HIAS-E-68

Some Empirical Evidence on Models of the Fisher Relation: Post-Data Comparison

Jae-Young Kim
Department of Economics, Seoul National University
Hitotsubashi Institute for Advanced Study, Hitotsubashi University

Woong Yong Park
Department of Economics, Seoul National Univeristy

April 2018

Hitotsubashi Institute for Advanced Study, Hitotsubashi University
2-1, Naka, Kunitachi, Tokyo 186-8601, Japan
tel:+81 42 580 8604 http://hias.ad.hit-u.ac.jp/

Some Empirical Evidence on Models of the Fisher Relation: Post-Data Comparison

Jae-Young Kim * Woong Yong Park †

April 2018

Abstract

The Fisher relation, describing a one-for-one relation between the nominal interest rate and the expected inflation, underlies many important results in economics and finance. Although it is a conceptually simple relation, the Fisher relation has more or less complicated with mixed results. There are several alternative models proposed in the empirical literature for the Fisher relation that have different implications. We evaluate those alternative models for the Fisher relation based on a post-data model determination method. Our results for data from the U.S. Japan and Korea show that models with both regimes/periods, a regime with nonstationary fluctuations and the other with stationary fluctuations, fit data best for the Fisher relation.

Keywords: Fisher relation, nonlinear behavior, post-data model determination.

JEL classification: C1, C22, C5.

*Department of Economics, Seoul National University and Hitotsubashi Institute for Advanced Studies (HIAS), Hitotsubashi University, Correspondence: Kwanak-ro 1 Kwanak-gu, Seoul, 08826, Korea. e-mail: jykim017@snu.ac.kr. The first author would like to thank HIAS for providing superior environment and support for finishing this work.

†Department of Economics, Seoul National University

1 Introduction

The Fisher relation underlies many important results in economics and finance. The relation explains that the nominal interest rate is determined as the sum of the expected inflation and the real interest rate which is constant or stable. The Fisher relation, thus, signifies that nominal interest rate has a statistical one-for-one relation with the expected rate of inflation. Although the Fisher relation looks a simple relation, there are several alternative models proposed in the literature for the Fisher relation that have different implications. In this paper, we examine the Fisher relation by evaluating those alternative models based on a post-data model determination method.

Empirical analysis on the Fisher relation was initiated by Fama (1975). Constancy of the real interest was studied by Nelson and Schwert (1977), Garbade and Wachtel (1978), Mishkin (1981, 1984), and Fama and Gibbons (1982). Correlation between the nominal interest rate and the inflation rate noted as the Fisher effect was studied by Summers (1982), Huizinga and Mishkin (1986), and Mishkin (1990). Also, Rose (1988), Atkins (1989), Mishkin (1992), and Wallace and Warner (1993) studied the real interest rate based on the concepts of a unit root and cointegration. On the other hand, it was studied by Clarida and Friedman (1984), Huizinga and Mishkin (1986), and Roley (1986) that change in the U.S. monetary policy in late 1970's through mid 1980's might have affected dynamics of interest rates and inflation. Evans and Lewis (1995) and Garcia and Perron (1996) used models of regime switch to analyze behavior of the U.S. real interest rate for post-war data including the period of policy regime change and oil shocks. Kim and Park (2016) studied the possibility of short-run instability as well as long-memory properties of the Fisher relation.

As described above there are several approaches and models for the Fisher relation proposed in the literature that have different implications. We, however, do not know which model is the most appropriate among the several models for the Fisher relation. This issue is an important one since different models have different implications for the Fisher relation, some of which are conflicting with each other. In this paper, we evaluate those alternative models for data from the U.S. Japan and Korea in the post war period before the 2007-2008 world financial crisis. For this purpose we apply a post-data model determination method to get the model that best fits the data. The post-data model selection method evaluates relative probability of each model among the alternatives. We use a Markov-Chain-Monte-Carlo (MCMC) method for computation of a criterion that quantifies the relative probability of each model. The model yielding the highest post-data probability is the one that best fits the data. We use Gibbs sampler to compute the relative post-data probabilities. Our results indicate that the best model is not the same for the three countries. However, models with both regimes/periods, a regime with nonstationary fluctuations and the other with stationary fluctuations, fit data best for the Fisher relation.

The paper is organized as follows. Section 2 introduces the Fisher relation and related issues. Section 3 explains several models for the Fisher relation proposed in the literature. In Section 4 we discuss how to select the model that is the most appropriate for the Fisher relation. Section 5 concludes the paper.

2 The Fisher Relation and Related Issues

The Fisher relation explains how the nominal interest rate is determined. Let π^e_{t+1} be the expected inflation from period t to period $t+1$. Also, let r^*_t and i_t, respectively, be the *ex-ante* real interest rate and the nominal interest rate at time t. Then, the Fisher relation defines that the nominal interest rate is equal to the real interest rate:

$$i^*_t = r_t + \pi^e_{t+1} + \varepsilon_t \tag{2.1}$$

allowing a temporary disturbance ε_t to the relation.

As is well explained in Kim and Park (2016) and others, the Fisher relation describes that the nominal interest rate has a one-for-one relation with the expected rate of inflation. In other words, the Fisher relation describes that there is a stable level of the "real interest rate" that is equal to the nominal interest rate minus expected inflation, allowing a temporary disturbance. In terms of *ex-ante* variables the relation writes as

$$r^*_t = i_t - \pi^e_{t+1} - \varepsilon_t. \tag{2.2}$$

We have an *ex-post* version of the Fisher relation as

$$r_t = i_t - \pi_{t+1} - \varepsilon_t. \tag{2.3}$$

where π_{t+1} and r_t are, respectively, *ex-post* inflation and *ex-post* real interest rate. Notice that we use the same notation for the disturbance ε_t in both cases of *ex-ante* and *ex-post* to save the notation.

Denoting by v_t the error of the inflation expectation: $v_t = \pi_t^e - \pi_t$, it is true that $r_t = r_t^* + v_t$. If v_t is stationary, which is the case under rational expectations, then the *ex ante* real interest rate r_t^* and the *ex post* real interest rate r_t have the same statistical properties. In this case one can analyze the Fisher relation based on the *ex post* interest rate as well as the *ex ante* rate. Existence of a stable Fisher relation is, in statistical sense, the same as that the real interest rate is a constant, or a stationary variable fluctuating around a constant mean. Thus, the Fisher relation is a simple relation in its concept. However, empirical analysis of the Fisher relation is more or less complicated with mixed results.

