to the full likelihood $f_k(\tilde{Y}|\Psi_k)$ raised to the power b. The *Bayes factor* with the training likelihood is called *Fractional Bayes factor*. *Fractional Bayes factor* for M_i against M_j is given by:

$$
B_{i,j}(\tilde{Y},b) = \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i)d_{\Psi_i}}{\int f_i(\tilde{Y}|\Psi_i)^b\pi_i(\Psi_i)d_{\Psi_i}} \; \frac{\int f_j(\tilde{Y}|\Psi_j)^b\pi_j(\Psi_j)d_{\Psi_j}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j)d_{\Psi_j}}
\tag{3.4}
$$

Bayesfactor in equation (3.1) and its alternatives in equation (3.2), (3.3), and (3.4) are not analytically tractable in practice. In the literature, several numerical methods have been proposed to estimate *Bayes factor*. Nevertheless, calculation of *Bayes factor* still remains a computationally complicated task. Alternative *Bayes factors* are even harder to evaluate in that they include larger sets of integrals than simple *Bayes factor*.

An central purpose of this chapter is to propose a simple and computationally efficient method for the alternative *Bayes factors*. An encompassing prior (EP) approach by Kluskist et al. (2005) and Hoijtink et al. (2008) is a recently proposed method to approximate *Bayes factor* in equation (1). Even if it applies only to nested models, it removes computational complications inherent in many other methods. By extending the EP approach to alternative *Bayes factors* in equation (2), (3), and (4), we show that the alternative *Bayes factors* can be significantly simplified in comparing nested models. This provides efficient and simple methods to conduct model comparisons for nested models, which is robust to prior specifications.

The chapter is organized as follows. We introduce general ideas and key tools in section 3.2. A simple Monte Carlo simulation is presented in section 3.3. Section 3.4 provides a summary and concluding remarks.

3.2. Alternative Bayes Factors with EP Approach

3.2.1 EP Approach with Inequality Constraints

In this section, we briefly introduce EP approach and explain our methods which are much simpler in calculating all the alternative *Bayes factors*. Our attention is limited to the issue of comparing nested models.

Suppose an encompassing model M_j has a parameter vector $\Psi_j = (\omega', \theta')'$ where a parameter vector with constraints is ω under a nested model and a vector of the remaining free parameters is θ. The prior distribution of the model parameters under the restricted model M_i is assumed as:

$$\pi_i(\omega, \theta) = \frac{\pi_j(\omega, \theta) I_{M_i}(\omega, \theta)}{\int \int \pi_j(\omega, \theta) I_{M_i}(\omega, \theta) d\omega d\theta} \tag{3.5}$$

where $I_{M_i}(\omega,\theta)$ is the indicator function of the model M_i. This implies that $I_{M_i}(\omega,\theta) = 1$ if parameter values are in accordance with the constraints imposed by model M_i. This prior assumption is not harmful because our training sample approach eventually gets rid of the sensitivity of alternative *Bayes factors* to prior assumptions.

With the above prior specification, Klugkist and Hoijtink (2007) show that the *Bayes Factor* in equation (1) can be easily estimated by sampling parameter values from the posterior and the prior distributions of the encompassing model M_j. The *Bayes Factor* under the prior specificantion in quation (3.5) can be approximated by [32]:

$$\begin{aligned} B_{i,j}(\tilde{Y}) &= \frac{\int\int \pi_j(\omega,\theta|\tilde{Y}) I_{M_i}(\omega,\theta) d\omega d\theta}{\int\int \pi_j(\omega,\theta) I_{M_i}(\omega,\theta) d\omega d\theta} \\ &\approx \frac{\frac{1}{m}\sum_{s=1}^{m} I_{M_i}(\omega^s,\theta^s|\tilde{Y},M_j)}{\frac{1}{n}\sum_{u=1}^{n} I_{M_i}(\omega^u,\theta^u|M_j)} \end{aligned} \tag{3.6}$$

where m and n are the total numbers of posterior and prior samples, respectively and $\pi_j(\omega,\theta|\tilde{Y})$ is the posterior distribution under M_j. Equation (3.6) implies that the numerator is the proportion of M_j's posterior samples for ω that are in accordance with the constraints by M_i, and the denominator is the proportion of M_j's prior samples for ω that are in accordance with the constraints by M_i.

We simplify the *Partial Bayes Factor* using EP approach for model comparison as below:

$$\begin{aligned} B_{i,j}(\tilde{Y}(-l)|\tilde{Y}(l)) &= \frac{\int f_i(\tilde{Y}(-l)|\Psi_i)\pi_i(\Psi_i|\tilde{Y}(l)) d_{\Psi_i}}{\int f_j(\tilde{Y}(-l)|\Psi_j)\pi_j(\Psi_j|\tilde{Y}(l)) d_{\Psi_j}} \\ &= \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i) d_{\Psi_i}}{\int f_i(\tilde{Y}(l)|\Psi_i)\pi_i(\Psi_i) d_{\Psi_i}} \; \frac{\int f_j(\tilde{Y}(l)|\Psi_j)\pi_j(\Psi_j) d_{\Psi_j}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j) d_{\Psi_j}} \\ &= \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i) d_{\Psi_i}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j) d_{\Psi_j}} \; \frac{\int f_j(\tilde{Y}(l)|\Psi_j)\pi_j(\Psi_j) d_{\Psi_j}}{\int f_i(\tilde{Y}(l)|\Psi_i)\pi_i(\Psi_i) d_{\Psi_i}} \\ &= B_{i,j}(\tilde{Y}) \frac{1}{B_{i,j}(\tilde{Y}(l))} \\ &= \frac{\int\int \pi_j(\omega,\theta|\tilde{Y}) I_{M_i}(\omega,\theta) d\omega d\theta}{\int\int \pi_j(\omega,\theta) I_{M_i}(\omega,\theta) d\omega d\theta} \frac{\int\int \pi_j(\omega,\theta) I_{M_i}(\omega,\theta) d\omega d\theta}{\int\int \pi_j(\omega,\theta|\tilde{Y}(l)) I_{M_i}(\omega,\theta) d\omega d\theta} \\ &\approx \frac{\frac{1}{m}\sum_{s=1}^{m} I_{M_i}(\omega^s,\theta^s|\tilde{Y},M_j)}{\frac{1}{m}\sum_{s=1}^{m} I_{M_i}(\omega^s,\theta^s|\tilde{Y}(l),M_j)} \end{aligned} \tag{3.7}$$

---

[32] The proof is shown in Appendix.

This is a very simple but powerful result. In the approximated *Partial Bayes Factor*, only the unrestricted model M_j is considered and it is not necessary to evaluate all the four integrals in equation (3.2). Therefore, if a researcher has a program for constructing the posterior distribution of the model M_j, using different data sets, $\tilde{Y}$ and $\tilde{Y}(l)$ the *Partial Bayes Factor* can be easily obtained from posterior and prior samplings. This result is applicable to model comparisons for any nested models. Extension to the *Intrinsic Bayes Factor* is straightforward because arithmetic and geometric *Intrinsic Bayes Factors* can be obtained from the approximated *Partial Bayes Factor* in equation (3.7).

Even if the simplified *Partial Bayes Factor* has some advantages in terms of prior sensitivity and computational efficiency, it is subject to the selection problem on a training sample. The *Intrinsic Bayes Factor* based on the simplified version of the *Partial Bayes Factor* resolves the dependency problem on a specific training sample at the cost of heavy computation. The *Fractional Bayes Factor* overcomes the dependency problem without a heavy computational cost. However, because it replaces the original likelihood with a training likelihood, implementing the *Fractional Bayes Factor* in practice is a complicating task not only analytically but also computationally. We employ EP approach to remove all the complicating issues in the use of the *Factional Bayes Factor* when comparing nested models within Bayesian framework.

It is, however, not straightforward to apply EP approach to the *Fractional Bayes Factor* since $f_j(\tilde{Y}|M_j)^b$ is not a usual likelihood. Suppose that $\int f_j(\tilde{Y}|M_j)^b d\tilde{Y}$ and $\int f_i(\tilde{Y}|M_i)^b d\tilde{Y}$ exist and are finite. The approximated *Fractional Bayes Factor* is given by:

$$\begin{aligned}
B_{i,j}(\tilde{Y},b) &= \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i)d_{\Psi_i}}{\int f_i(\tilde{Y}|\Psi_i)^b\pi_i(\Psi_i)d_{\Psi_i}} \; \frac{\int f_j(\tilde{Y}|\Psi_j)^b\pi_j(\Psi_j)d_{\Psi_j}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j)d_{\Psi_j}} \\
&= \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i)d_{\Psi_i}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j)d_{\Psi_j}} \; \frac{\int f_j(\tilde{Y}|\Psi_j)^b\pi_j(\Psi_j)d_{\Psi_j}}{\int f_i(\tilde{Y}|\Psi_i)^b\pi_i(\Psi_i)d_{\Psi_i}} \\
&= \frac{\int\int \pi_j(\omega,\theta|\tilde{Y})I_{M_i}(\omega,\theta)d\omega d\theta}{\int\int \pi_j(\omega,\theta)I_{M_i}(\omega,\theta)d\omega d\theta} \frac{\int\int \pi_j(\omega,\theta)I_{M_i}(\omega,\theta)d\omega d\theta}{\int\int \pi_{j,b}(\omega,\theta|\tilde{Y})I_{M_i}(\omega,\theta)d\omega d\theta} \\
&\approx \frac{\frac{1}{m}\sum_{s=1}^{m} I_{M_i}(\omega^s,\theta^s|\tilde{Y},M_j)}{\frac{1}{m}\sum_{s=1}^{m} I_{M_i}(\omega^s,\theta^s|\tilde{Y}(l),M_j,b)}
\end{aligned} \tag{3.8}$$

where $0 \le b \le 1$; $\pi_{j,b}(\omega,\theta|\tilde{Y})$ is the posterior distribution from the training likelihood and the prior distribution under M_j; $I_{M_i}(\omega, \theta|\tilde{Y}, M_j, b)$ is the indicator function of model M_i conditional on the modified posterior distribution $\pi_{j,b}(\omega, \theta|\tilde{Y})$ under M_j. By treating $\frac{f_j(\tilde{Y}|\Psi_j)}{c(b,\Psi_j)}$ and $\frac{f_i(\tilde{Y}|\Psi_i)}{c(b,\Psi_i)}$ as new likelihoods, where $c(.)$ is a normalizing constant, we can show the justification to apply EP approach in the presence of the training likelihood. Details of the proof are given in Appendix.