There are several alternative models proposed in the literature for the Fisher relation. They have different implications on the stationary and nonstationary behavior of the real interest rate. Therefore, the issue of which model is the most appropriate for the empirical Fisher relation is a very important problem. We evaluate those alternative models for the Fisher relation based on a post-data model determination method.

3 Models for the Real Interest Rate

In the following discussion we use the variable y_t for the real interest rate. Also, let $\mathfrak{T}^T = \{1, \cdots, T\}$ be the sample period. We have all four alternative models for the Fisher relation in the following, M_i, $i = 0, 1, 2, 3$.

3.1 An Autogression: M_0

The basic model is the p^{th} order autoregression in y_t:

$$(y_t - \mu) = \sum_{s=1}^{p} \phi_s (y_{t-s} - \mu) + \varepsilon_t, \tag{3.1}$$

where $\varepsilon_t \sim \text{iid}N(0, \sigma^2)$, and all roots of the characteristic equation $1 - \phi_1 z - \cdots - \phi_p z^p = 0$ lie outside the unit circle.

Notice that we can rewrite the model (3.1) as in the following:

$$y_t = \left(1 - \sum_{s=1}^{p} \phi_s\right) \mu + \sum_{s=1}^{p} \phi_s y_{t-s} + \varepsilon_t, \tag{3.2}$$

which is a more commonly used form in the usual time series analysis. We use the mean-deviated form (3.1) instead of the more common one (3.2) since the former is more convenient for adopting standard regime switching models in our study.

3.2 A Model with Partial-sample Instability: M_1

We incorporate the possibility of partial-sample (or short-run) instability in M_0 following the suggestion in Kim (2003), Andrews and Kim (2008) and Kim and Park (2016). Suppose that in a relatively short period $\mathfrak{T}_B^T$ ($\subset \mathfrak{T}^T$) the process y_t becomes unstable, having properties of a nonstationary unit root or of higher volatility. In the model M_1 we assume that $\mathfrak{T}_B^T$ is identified a priori, which is different from the model M_3 below. Then, we have the following model for y_t:

$$(y_t - \tau_t) = \sum_{s=1}^{p} [\phi_s \cdot I(t \in \mathfrak{T}_S) + \zeta_s \cdot I(t \in \mathfrak{T}_B)](y_{t-s} - \tau_{t-s}) + \varepsilon_t \tag{3.3}$$

where $\mathfrak{T}_S = \mathfrak{T} \setminus \mathfrak{T}_B$; $\varepsilon_t \sim \mathrm{N}(0, \nu_0^2)$; $I(\cdot)$ is the indicator function, and ϕ_s and ζ_s are parameters. For $t \in \mathfrak{T}_S$ we assume that the mean of y_t is $\tau_t = \mu_0$ and for $t \in \mathfrak{T}_B$ $\tau_t = \mu_1 + y_{t-1}$. This means that in the period $\mathfrak{T}_B$ the process y_t has a unit root and its first difference Δy_t is a stationary autoregressive process.

3.3 Markov Regime Switching Model: M_2

Assume that the variable y_t follows regime switching dynamics across K states $s_t = 1, 2, \cdots, K$:

$$(y_t - \tau_t) = \sum_{s=1}^{p} \phi_s (y_{t-s} - \tau_{t-s}) + \varepsilon_t. \tag{3.4}$$

for $\varepsilon_t \sim \mathrm{N}(0, \nu_t^2)$, where $\tau_t = \mu_{s_t}$ and $\nu_t^2 = \sigma_{s_t}^2$ in state s_t. We assume that $\mu_1 < \cdots < \mu_K$ for identification. The state variable s_t follows the first order Markov process with the transition probability from the state i to j, $p_{ij} = \mathrm{P}[s_t = j | s_{t-1} = i]$ for $i, j = 1, \cdots, K$. Garcia and Perron (1996) used models of regime switch of the M_2 type to anlyze behavior of the U.S. real interest rate for post-war data.

3.4 An Extended Markov Switching Model: M_3

We now consider a Markov switching model that contains a state of nonstationarity. It is an extension of M_2 in the above subsection that only contains K stationary regimes. In this extended model M_3, the last K^{th} state is set to be a nonstationary state. Then, the extended model M_3 is

$$(y_t - \tau_t) = \sum_{s=1}^{p} [\phi_s \cdot I(s_t \neq K) + \zeta_s \cdot I(s_t = K)](y_{t-s} - \tau_{t-s}) + \varepsilon_t \tag{3.5}$$

where $\varepsilon_t \sim \mathrm{N}(0, \nu_t^2)$. The mean of y_t in a stationary state is $\tau_t = \mu_{s_t}$ and that in the nonstationary state is $\tau_t = \mu_K + y_{t-1}$. This implies that the variable y_t has a unit root in the state s_K.

Remark: There is an alternative modeling scheme with the regime switching due to its own lagged variable, known as the self-exciting threshold regression. It was introduced by Tong and Lim (1980) and studied by Seo (2008), for instance, in relation to the unit root testing for the model. This alternative specification may be a relevant option for modeling the Fisher relation. We do not, however, consider this specification in this paper since our objective is to evaluate the existing models of the Fisher relation. This modeling scheme can be applied to the Fisher relation in any future work.

4 Model Selection For the Fisher Relation

In this section we discuss how to compare different models and select the one that best fits data for the Fisher relation. Then, we provide the result of selecting a model among those explained in Section 3.

4.1 Post-data Model Selection

In this subsection we explain how to determine the best model for the real interest rate out of several alternatives. Our approach is a post data model selection method developed in the Bayesian framework. Thus, it is a generalized version of the Bayesian information criterion. In the method we can evaluate relative merit of each model among several alternatives and select a model that best fits the data. We use a Markov-Chain-Monte-Carlo (MCMC) method for the computation of the criterion that quantifies the relative merit of each model.

Denote by $Y_T = (y_1, \cdots, y_T)$, a sample of T observations for the process y_t. A family $\mathcal{M}$ consists of candidate models for Y_T in the presence of uncertainty of the true model. A model $m_i \in \mathcal{M}$ is associated with a parameter space Θ^i of dimension p_i for $i \in \mathcal{I}$ where $\mathcal{I} = \{1, ..., I\}$. Assume that for each m_i a family $Q^i_T(\theta^i, Y_T)$ of distribution functions, with a density $q^i_T(\theta^i, Y_T)$, is defined.