Drawing samples from the modified posterior with the training likelihood may not be straightforward because it does not seem to be a known distribution. However, in MCMC algorithm many conditional posterior distributions in the exponential family are still the same distributions after taking a power b to a likelihood function. Consequently, posterior sampling from $\pi_{j,b}(\omega, \theta|\tilde{Y})$ can be achieved with well-known distributions in most cases. When the corresponding conditional posteriors with a training likelihood are not well-known distributions, Metropolis-Hastings algorithm can be used for posterior sampling. The suggested method allows us to enjoy a simple and efficient method to conduct model comparison for any nested models.

3.2.2 EP Approach with Equality Constraints

In some situations, equality constraints are required for model comparison. However, EP approach dose not extend to the case of the exact equality constraints in a straightforward manner. The proportion of prior and posterior samples that are agreement with the exact constraints is zero which is called the *Borel-Kolmogorov* paradox. Laudy (2006) and Klugkist (2008) suggest an iterative procedure of EP approach for exact equality constraints. Each step of the procedure is presented in Appendix.

Rubb et al. (2010) point out that EP approach generally works well when nested models have inequality constraints. However, when it comes to exact constraints, EP approach with the suggested iterative procedure becomes vulnerable to the dependency on constraint parameterizations. For example, suppose that we want to test whether two random variables x and y are same. The exact constraint can be expressed in two different ways:

$$i)\ \ x - y = 0$$

$$ii)\ \ \frac{x}{y} = 1$$

They show EP approach with exact constraints could result in different values depending on constraint parameterizations by using the simple example. However, integrating EP approach and alternative Bayes factors allows us to easily overcome this problem. The problem of the iterative procedure is caused mainly by the difference in the convergence speeds of the prior and the posterior distributions of equation (6). Since only posterior distributions are considered in our proposal, the problem is no longer our concern. As a result, these two lines of research blend well together to overcome the problems.

The iterative procedure by Laudy (2006) and Klugkist (2008), however, is only approximation and time consuming process. Rubb et al. (2010) show that EP approach with exact constraints turns out to be identical to Savage-Dickey density ratio which is more principled and fast. Hence, we show that combing alternative Bayes factors with Savage-Dickey density ratio offers an nicer way to achieve our goal with the case of exact constraints.

When a nested model M_i has exact constraints, we can employ *Savage-Dickey density ratio* by Verdinelli and Wasserman (1995) under the following prior specification:

$$\pi_j(\theta|\omega = \omega_0) = \pi_i(\theta) \tag{3.9}$$

Then, the *Bayes Factor* in equation (1) comparing M_i and M_j, has the form:

$$B_{i,j}(\tilde{Y}) = \frac{\pi_j(\omega = \omega_0|\tilde{Y})}{\pi_j(\omega = \omega_0)} \tag{3.10}$$

where $\pi_j(\omega = \omega_0|\tilde{Y})$ and $\pi_j(\omega = \omega_0)$ are the posterior and the prior evaluated at ω_0 under M_j.

We can simplify the *Partial Bayes Factor* using *Savage-Dickey density ratio* for model comparison between M_i and M_j as in the case of inequality constraints:

$$\begin{aligned} B_{i,j}(\tilde{Y}(-l)|\tilde{Y}(l)) &= B_{i,j}(\tilde{Y})\frac{1}{B_{i,j}(\tilde{Y}(l))} \\ &= \frac{\pi_j(\omega = \omega_0|\tilde{Y})}{\pi_j(\omega = \omega_0)}\frac{\pi_j(\omega = \omega_0)}{\pi_j(\omega = \omega_0|\tilde{Y}(l))} \\ &= \frac{\pi_j(\omega = \omega_0|\tilde{Y})}{\pi_j(\omega = \omega_0|\tilde{Y}(l))} \end{aligned} \tag{3.11}$$

$\pi_j(\omega = \omega_0|\tilde{Y})$ and $\pi_j(\omega = \omega_0|\tilde{Y}(l))$ can be approximated by numerical integration from posterior samplings. Arithmetic and geometric *Intrinsic Bayes Factors* can be easily obtained from the above simplified *Partial Bayes Factor*.

The *Fractional Bayes Factor* is simplified as:

$$\begin{aligned}
B_{i,j}(\tilde{Y}, b) &= \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i)d_{\Psi_i}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j)d_{\Psi_j}} \frac{\int f_j(\tilde{Y}|\Psi_j)^b\pi_j(\Psi_j)d_{\Psi_j}}{\int f_i(\tilde{Y}|\Psi_i)^b\pi_i(\Psi_i)d_{\Psi_i}} \\
&= \frac{\pi_j(\omega = \omega_0|\tilde{Y})}{\pi_j(\omega = \omega_0)} \frac{\pi_j(\omega = \omega_0)}{\pi_{j,b}(\omega = \omega_0|\tilde{Y})} \\
&= \frac{\pi_j(\omega = \omega_0|\tilde{Y})}{\pi_{j,b}(\omega = \omega_0|\tilde{Y})}
\end{aligned} \tag{3.12}$$

where $\pi_{j,b}(\omega = \omega_0|\tilde{Y})$ is the modified posterior distributions with a training likelihood evaluated at ω_0 under of M_j. With the above results, model comparison for nested models can be done in a simple and efficient way.