Let $\Pr(m_i|Y_T)$ be the post-data (posterior) probability that m_i is true. By the Bayes' rule, we have

$$\Pr(m_i|Y_T) = \frac{q_T(Y_T|m_i)\Pr(m_i)}{\sum_{j\in\mathcal{I}} q_T(Y_T|m_j)\Pr(m_j)} \tag{4.1}$$

where $\Pr(m_i)$ is the prior probability that m_i is true and $q_T(Y_T|m_j) = q^i_T(Y_T)$. But, notice that

$$q_T(Y_T|m_j) = \int q_T(Y_T|\theta^j, m_j)\varphi(\theta^j|m_j)d\theta^i = E_j[q_T(Y_T|\theta^j)], \tag{4.2}$$

where $\varphi(\theta^j|m_j)$ is the prior density associated with the model m_j. If we further assume that $\Pr(m_j)$ is the same for all j, the model selection rule is to choose m_i for which $E_i[q_T(Y_T|\theta^i)]$ is the largest.

The quantity in (4.2) can be alternatively interpreted as the marginal likelihood. The marginal likelihood $q_T(Y_T|m_j) = E_i[q_T(Y_T|\theta^i)]$ can be rewritten as

$$q_T(Y_T) = \frac{q_T(Y_T|\theta)\varphi(\theta)}{\varphi(\theta|Y_T)} \tag{4.3}$$

where the script j and m_j are omitted for convenience, and $\varphi(\theta|Y_T)$ is the posterior density of θ. The equation (4.3) is a reversed version of the Bayes' rule. Since (4.3) holds for any θ, we may evaluate $q_T(Y_T)$ for a convenient θ, say $\theta = \theta^*$ the posterior mean. Taking the logarithm of (4.3) for $\theta = \theta^*$, we have

$$\ln q_n(Y_T) = \ln q_T(Y_T|\theta^*) + \ln \varphi(\theta^*) - \ln \varphi(\theta^*|Y_T) \tag{4.4}$$

Our decision rule is to choose the model m_j that yields the highest value of (4.4). Calculation of the log-likelihood and the log-prior at $\theta = \theta^*$ is relatively easy. However, calculation of the posterior $\varphi(\theta^*|Y_T)$ is not easy. To compute $\varphi(\theta^*|Y_T)$ we can use an MCMC method such as Gibbs sampler as in Chib (1995).[1] Computation of the marginal likelihood or the posterior $\varphi(\theta^*|Y_T)$ for the models in Section 3 is a very demanding work with highly sophisticated programming.[2]

We use standard priors for the parameters that are used in the literature. That is, the regression coefficients ϕ and ζ have normal priors. The error variance σ^2 has an inverted gamma distribution. This normal-inverted gamma prior distribution is a conjugate prior and is standard one used in the literature. The mean μ_i has a normal prior, and the transition probability $\{p_{ij}\}$ has a prior of Dirichlet distribution, which are also conjugate priors. The prior for the transition probability reflects information on duration of a state. For example, if it is reasonable to think that duration of the state i is four quarters, then $1/(1-p_{ii}) = 4$, so that $p_{ii} = 0.75$ for the prior.

4.2 Data and Empirical Results

As in many existing works we use the 3-month Treasury bill rate or equivalence for the nominal interest rate and the consumer price index (CPI) for the price level to compute the inflation rate. We get the U.S. data of the T-bill rate and the CPI from the Federal Reserve Board and the Bureau of Labor Statistics, respectively. The Japan and Korean data are from the International Financial Statistics (IFS). All the data are seasonally adjusted. The data period is 1957:Q1-2006:Q1 for Japan, 1953:Q1-2006:Q1 for the U.S. and 1976:Q3-2006:Q1 for Korea.

We consider most possible aspects of each model for the model selection: the number of states K up to 5 and the order of the autoregression up to 5 for each M_i, $i = 0, 1, 2, 3$. Table 1 shows results of model selection by the method of post-data model selection.

As shown in the Tables 1 the best model selected is different for the three countries: M_1 is the best model for the U.S. and M_3 for Japan and Korea. However, we can see that for data from the three countries the models with both regimes/periods, a regime with stationary fluctuations and the other with nonstationary/unstable fluctuations, (M_1 and M_3) are selected as the most appropriate models. This result does not apparently confirm

[1]Good references for the MCMC method and the Gibbs sampler are Gelman, Carlin, Stern and Rubin (2000), Chib and Greenberg (1996), and Casella and George (1992), among others.

[2]A detailed explanation of the MCMC used for our models is provided in Appendix A.

Table 1

The result of model selection

	Selected model	Number of states	lag order
Japan	M_3	5	4
Korea	M_3	4	4
U.S.	M_1	-	3

the existence of Fisher relation since the Fisher relation implies that the real interest rate is stable with stationary fluctuations around a constant. However, we can say that the Fisher relation prevails if the nonstationary fluctuations occur only temporarily. Kim and Park (2016) have shown that the length of the period of nonstationary fluctuations in the real interest rate is relatively short for a data set similar to that used in this paper.

5 Concluding Remarks

We have compared several alternative models of the Fisher relation for data from Japan, Korea and the U.S based on the Bayesian model determination method. Among four alternative models for the Fisher relation our result shows that models with both regimes/periods of nonstationarity and stationarity fit data best although the best model is not the same for the three countries. It is a new and interesting result about the Fisher relation and would motivate further investigation of related issues.

References

[1] Andrews, D.W.K. and J.Y. Kim (2006), "Tests for Cointegration Breakdown over a Short Time Period," *Journal of Business & Economic Statistics*

[2] Atkins, F. (1989), "Co-integration, Error Correction and the Fisher Effect," *Applied Economics* 21, 1611–1620.

[3] Casella, G. and George, E. I. (1992), "Explaining the Gibbs Sampler," *American Statistician* 46, 167–174.

[4] Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of American Statistical Association* 90, 1313–1321.

[5] Chib, S. and Greenberg, E. (1996), "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory* 12, 409–431.

[6] Evans, M. D. D. and Lewis, K. K. (1995), "Do Expected Shifts in Inflation Affect Estimates of the Long-Run Fisher Relation?," *Journal of Finance* 50.

[7] Fama, E. F. (1975), "Short-term Interest Rates as Predictors of Inflation," *American Economic Review* 65, 269–282.

[8] Fama, E. F. and Gibbons, M. R. (1982), "Inflation, Real Returns, and Capital Investment," *Journal of Monetary Economics* 9, 297–323.

[9] Garbade, K. and Wachtel, P. (1978), "Time Variation in the Relationship between Inflation and Interest Rates," *Journal of Monetary Economics* 4, 755–765.

[10] Garcia, R. and Perron, P. (1996), "An Analysis of the Real Interest Rate under Regime Shifts," *Review of Economics and Statistics* 78, 111–125.

[11] Gelman, A., Carlin, C. B., Stern, H. S. and Rubin, D. B. (2000), *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, FL.

[12] Huizinga, J. and Mishkin, F. S. (1986), "Monetary Policy Regime Shifts and the Unusual Behavior of Real Interest Rates," *Carnegie-Rochester Conference Series on Public Policy* 24, 231–274.

[13] Kim, J.Y. (2003), "Inference on Segmented Cointegration," *Econometric Theory* 19, 620–639.

[14] Kim, J.Y. and W.Y. Park (2016), "New Empirical Evidence for the Fisher Relation: Integration and Short-run Instability," *Seoul Journal of Economics.*

[15] Mishkin, F. S. (1981), "The Real Interest Rate: An Empirical Investigation," *Carnegie-Rochester Conference Series on Public Policy* 15, 151–200.

[16] Mishkin, F. S. (1984), "The Real Interest Rate: A Multi-country Empirical Study," *Canadian Journal of Economics* 17, 283–311.

[17] Mishkin, F. S. (1990), "What does the Term Structure of Interest Rates Tell Us about Future Inflation?," *Journal of Monetary Economics* 25, 77–95.

[18] Mishkin, F. S. (1992), "Is the Fisher Effect for Real?," *Journal of Monetary Economics* 30, 195–215.

[19] Nelson, C. R. and Schwert, G. W. (1977), "Short-term Interest Rates as Predictors of Inflation: On Testing the Hypothesis that the Real Rate of Interest Is Constant," *American Economic Review* 67, 478–486.

[20] Roley, V. V. (1986), "The Response of Interest Rates to Money Announcements under Alternative Operating Prosedures and Reserve Requirement Systems". NBER Working Paper No.1812.

[21] Rose, A. K. (1988), "Is the Real Interest Rate Stable?," *Journal of Finance* 43, 1095–1112.

[22] Seo, M. H. (2008), "Unit root test in a threshold autoregression: asymptotic theory and residual-based block bootstrap." *Econometric Theory*, 24, 6, 1699-1716.

[23] Summers, L. H. (1982), "The Non-Adjustment of Nominal Interest Rates: A Study of the Fisher Effect," *NBER Working Paper No. 836* .

[24] Tong, H. and K. S. Lim (1980), "Threshold autoregression, limit cycles and cyclical data." *Journal of the Royal Statistical Society, Series B (Methodological)*, 245-292.

[25] Wallace, M. S. and Warner, J. T. (1993), "The Fisher Effect and the Term Structure of Interest Rates: Tests of Cointegration," *Review of Economics and Statistics* 75, 320–324.

Appendix A: Inference on Posterior

1 An Autoregression: M_0

Assume that the initial p observations $\mathbf{y}_0 = \{y_{-p+1}, \cdots, y_0\}$ are given. Let $\mathbf{y}_T = \{y_1, \cdots, y_T\}$. Rewrite the given linear autoregression

$$(y_t - \mu) = \sum_{s=1}^{p} \phi_s (y_{t-s} - \mu) + \varepsilon_t,$$

as in the following:

$$y_t = \phi_0 + \sum_{s=1}^{p} \phi_s y_{t-s} + \varepsilon_t,$$

where $\phi_0 = (1 - \sum_{s=1}^{p} \phi_s)\mu$, $\varepsilon_t \sim N(0, \sigma^2)$. Writing it in a matrix form, we have

$$\mathbf{y}_t = \mathbf{X}_T \phi + \varepsilon$$

where $\mathbf{X}_T = \{\mathbf{x_1}, \cdots, \mathbf{x_T}\}'$ for $\mathbf{x_t} = \{1, y_{t-1}, \cdots, y_{t-p}\}'$, and $\phi = \{\phi_\mathbf{0}, \phi_\mathbf{1}, \cdots, \phi_\mathbf{p}\}$. Also, $\varepsilon = \{\varepsilon_1, \cdots, \varepsilon_T\}'$.

The posterior of this model is well known. We divide prior of ϕ and σ^2 as in the following

$$p(\phi, \sigma^2) = p(\phi|\sigma^2) p(\sigma^2) \tag{B.1}$$

Assume that, given σ^2 prior of ϕ is normal $N(\underline{\phi}, \sigma^2 \underline{H}^{-1})$ and prior of σ^2 is inverted-gamma $IG(\underline{\nu}/2, \underline{s}^2/2)$, that is, $1/\sigma^2$ has a gamma distribution. This prior is a conjugate prior whose family of distributions is the same as that of its posterior. Now, the prior (B.1) can be written as

$$\begin{aligned}
p(\phi, \sigma^2) = {} & p(\phi|\sigma^2) p(\sigma^2) \\
& = \left(\frac{1}{\sqrt{2\pi}}\right)^p \left(\frac{1}{\sigma^2}^{k/2}\right) |\underline{H}|^{1/2} exp\left[-\frac{1}{2\sigma^2}(\phi - \underline{\phi})' \underline{H} (\phi - \underline{\phi})\right] \\
& \times \frac{1}{\Gamma(\underline{\nu}/2)} \left(\frac{\underline{s}^2}{2}\right)^{\underline{\nu}/2} exp\left(-\frac{s^2}{2\sigma^2}\right) \\
& \propto \left(\frac{1}{\sigma^2}\right)^{(\underline{\nu}+k)/2+1} exp\left[-\frac{1}{2\sigma^2}[\underline{s}^2 + (\phi - \underline{\phi}' \underline{H}(\phi - \underline{\phi})]\right],
\end{aligned}$$