3.3. Monte-Carlo Simulation

Consider the competing 2 models to check reliability of the simplified *Fractional Bayes Factor*:

Model #1: Encompassing Model

$$y_t = x_{1,t}\beta_1 + x_{2,t}\beta_2 + \epsilon_t \tag{3.13}$$

Model #2: Constrained Model

$$y_t = x_{1,t}\beta_1 + x_{2,t}\beta_2 + \epsilon_t, \quad \beta_1 < \beta_2 \tag{3.14}$$

where ϵ_t follows $N(0, \sigma^2)$ in both equation (3.13) and (3.14). Since *Model #2* is nested in *Model #1*, we can employ EP approach to simplify the *Fractional Bayes Factor* with the prior assumption in equation (3.5). We generate 10,000 samples of 100 observations from $Model\#1$ assuming different values of β_2, and for each sample we compute the *Fractional Bayes Factor* with EP approach. Based on the computed *Fractional Bayes Factors* we record the proportion of the samples for which resulting *Fractional Bayes Factor* supports the nested model. Table 3.1 shows the result of our Monte-Carlo experiment. In the most of times, when the true model is *Model #2*, resulting *Fractional Bayes Factors* prefer *Model #2*. As the parameter β_2 moves away from the constraint by *Model #2*, the proportion falls dramatically, which indicates that the approximated *Fractional Bayes Factor* performs well in comparing the competing models.

3.4 Summary and Conclusion

Bayesian model comparison is often achieved by the *Bayes Factor* which is sensitive to prior assumptions. Various alternative *Bayes Factors* such as the *Intrinsic* and *Fractional Bayes Factors* have been proposed to overcome this problem. However, practical problems arise since they include many marginal likelihoods which are not analytically tractable in most cases. An encompassing prior approach (EP) is a recently proposed method to approximate the *Bayes Factor* numerically in comparing nested models. We extend EP approach to the alternative *Bayes factors* in this paper. Our method provides a simple and elegant way to conduct robust model comparisons to the prior sensitivity for nested models.

Table 3.1. Proportion of Nested Model Supported

$$Model\ \#1: \quad y_t = x_{1,t}\beta_1 + x_{2,t}\beta_2 + \varepsilon_t,$$

$$Model\ \#2: \quad y_t = x_{1,t}\beta_1 + x_{2,t}\beta_2 + \varepsilon_t, \quad where\ \beta_1 < \beta_2$$

$$\beta_1 = 1.3,\ \sigma^2 = 0.001$$

β_2	*training sample size:1, b = 0.01*	*training sample size:5, b = 0.05*	*training sample size:10, b = 0.1*
1.1	0.0063	0.0067	0.0057
1.15	0.0280	0.0283	0.0275
1.2	0.1010	0.0993	0.1052
1.25	0.2556	0.2572	0.2620
1.3	0.5054	0.4989	0.4974
1.35	0.7402	0.7389	0.7306
1.4	0.9056	0.8934	0.8981
1.45	0.9705	0.9719	0.9706

Note:
1. Sample size = 100.
2. A diffuse prior is used for the simulation.

References

[1] Albert, James H., and Siddhartha Chib. “Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts.” *Journal of Business and Economic Statistics* 11.1 (1993): 1-15.

[2] Anderson, Theodore Wilbur, and Akimichi Takemura. “WHY DO NONINVERTIBLE ESTIMATED MOVING AVERAGES OCCUR?” *Journal of Time Series Analysis* 7.4 (1986): 235-254.

[3] Ansley, Craig F., and Paul Newbold. “Finite sample properties of estimators for autoregressive moving average models.” *Journal of Econometrics* 13.2 (1980): 159-183.

[4] Antoncic, Madelyn. “High and volatile real interest rates: Where does the Fed fit in?.” *Journal of Money, Credit and Banking* (1986): 18-27.

[5] Bai, Jushan, and Pierre Perron. “Computation and analysis of multiple structural change models.” *Journal of applied econometrics* 18.1 (2003): 1-22.

[6] Berger, James O., Brunero Liseo, and Robert L. Wolpert. “Integrated likelihood methods for eliminating nuisance parameters.” *Statistical Science* 14.1 (1999): 1-28.

[7] Berger, James O., and Luis R. Pericchi. “The intrinsic Bayes factor for model selection and prediction.” *Journal of the American Statistical Association* 91.433 (1996): 109-122.

[8] Beveridge, Stephen, and Charles R. Nelson. “A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’.” *Journal of Monetary economics* 7.2 (1981): 151-174.

[9] Billio, Monica, Alain Monfort, and Christian P. Robert. “Bayesian estimation of switching ARMA models.” *Journal of econometrics* 93.2 (1999): 229-255.

[10] Campbell, John Y., and N. Gregory Mankiw. “Are output fluctuations transitory?.” *The Quarterly Journal of Economics* 102.4 (1987): 857-880.

[11] Caporale, Tony, and Kevin B. Grier. “Political Regime Change and the Real Interest Rate.” *Journal of Money, Credit and Banking* 32.3 (2000): 320-334.

[12] Caporale, Tony, and Kevin Grier. “Inflation, presidents, Fed chairs, and regime shifts in the US real interest rate.” *Journal of Money, Credit, and Banking* 37.6 (2005): 1153-

1163.

[13] Carter, Chris K., and Robert Kohn. “On Gibbs sampling for state space models.” *Biometrika* 81.3 (1994): 541-553.