where $k = dim(\phi) = p+1$. The probability density function of $\mathbf{y}_T$, that is, the likelihood of ϕ and σ^2 is

$$\begin{aligned} l(\phi, \sigma^2|\mathbf{y}_T, \mathbf{y}_0) &= \left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\frac{1}{\sigma^2}\right)^{T/2} exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}_T - \mathbf{X}_T\phi)'(\mathbf{y}_T - \mathbf{X}_T\phi)\right] \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\frac{1}{\sigma^2}\right)^{T/2} \\ &\quad \times exp\left[-\frac{1}{2\sigma^2}[s^2 + (\phi - \hat{\phi})'\mathbf{X}_T'\mathbf{X}_T(\phi - \hat{\phi})\right] \end{aligned}$$

where $\hat{\phi} = (\mathbf{X}_T'\mathbf{X}_T)^{-1}\mathbf{X}_T'\mathbf{y}_T$, and $s^2 = (\mathbf{y}_T - \mathbf{X}_T\phi)'(\mathbf{y}_T - \mathbf{X}_T\phi)$. The posterior density of ϕ and σ^2 can be obtained as in the following:

$$\begin{aligned} &p(\phi, \sigma^2|\mathbf{y}_T, \mathbf{y}_0) \\ &\quad \propto l(\phi, \sigma^2|\mathbf{y}_T, \mathbf{y}_0)p(\phi, \sigma^2) \\ &\quad \propto \left(\frac{1}{\sigma^2}\right)^{(\underline{\nu}+k)/2+1} \\ &\quad \times exp\left[-\frac{1}{2\sigma^2}[\underline{s}^2 + s^2 + (\phi - \underline{\phi})'\underline{H}(\phi - \underline{\phi}) + (\phi - \hat{\phi})'\mathbf{X}_T'\mathbf{X}_T(\phi - \hat{\phi})]\right] \\ &\quad = \left(\frac{1}{\sigma^2}\right)^{(\underline{\nu}+k)/2+1} exp\left[-\frac{1}{2\sigma^2}[s^{*2} + (\phi - \phi^*)'[\underline{H} + \mathbf{X}_T'\mathbf{X}_T](\phi - \phi^*)\right] \end{aligned} \tag{B.2}$$

where

$$s^{*2} = \underline{s}^2 + s^2 + (\underline{\phi} - \hat{\phi})'[\underline{H}^{-1} + (\mathbf{X}_T'\mathbf{X}_T)^{-1}]^{-1}(\underline{\phi} - \hat{\phi})]$$

and

$$\phi^* = [\underline{H} + \mathbf{X}_T'\mathbf{X}_T]^{-1}[\underline{H}\underline{\phi} + (\mathbf{X}_T'\mathbf{X}_T)\hat{\phi}].$$

Expansion from the third row to the fourth row of (B.2) is possible by the following fact:

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[DA^{-1}B + C^{-1}]^{-1}DA^{-1}.$$

From (B.2) we know that the posterior of $\phi|\sigma^2$ and that of σ^2 are, respectively,

$$\phi|\sigma^2, \mathbf{y}_T, \mathbf{y}_0 \sim N(\phi^*, \sigma^2[\underline{H} + \mathbf{X}_T'\mathbf{X}_T]^{-1})$$

$$\sigma^2|\mathbf{y}_T, \mathbf{y}_0 \sim IG(\nu^*/2, s^{*2}/2)$$

where $\nu^* = \underline{\nu} + T$. We need posterior of μ and not that of ϕ_0. We rewrite the mean and variance of the conditional posterior of ϕ, ϕ^* and $\sigma^2[\underline{H} + \mathbf{X}_T'\mathbf{X}_T]^{-1}$, respectively, as

$$\phi^* = (\phi_0^*; \phi_{-0}^*)'$$

and

$$\sigma^2[\underline{H} + \mathbf{X}_T'\mathbf{X}_T]^{-1} = \begin{pmatrix} h_{11} & H_{12} \\ H_{12}' & H_{22} \end{pmatrix}$$

where $\phi_{-0} = (\phi_1, \cdots, \phi_p)'$, $h_{11} = Var(\phi_0)$, and $H_{12} = Cov(\phi_0, \phi_{-\mathbf{0}})$, $H_{22} = Var(\phi_0)$. From properties of multivariate normal distribution, we have (Poirier 1995)

$$\phi_0|\phi_{-0}, \sigma^2, \mathbf{y}_T, \mathbf{y}_0 \sim N({\phi_0}^* + H_{12}H_{22}^{-1}(\phi_{-0}^* - \phi_{-0}), h_{11} - H_{12}H_{22}^{-1}H_{12}').$$

Therefore, we get posterior of $\mu|\phi_{-0}, \sigma^2$ as

$$\mu|\phi_{-0}, \sigma^2, \mathbf{y}_T, \mathbf{y}_0 \sim N\left(\frac{{\phi_0}^* + H_{12}H_{22}^{-1}(\phi_{-0}^* - \phi_{-0})}{1 - \sum_{s=1}^p \phi_s}, \frac{h_{11} - H_{12}H_{22}^{-1}H_{12}'}{(1 - \sum_{s=1}^p \phi_s)^2}\right).$$

On the other hand, it is easy to show that posterior of $\phi_{-0}|\sigma^2$ is

$$\phi_{-0}|\sigma^2, \mathbf{y}_T, \mathbf{y}_0 \sim N(\phi_{-0}^*, H_22).$$

2 A Model with Partial-sample Instability: M_1

Posterior of model M_1 can be directly obtained from the extended Markov switching regime change model M_3 by assuming that the state variable is known *a priori*.

3 Markov Regime Switching Model: M_2

We explain how to simulate the posterior distribution of the Markov switching regime change model based on the Gibbs sampling.

To get the posterior of the Markov switching regime change model M_2 by the Gibbs sampler we have together the parameters of the model, vector of the state variables $\{s_1, \cdots, s_T\}$, mean $\mu_1, \cdots, \mu_K$, variance ${\sigma_1}^2, \cdots, {\sigma_K}^2$, autoregression coefficients $\phi_1, \cdots, \phi_p$, and $\phi = \{\phi_1, \cdots, \phi_p\}$. In Gibbs sampling the posterior of parameters in a part is generated from a conditional distribution of the other part.