[14] Cheung, Yin-Wong, and Menzie D. Chinn. “Further investigation of the uncertain unit root in GNP.” *Journal of Business and Economic Statistics* 15.1 (1997): 68-73.

[15] Chib, Siddhartha. “Marginal likelihood from the Gibbs output.” *Journal of the American Statistical Association* 90.432 (1995): 1313-1321.

[16] Chib, Siddhartha. “Estimation and comparison of multiple change-point models.” *Journal of econometrics* 86.2 (1998): 221-241.

[17] Chib, Siddhartha, and Edward Greenberg. “Bayes inference in regression models with ARMA (p, q) errors.” *Journal of Econometrics* 64.1 (1994): 183-206.

[18] Chib, Siddhartha, and Edward Greenberg. “Understanding the metropolis-hastings algorithm.” *The American Statistician* 49.4 (1995): 327-335.

[19] Christiano, Lawrence J. “Searching for a Break in GNP.” *Journal of Business and Economic Statistics* 10.3 (1992): 237-250.

[20] Clark, Peter K. “The cyclical component of US economic activity.” *The Quarterly Journal of Economics* 102.4 (1987): 797-814.

[21] Cosslett, Stephen R., and Lung-Fei Lee. “Serial correlation in latent discrete variable models.” *Journal of Econometrics* 27.1 (1985): 79-97.

[22] Crowder, William J., and Dennis L. Hoffman. “The long-run relationship between nominal interest rates and inflation: the Fisher equation revisited.” *Journal of Money, Credit and Banking* (1996): 102-118.

[23] Davis Richard A., Meiching Chen, and William T. M. Dunsmuir. “Inference for MA(1) Processes with a Root on or near the Unit Circle,” *Probability and Mathematical Statistics* 15 (1995): 227-242.

[24] Davidson, James EH. “Problems with the estimation of moving average processes.” *Journal of Econometrics* 16.3 (1981): 295-310.

[25] DeJong, David N., John C. Nankervis, N. Eugene Savin, and Charles H. Whiteman. “Integration versus trend stationary in time series.” *Econometrica: Journal of the Econometric Society* (1992): 423-433.

[26] DeJong, David N., and Charles H. Whiteman. "Reconsidering 'Trends and random walks in macroeconomic time series'." *Journal of Monetary Economics* 28.2 (1991): 221-254.

[27] DeJong, David N., and Charles H. Whiteman. "Estimating moving average parameters: classical pileups and Bayesian posteriors." *Journal of Business and Economic Statistics* 11.3 (1993): 311-317.

[28] DeJong, Piet, and Neil Shephard. "The simulation smoother for time series models." *Biometrika* 82.2 (1995): 339-350.

[29] Diebold, Francis X., and Abdelhak S. Senhadji. "The uncertain unit root in real GNP: Comment." *The American Economic Review* (1996): 1291-1298.

[30] Diebolt, Jean, and Christian P. Robert. "Estimation of finite mixture distributions through Bayesian sampling." *Journal of the Royal Statistical Society. Series B (Methodological)* (1994): 363-375.

[31] Fama, Eugene F. "Short-term interest rates as predictors of inflation." *The American Economic Review* (1975): 269-282.

[32] Gali, Jordi. "How well does the IS-LM model fit postwar US data?." *The Quarterly Journal of Economics* 107.2 (1992): 709-738.

[33] Garbade, Kenneth, and Paul Wachtel. "Time variation in the relationship between inflation and interest rates." *Journal of Monetary Economics* 4.4 (1978): 755-765.

[34] Garcia, Rene, and Pierre Perron. "An analysis of the real interest rate under regime shifts." *The Review of Economics and Statistics* (1996): 111-125.

[35] Gospodinov, Nikolay. "Bootstrap-based inference in models with a nearly noninvertible moving average component." *Journal of Business and Economic Statistics* 20.2 (2002): 254-268.

[36] Hamilton, James D. "Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates." *Journal of Economic Dynamics and Control* 12.2 (1988): 385-423.

[37] Hamilton, James D. "A new approach to the economic analysis of nonstationary time series and the business cycle." *Econometrica: Journal of the Econometric Society* (1989): 357-384.

[38] Harrison, P.J. and C.F. Stevens. "Bayesian forecasting". *Journal of the Royal Statistical*

Society B, 38 (1976): 205-247.

[39] Harvey, Andrew C. “Trends and cycles in macroeconomic time series.” *Journal of Business and Economic Statistics* 3.3 (1985): 216-227.

[40] Hoijtink, Herbert, Irene Klugkist, and Paul A. Boelen. “Bayesian evaluation of informative hypotheses.” *New York: Springer* (2008).

[41] Huizinga, John, and Frederic S. Mishkin. “Monetary policy regime shifts and the unusual behavior of real interest rates.” *Carnegie-Rochester Conference Series on Public Policy. Vol. 24. North-Holland*, (1986).

[42] Kang, K. M. “A Comparison of Estimators for Moving Average Processes.” *Unpublished Paper, Australian Bureau of Statistics* (1975).

[43] Kim, Chang-Jin. “Dynamic linear models with Markov-switching.” *Journal of Econometrics* 60.1 (1994): 1-22.