1) Generation of $\boldsymbol{S}_T$: Given μ, σ^2, ϕ, P we can write the posterior of the state vector S_T as in the following:

$$\begin{aligned} p(S_T|\mathbf{y}_T) &= p(s_1, \cdots, s_T|\mathbf{y}_T) \\ &= p(s_T|\mathbf{y}_T) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, \cdots, s_T, \mathbf{y}_T) \\ &= p(s_T|\mathbf{y}_T) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, \mathbf{y}_T). \end{aligned} \tag{B.3}$$

In (B.3) derivation of the second line from the first line is based on a basic property of conditional probability, and the third line is obtained from the first order Markov properties of state variables. (B.3) shows that we generate s_T, assuming that $\mathbf{y}_T$ is given, and then generate s_t for $t = T-1, T-2, \cdots, 1$ successively conditioned on the pre-generated s_{t+1} and $\mathbf{y}_T$, by which S_T can be generated. The problem is how to get $p(s_t|s_{t+1}, \mathbf{y}_t)$, which can be easily available by applying the Hamilton(1989)'s filter. Assuming that the other parameters are given, we can get the probability of current state $p(s_t|\mathbf{y}_t)$ by Hamilton filter if information up tp t is available. Since, for $t = T-1, T-2, \cdots, 1$,

$$\begin{aligned} p(s_t|s_{t+1}, \mathbf{y}_t) &= \frac{p(s_t, s_{t+1}|\mathbf{y}_t)}{s_{t+1}|\mathbf{y}_t)} \\ &= \frac{p(s_{t+1}|s_t, \mathbf{y}_t)p(s_t|\mathbf{y}_t)}{p(s_{t+1}|\mathbf{y}_t)} \\ &= \frac{p(s_{t+1}|s_t)p(s_t|\mathbf{y}_t)}{p(s_{t+1}|\mathbf{y}_t)} \end{aligned}$$

from the assumed transition probability and probability obtained from the Hamilton's filter, we can get $p(s_t|s_{t+1}, y_t)$. From this result we get, for $t = T-1, T-2, \cdots, 1$,

$$P[s_t = i|s_{t+1}, y_t] = \frac{p(s_{t+1}|s_t = i)p(s_t = i|y_t)}{\sum_{j=1}^{M} p(s_{t+1}|s_t = i)p(s_t = i|y_t)}. \tag{B.4}$$

Given s_{t+1}, $p(s_{t+1}|\mathbf{y}_t)$ is determined independently of s_t, so that it is cancelled out in the numerator and denominator. Generation of the state variable from (B.4) can be done based on the uniform distribution. On the other hand, $p(s_T|\mathbf{y}_T)$ can be immediately obtained from the Hamilton filter.

2) Generation of $\boldsymbol{\mu}$: Given S_T, σ^2, ϕ, P modify M_2 as in the following:

$$\frac{\tilde{y_t}}{\nu_t} = \mu_1 \frac{\tilde{s_{1t}}}{\nu_t} + \cdots + \mu_M \frac{\tilde{s_{M,t}}}{\nu_t} + \frac{\varepsilon_t}{\nu_t} \tag{B.5}$$

Here $\tilde{y_t} = y_t - (\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p})$ and $\tilde{s}_{i,t} = s_{i,t} - (\phi_1 s_{i,t-1} + \cdots + \phi_p s_{i,t-p})$. Also, $s_{i,t} = I(s_t = i)$. Letting $u_t = \varepsilon_t/\nu_t$, we have that $u_t \sim N(0, 1)$. Now, rewriting (B.5),

$$\tilde{\mathbf{y}} = \tilde{\mathbf{s}}\boldsymbol{\mu} + \mathbf{u} \tag{B.6}$$

where $\tilde{\mathbf{y}} = \{\tilde{y}_1/\nu_t, \cdots, \tilde{y}_t/\nu_t\}'$, $\tilde{s}_i = \{\tilde{s}_{i,1}, \cdots, \tilde{s}_{i,T}\}'$, $i = 1, \cdots, M)$, and $\tilde{\mathbf{s}} = \{\tilde{s}_1, \cdots, \tilde{s}_M\}$. Also, $\mathbf{u} = \{u_1, \cdots, u_T\}'$ and $\mathbf{u} \sim N(\mathbf{0}_T, I_T)$. Then, the likelihood of (B.6) becomes

$$l(\boldsymbol{\mu}|\tilde{\mathbf{y}}, \mathbf{S_T}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \mathbf{P}) = \left(\frac{1}{\sqrt{2\pi}}\right)^T exp\left[-\frac{1}{2}(\tilde{\mathbf{y}} - \tilde{\mathbf{s}}\boldsymbol{\mu})'(\tilde{\mathbf{y}} - \tilde{\mathbf{s}}\boldsymbol{\mu})\right]. \tag{B.7}$$

Now, assume that the prior distribution of $\boldsymbol{\mu}$ is independent of the other parameters and is Gaussian with mean a_0 and variance ${m_0}^{-1}$:

$$\boldsymbol{\mu}|\mathbf{S_T}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \mathbf{P} \sim N(a_0, {m_0}^{-1})_{I(\mu_1 < \cdots < \mu_K)} \tag{B.8}$$

where $I(\mu_1 < \cdots < \mu_K)$ is a restriction for the identification of states. With this restriction the prior of $\boldsymbol{\mu}$ is a truncated normal distribution. From (B.7) and (B.8) we get a conditional posterior of $\boldsymbol{\mu}$:

$$\begin{aligned}
&p(\boldsymbol{\mu}|S_T, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{y}_T) \\
&\quad \propto l(\boldsymbol{\mu}|\tilde{\mathbf{y}}, S_T, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \mathbf{P})p(\boldsymbol{\mu}|S_T, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \mathbf{P}) \\
&\quad \propto |{m_0}^{-1}|^{-1/2} exp\left[-\frac{1}{2}[(\boldsymbol{\mu} - a_0)' m_0 (\boldsymbol{\mu} - a_0) + (\tilde{\mathbf{y}} - \tilde{\mathbf{s}}\boldsymbol{\mu})'(\tilde{\mathbf{y}} - \tilde{\mathbf{s}}\boldsymbol{\mu})]\right] \\
&\quad \propto |m^*|^{-1/2} exp\left[-\frac{1}{2}[(\boldsymbol{\mu} - \boldsymbol{\mu}^*)'(\boldsymbol{\mu} - \boldsymbol{\mu}^*)\right],
\end{aligned}$$

where $m^* = (m_0 + \tilde{\mathbf{s}}'\tilde{\mathbf{s}})^{-1}$, $\boldsymbol{\mu}^* = m * (m_0 a_0 + \tilde{\mathbf{s}}'\tilde{\mathbf{y}})$. Therefore, the conditional posterior of $\boldsymbol{\mu}$ is a normal distribution:

$$\boldsymbol{\mu}|S_T, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \mathbf{P}, \mathbf{y}_T \sim N(\tilde{\boldsymbol{\mu}}, m^*)_{I(\mu_1 < \cdots < \mu_K)}$$

where $I(\mu_1 < \cdots < \mu_K)$ is a restriction for the identification of states. After generation of $\boldsymbol{\mu}$ we take only those satisfying this restriction.

3) Generation of $\boldsymbol{\sigma^2}$: Given $S_T, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{P}$, we generate ${\sigma_1}^2, \cdots, {\sigma_K}^2$ individually one by one since they are independent of each other. First, for

$$\varepsilon_t = (y_t - \tau_t) - \sum_{p=1}^{L} \phi_p (y_{t-p} - \tau_{t-p})$$

let

$$\tilde{\boldsymbol{\varepsilon}}_i = \{\varepsilon_t | s_t = i\}'$$

for $i = 1, \cdots, K$. With ε_t having a normal distribution with mean 0 and variance $\sigma^2_{s_t}$, we have the likelihood of $\tilde{\boldsymbol{\varepsilon}}_i$ as

$$l(\sigma_i^2 | \tilde{\boldsymbol{\varepsilon}}_i, S_T, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{P}) \propto \left(\frac{1}{\sigma_i^2}\right)^{\frac{T_i}{2}} exp\left[-\frac{1}{\sigma_i^2} \tilde{\boldsymbol{\varepsilon}}_i{}' \tilde{\boldsymbol{\varepsilon}}_i\right]$$

where T_i is the sample size of the i^{th} states $(s_t = i)$. Assuming that the prior of σ_i^2 is an inverted Gamma, $IG(\nu_i/2, \delta_i/2)$,

$$p(\sigma_i^2 | S_T, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{P}) \propto \left(\frac{1}{\sigma_i^2}\right)^{\nu_i/2-1} exp\left(-\frac{\delta_i}{2\sigma_i^2}\right),$$

we know that a conditional posterior of σ_i^2 is

$$p(\sigma_i^2 | S_T, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{P}, \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}}) \propto \left(\frac{1}{\sigma_i^2}\right)^{T_i+\nu_i/2-1} exp\left(-\frac{\delta_i + \tilde{\boldsymbol{\varepsilon}}_i{}' \tilde{\boldsymbol{\varepsilon}}_i}{2\sigma_i^2}\right),$$

which is another inverted Gamma,

$$\sigma_i^2 | S_T, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{P}, \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{i}} \sim IG\left(\frac{T_i + \nu_i}{2}, \frac{\delta_i + \tilde{\boldsymbol{\varepsilon}}_i{}' \tilde{\boldsymbol{\varepsilon}}_i}{2}\right)$$

We can generated the posterior of σ_i^2 from the fact that the inverse of a Gamma has an inverted Gamma distribution.

4) Generation of $\boldsymbol{\phi}$: Given S_T and $\boldsymbol{\sigma^2}, \boldsymbol{\mu}, \boldsymbol{P}$, let $\omega_t = y_t - \tau_t$ and rewrite M_2 as follows:

$$\frac{\tilde{\omega}_t}{\nu_t} = \phi_1 \frac{\omega_{t-1}}{\nu_t} + \cdots + \phi_p \frac{\omega_{t-p}}{\nu_t} + \frac{\varepsilon_t}{\nu_t} \tag{B.9}$$

Letting $u_t = \varepsilon_t / \nu_t$, we have $u_t \sim N(0, 1)$. Now, write (B.9) as

$$\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{u} \tag{B.10}$$

where $\boldsymbol{z} = \{\omega_1/\nu_1, \cdots, \omega_T/\nu_T\}'$, and for $x_t = \{\omega_{t-1}/\nu_1, \cdots, \omega_{t-1}/\nu_t\}'$ $\boldsymbol{X} = \{x_1, \cdots, x_T\}'$. Also, for $\boldsymbol{u} = \{u_1, \cdots, u_T\}'$, $\boldsymbol{u} \sim N(\boldsymbol{0}_T, I_T)$. Then, the likelihood of (B.10) is

$$l(\boldsymbol{\phi}|\boldsymbol{z}, S_T, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{P}) = \left(\frac{1}{\sqrt{2\pi}}\right)^T exp\left[-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{X\phi})'(\boldsymbol{z} - \boldsymbol{X\phi})\right]. \tag{B.11}$$

Now, assume that the prior of $\boldsymbol{\phi}$ is independent of the other parameters and has a normal distribution with mean b_0 and variance B_0^{-1},

$$\boldsymbol{\phi}|S_T, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{P} \sim N(b_0, {B_0}^{-1})_{I(\boldsymbol{\phi} \in \Phi_S)}. \tag{B.12}$$

where $I(\boldsymbol{\phi} \in \Phi_S)$ is a stability condition for $\boldsymbol{\phi}$ where Φ_S is the space of ϕ such that roots of the characteristic equation $1 - \phi_1 z - \cdots - \phi_p z^p = 0$ lie outside the unit circle. From (B.11) and (B.12) we get a conditional posterior of ϕ:

$$\begin{aligned}
&p(\boldsymbol{\phi}|S_T, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{P}, \boldsymbol{z}) \\
&\quad \propto l(\boldsymbol{\phi}|\boldsymbol{z}, S_T, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{P}) p(\boldsymbol{\phi}|S_T, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{P}) \\
&\quad \propto |{B_0}^{-1}|^{-1/2} exp\left[-\frac{1}{2}[(\boldsymbol{\phi} - b_0)' B_0 (\boldsymbol{\phi} - b_0) + (\boldsymbol{z} - \boldsymbol{X\phi})'(\boldsymbol{z} - \boldsymbol{X\phi})]\right] \\
&\quad \propto |B^*|^{-1/2} exp\left[-\frac{1}{2}[(\boldsymbol{\phi} - \phi^*)' B_0 (\boldsymbol{\phi} - \phi^*)\right]
\end{aligned}$$

,

where $B^* = (B^0 + \boldsymbol{X'X})^{-1}$ and $\phi^* = B^*(B_0 b_0 + \boldsymbol{X'z})$. Therefore the conditional posterior of ϕ is a Gaussian as in the following:

$$\boldsymbol{\phi}|S_T, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{P}, \boldsymbol{z} \sim N(\hat{\boldsymbol{\phi}}, B^*)_{I(\phi \in \Phi)}$$

which is a truncated normal. As for μ we take only the generated values satisfying these restrictions.