[44] Kim, Chang-Jin and Jaeho Kim. “Bayesian Inference of Regime-Switching ARMA Models with Absorbing States: Dynamics of Ex-Ante Real Interest Rate Under Structural Breaks.” *Working Paper, University of Washington* (2013).

[45] Kim, Chang-Jin, and Charles R. Nelson. “Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle.” *Review of Economics and Statistics* 81.4 (1999): 608-616.

[46] Kim, Sangjoon, Neil Shephard, and Siddhartha Chib. “Stochastic volatility: likelihood inference and comparison with ARCH models.” *The Review of Economic Studies* 65.3 (1998): 361-393.

[47] King, Robert, Charles I. Plosser, James H. Stock, and Mark W. Watson. ”Stochastic trends and economic fluctuations.” *American Economic Review* 81.4 (1991): 819-840.

[48] Klugkist, Irene, Bernet Kato, and Herbert Hoijtink. “Bayesian model selection using encompassing priors.” *Statistica Neerlandica* 59.1 (2005): 57-69.

[49] Klugkist, Irene, and Herbert Hoijtink. “The Bayes factor for inequality and about equality constrained models.” *Computational Statistics and Data Analysis* 51.12 (2007): 6367-6379.

[50] Klugkist, Irene, Olav Laudy, and Herbert Hoijtink. “Inequality constrained analysis of variance: a Bayesian approach.” *Psychological Methods* 10.4 (2005): 477.

[51] Koop, G. "Bayesian econometrics." *John Wiley and Sons* (2003).

[52] Koustas, Zisimos, and Apostolos Serletis. "On the Fisher effect." *Journal of Monetary Economics* 44.1 (1999): 105-130.

[53] Laudy, O. "Bayesian inequality constrained models for categorical data". *Ph.D. Thesis, Utrecht university* (2006).

[54] Liu, Jun S., Wing Hung Wong, and Augustine Kong. "Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes." *Biometrika* 81.1 (1994): 27-40.

[55] Luo, Sui, and Richard Startz. "Is It One Break or Ongoing Permanent Shocks That Explains U.S. Real GDP?" *Working Paper* (2013)

[56] McConnell, Margaret M., and Gabriel Perez-Quiros. "Output fluctuations in the United States: What has changed since the early 1980's?" *American Economic Review* (2000): 1464-1476.

[57] Mishkin, Frederic S. "The real interest rate: an empirical investigation." *Carnegie-Rochester Conference Series on Public Policy. Vol. 15. North-Holland* (1981).

[58] Mishkin, Frederic S. "Is the Fisher effect for real?: A reexamination of the relationship between inflation and interest rates." *Journal of Monetary Economics* 30.2 (1992): 195-215.

[59] Morley, James C., Charles R. Nelson, and Eric Zivot. "Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?" *Review of Economics and Statistics* 85.2 (2003): 235-243.

[60] Murray, Christian J., and Charles R. Nelson. "The uncertain trend in US GDP." *Journal of Monetary Economics* 46.1 (2000): 79-95.

[61] Murray, Christian J., and Charles R. Nelson. "The great depression and output persistence." *Journal of Money, Credit and Banking* (2002): 1090-1098.

[62] Nakatsuma, Teruo. "Bayesian analysis of ARMAGARCH models: A Markov chain sampling approach." *Journal of Econometrics* 95.1 (2000): 57-69.

[63] Neely, Christopher J., and David E. Rapach. "Real interest rate persistence: evidence and implications." *Federal Reserve Bank of St. Louis Review* 90.November/December (2008).

[64] Nelson, Charles R., and Charles R. Plosser. "Trends and random walks in macroeconmic time series: some evidence and implications." *Journal of monetary economics* 10.2 (1982): 139-162.

[65] Nelson, Charles R., and G. William Schwert. "Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant." *The American Economic Review* (1977): 478-486.

[66] Newbold, Paul, Stephen Leybourne, and Mark E. Wohar. "Trend-stationarity, difference-stationarity, or neither: further diagnostic tests with an application to US Real GNP, 18751993." *Journal of Economics and Business* 53.1 (2001): 85-102.

[67] O'Hagan, Anthony. "Fractional Bayes factors for model comparison." *Journal of the Royal Statistical Society. Series B (Methodological)* (1995): 99-138.

[68] Perron, Pierre. "The great crash, the oil price shock, and the unit root hypothesis." *Econometrica: Journal of the Econometric Society* (1989): 1361-1401.

[69] Perron, Pierre. "Testing for a unit root in a time series with a changing mean." *Journal of Business and Economic Statistics* 8.2 (1990): 153-162.

[70] Perron, Pierre, and Tatsuma Wada. "Let's take a break: Trends and cycles in US real GDP." *Journal of Monetary Economics* 56.6 (2009): 749-765.

[71] Pierce, David A. "Least squares estimation in the regression model with autoregressive-moving average errors." *Biometrika* 58.2 (1971): 299-312.

[72] Plosser, Charles I., and G. William Schwert. "Estimation of a non-invertible moving average process: The case of overdifferencing." *Journal of Econometrics* 6.2 (1977): 199-224.

[73] Rapach, David E., and Christian E. Weber. "Are real interest rates really nonstationary? New evidence from tests with good size and power." *Journal of Macroeconomics* 26.3 (2004): 409-430.

[74] Rose, Andrew K. "Is the real interest rate stable?." *The Journal of Finance* 43.5 (1988): 1095-1112.

[75] Sargan, John D., and Alok Bhargava. "Maximum likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle." *Econometrica: Journal of the Econometric Society* (1983): 799-820.

[76] Scott, Steven L. "Bayesian methods for hidden Markov models." *Journal of the American Statistical Association* 97.457 (2002).

[77] Shephard, Neil. "Partial non-Gaussian state space." *Biometrika* 81.1 (1994): 115-131.

[78] Shephard, Neil G., and Andrew C. Harvey. "On the probability of estimating a deterministic component in the local level model." *Journal of time series analysis* 11.4 (1990): 339-347.

[79] Shephard, Neil. "Maximum likelihood estimation of regression models with stochastic trend components." *Journal of the American Statistical Association* 88.422 (1993): 590-595.

[80] Sims, Christopher A., and Harald Uhlig. "Understanding unit rooters: A helicopter tour." *Econometrica: Journal of the Econometric Society* (1991): 1591-1599.

[81] Smith, Richard L., and J. C. Naylor. "A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution." *Applied Statistics* 36.3 (1987): 358-369.

[82] Stock, James H. "Unit roots, structural breaks and trends." *Handbook of econometrics* 4 (1994): 2739-2841.

[83] Stock, James H., and Mark W. Watson. "Median unbiased estimation of coefficient variance in a time-varying parameter model." *Journal of the American Statistical Association* 93.441 (1998): 349-358.

[84] Sun, Yixiao, and Peter CB Phillips. "Understanding the Fisher equation." *Journal of Applied Econometrics* 19.7 (2004): 869-886.

[85] Tanaka, Katsuto, and Stephen E. Satchell. "Asymptotic properties of the maximum-likelihood and nonlinear least-squares estimators for noninvertible moving average models." *Econometric Theory* 5.03 (1989): 333-353.

[86] Verdinelli, Isabella, and Larry Wasserman. "Computing Bayes factors using a generalization of the Savage-Dickey density ratio." *Journal of the American Statistical Association* 90.430 (1995): 614-618.

[87] Wetzels, Ruud, Raoul PPP Grasman, and Eric-Jan Wagenmakers. "An encompassing prior generalization of the SavageDickey density ratio." *Computational Statistics and Data Analysis* 54.9 (2010): 2094-2102.

[88] Zivot, Eric, and Donald W. K. Andrews. “Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis.” *Journal of Business and Economic Statistics* 20.1 (2002): 25-44.

[89] Zivot, Eric, and Donald WK Andrews. “Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis.” *Journal of Business and Economic Statistics* 10.3 (1992): 251-270.

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:

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

$$y_t^* = y_t - \Sigma_{i=1}^{p}\phi_i \ y_{t-i} - \Sigma_{j=1}^{q}\theta_i \ y_{t-j}^*,$$

$$x_t^* = x_t - \Sigma_{i=1}^{p}\phi_i \ x_{t-i} - \Sigma_{j=1}^{q}\theta_j \ x_{t-j}^*,$$

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:

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

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

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

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

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

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:

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

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

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

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

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

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

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).

Appendix C. Encompassing Prior Approach

In this appendix, we will show how the encompassing prior approach can approximate the default *Bayes Factor* in equation (3.1). This is originally shown by Klugkist and Hoijtink (2007). Consider the following marginal likelihood under a model, M_a:

$$p(\tilde{Y}|M_a) = \frac{f(\tilde{Y}|\Psi_a)\pi_a(\Psi_a)}{\pi_a(\Psi_a|\tilde{Y})} \qquad (C.1)$$

for all $\Psi_a \in \Theta_a$, where Θ_a denotes the parameter space of model M_a for $a = i, j$. M_j and M_i represent the encompassing model and the constrained model, respectively. The numerator is the product of the likelihood and the prior and the denominator is the posterior distribution as shown by Chib (1995). According to equation (C.1), the default *Bayes Factor* in equation (1) for M_i against M_j is given by:

$$\begin{aligned} B_{i,j}(\tilde{Y}) &= \frac{\int f_i(\tilde{Y}|\Psi_i)\pi_i(\Psi_i)d_{\Psi_i}}{\int f_j(\tilde{Y}|\Psi_j)\pi_j(\Psi_j)d_{\Psi_j}} \\ &= \frac{p(\tilde{Y}|M_i)}{p(\tilde{Y}|M_j)} \\ &= \frac{f(\tilde{Y}|\Psi_i)\pi_i(\Psi_i)}{\pi_i(\Psi_i|\tilde{Y})} \quad \frac{\pi_j(\Psi_j|\tilde{Y})}{f(\tilde{Y}|\Psi_j)\pi_j(\Psi_j)} \\ &= \frac{\pi_i(\Psi_i)/\pi_i(\Psi_i|\tilde{Y})}{\pi_j(\Psi_j)/\pi_j(\Psi_j|\tilde{Y})} \end{aligned} \qquad (C.2)$$

Form the third line to the fourth line, we use the fact that if $\Psi^* \in \Theta_i$, then $\Psi^* \in \Theta_j$ and $f_i(\tilde{Y}|\Psi^*) = f_j(\tilde{Y}|\Psi^*)$.

Note that, Ψ^* consists of two sets of parameters, ω^*, θ^* where ω^* is the set of constrained parameters. Applying the assumption in equation (3.5), we can get the following prior:

$$\pi_i(\omega^*, \theta^*) = \frac{I_{M_i}(\omega^*, \theta^*)}{\int\int \pi_j(\omega^*, \theta^*)I_{M_i}(\omega^*, \theta^*)d\omega^* d\theta^*}\pi_j(\omega^*, \theta^*) \qquad (C.3)$$

and

$$\pi_i(\omega^*, \theta^*|\tilde{Y}) = \frac{I_{M_i}(\omega^*, \theta^*)}{\int\int \pi_j(\omega^*, \theta^*|\tilde{Y})I_{M_i}(\omega^*, \theta^*)d\omega^* d\theta^*}\pi_j(\omega^*, \theta^*|\tilde{Y}) \qquad (C.4)$$

Therefore, equation (C.2) can be further simplified as:

$$\begin{aligned} B_{i,j}(\tilde{Y}) &= \frac{\int\int \pi_j(\omega^*, \theta^* | \tilde{Y}) I_{M_i}(\omega^*, \theta^*) d\omega^* d\theta^*}{\int\int \pi_j(\omega^*, \theta^*) I_{M_i}(\omega^*, \theta^*) d\omega^* d\theta^*} \\ &\approx \frac{\frac{1}{m}\sum_{s=1}^{m} I_{M_i}(\omega^s, \theta^s | \tilde{Y}, M_j)}{\frac{1}{m}\sum_{s=1}^{m} I_{M_i}(\omega^s, \theta^s | M_j)} \end{aligned} \tag{C.5}$$

where $\int\int \pi_j(\omega^*, \theta^* | \tilde{Y}) I_{M_i}(\omega^*, \theta^*) d\omega^* d\theta^*$ and $\int\int \pi_j(\omega^*, \theta^*) I_{M_i}(\omega^*, \theta^*) d\omega^* d\theta^*$ represent the portions of the posterior and prior distributions that are in agreement with the constraint under M_i; m is the total number of MCMC samples for the posterior and the prior of Ψ; $I_{M_i}(\omega, \theta | M_j)$ is the indicator function of model M_i conditional on the prior distribution under the model M_j; $I_{M_i}(\omega, \theta | \tilde{Y}, M_j)$ is the indicator function of model M_i conditional on the posterior distribution under the model M_j.

Appendix D. Fractional Bayes Factor with EP Approach

Suppose that $\int f_j(\tilde{Y}|\Psi_j)^b d\tilde{Y}$ and $\int f_i(\tilde{Y}|\Psi_i)^b d\tilde{Y}$ exist and are finite. Define the integrals as:

$$\begin{aligned} c(b,\Psi_i) &= \int f_i(\tilde{Y}|\Psi_i)^b d\tilde{Y} \\ c(b,\Psi_j) &= \int f_j(\tilde{Y}|\Psi_j)^b d\tilde{Y} \end{aligned} \tag{D.1}$$

From the second term of the *Fractional Bayes Factor* in the second line of equation (3.8),

$$\begin{aligned} \frac{\int f_j(\tilde{Y}|\Psi_j)^b \pi_j(\Psi_j) d\Psi_j}{\int f_i(\tilde{Y}|\Psi_i)^b \pi_i(\Psi_i) d\Psi_i} &= \frac{c(b,\Psi_j)\int \frac{f_j(\tilde{Y}|\Psi_j)^b}{c(b,\Psi_j)}\pi_j(\Psi_j)d\Psi_j}{c(b,\Psi_i)\int \frac{f_i(\tilde{Y}|\Psi_i)^b}{c(b,\Psi_i)}\pi_i(\Psi_i)d\Psi_i} \\ &= \frac{c(b,\Psi_j)\int f_j(\tilde{Y}|\Psi_j)^* \pi_j(\Psi_j)d\Psi_j}{c(b,\Psi_i)\int f_i(\tilde{Y}|\Psi_i)^* \pi_i(\Psi_i)d\Psi_i} \\ &= \frac{c(b,\Psi_j)}{c(b,\Psi_i)}\frac{f(\tilde{Y}|\Psi_i)^*\pi_i(\Psi_i)}{\pi_i(\Psi_i|\tilde{Y})} \quad \frac{\pi_j(\Psi_j|\tilde{Y})}{f(\tilde{Y}|\Psi_j)^*\pi_j(\Psi_j)} \\ &= \frac{\pi_i(\Psi_i)/\pi_i(\Psi_i|\tilde{Y})^*}{\pi_j(\Psi_j)/\pi_j(\Psi_j|\tilde{Y})^*} \end{aligned} \tag{D.2}$$

where $f_a(\tilde{Y}|\Psi_a)^*$ is a newly defined probability density for $a = i, j$. From the second to the fourth line, we apply the same fact used in equation (C.2). Applying the assumption in equation (3.5), finally we can get the following result:

$$\frac{\int f_j(\tilde{Y}|\Psi_j)^b \pi_j(\Psi_j) d\Psi_j}{\int f_i(\tilde{Y}|\Psi_i)^b \pi_i(\Psi_i) d\Psi_i} = \frac{\int\int \pi_j(\omega^*,\theta^*|\tilde{Y})^* I_{M_i}(\omega^*,\theta^*)d\omega^* d\theta^*}{\int\int \pi_j(\omega^*,\theta^*) I_{M_i}(\omega^*,\theta^*)d\omega^* d\theta^*} \tag{D.3}$$

Combining