5) Generation of $\boldsymbol{P}$: Given $S_T, \boldsymbol{\sigma^2}, \boldsymbol{\mu}, \boldsymbol{\phi}$, S_T being given conditionally, we can count the number of transitions between states in a sample. Thus, let n_{ij} be the number of transitions from state i to state j. Then, $\sum_{i=1}^K \sum_{j=1}^K n_{ij} = T$. Since for $i = 1, \cdots, K$ the number of transitions to states $j = 1, \cdots, K$ follows a multinomial distribution, the likelihood of p_{ij} $(j = 1, \cdots, K)$ is

$$l(p_{i1}, \cdots, p_{iK}|\boldsymbol{y}_T) = \left(\prod_{j=1}^{K-1} p_{ij}^{n_{ij}}\right)\left(1 - \sum_{j=1}^{K-1} p_{ij}\right)^{n_{iK}}.$$

Since $p_{iK} = 1 - (p_{i1} + p_{i(K-1)})$ we do not consider p_{iM}. For the prior of transition probability we use a Dirichlet distribution. The pdf of p_{ij} having a Dirichlet distribution with parameters $\alpha_{i1}, \cdots, \alpha_{i(K-1)}$ is as follows

$$p(p_{i1}, \cdots, p_{i(K-1)}|S_T) = \left(\prod_{j=1}^{K-1} p_{ij}^{\alpha_{ij}-1}\right)\left(1 - \sum_{j=1}^{K-1} p_{ij}\right)^{\alpha_{iK}-1}.$$

Therefore, a conditional posterior of p_{ij} is

$$p(p_{i1}, \cdots, p_{i(K-1)}|S_T, \boldsymbol{y}_T) = \left(\prod_{j=1}^{K-1} p_{ij}^{n_{ij}+\alpha_{ij}-1}\right)\left(1 - \sum_{j=1}^{K-1} p_{ij}\right)^{n_{iK}+\alpha_{iK}-1}.$$

That is, p_{ij} has a Dirichlet distribution with parameters $n_{i1} + \alpha_{i1}, \cdots, n_{i(K-1)} + \alpha_{i(K-1)}$:

$$p_{i1}, \cdots, p_{i(K-1)}|S_T, \boldsymbol{y}_T \sim \mathit{Dirichlet}(n_{i1} + \alpha_{i1}, \cdots, n_{i(K-1)} + \alpha_{i(K-1)}).$$

We generate Dirichlet random numbers by a Gamma distribution. That is, with x_{ij} generated from $\mathit{Gamma}(\alpha_{ij}, 1)$ we have p_{ij} as $x_{ij}/\sum x_{ij}$.

4 An Extended Markov Switching Model: M_3

Inclusion of instability state(s) in M_2 does not substantially affect estimation of posterior distribution by Gibbs sampler. However, given S_T, $\boldsymbol{\sigma^2}, \boldsymbol{\phi}, \boldsymbol{P}$ we need to modify the part of generation of $\boldsymbol{\mu}$. Since, in M_3, $\tau_t = \mu_1 s_{1,t} + \cdots + \mu_{K-1}s_{K-1,t} + (\mu_K + y_{t-1})s_{K,t}$, we can rewrite M_3 as in the following with $s_{i,t} = I(s_t = i)$:

$$\begin{aligned} & y_t - y_{t-1}s_{K,t} - \sum_{s=1}^{p} \psi_{s,t}(y_{t-s} - y_{t-s-1}s_{K,t-s}) \\ &\quad = \mu_1[s_{1,t} - \sum_{s=1}^{p} \psi_{s,t}s_{1,t-s}] + \cdots + \mu_K[s_{K,t} - \sum_{s=1}^{p} \psi_{s,t}s_{K,t-s}] + \varepsilon_t \end{aligned}$$

where $\psi_{s,t} = [\phi_s \cdot I(s_t \neq K) + \zeta_s \cdot I(s_t = K)]$. Thus, having $\tilde{y}_t = y_t - y_{t-1}s_{K,t} - \sum_{s=1}^{p} \psi_{s,t}(y_{t-s} - y_{t-s-1}s_{K,t-s})$ in (B.5) we can generate $\boldsymbol{\mu}$ by the same method. In this case the identifying restriction for the distribution of $\boldsymbol{\mu}$ becomes $I(\mu_1 < \cdots < \mu_{K-1})$. We can generate all the other parameters except τ_t by the same method as in M_2.

Appendix B: Estimation of Marginal Likelihood

Estimation of marginal likelihood can be done by Chib(1995)'s algorithm. This algorithm calculates a posterior density based on the fact that Gibbs sampler generates parameters of a posterior. For Gibbs sampling we need to know the exact form of a posterior density.

Suppose that we consider computation of the following posterior density at $\boldsymbol{\theta^*} = (\boldsymbol{\theta_1^*}, \cdots, \boldsymbol{\theta_B^*})$:

$$p(\boldsymbol{\theta}^*|\boldsymbol{y}_T) = p(\boldsymbol{\theta_1^*}|\boldsymbol{y}_T)p(\boldsymbol{\theta_2^*}|\boldsymbol{y}_T, \boldsymbol{\theta_1^*}) \times \cdots \times p(\boldsymbol{\theta_B^*}|\boldsymbol{y}_T, \boldsymbol{\theta_1^*}, \cdots, \boldsymbol{\theta_{B-1}^*}) \tag{C.1}$$

where $\boldsymbol{\theta}^*(r = 1, \cdots, B)$ can be a vector. Suppose that there exist unobserved state vector $S_T = \{s_1, \cdots, s_T\}$ in this model. The first term can be computed by conditional densities of the other parameters

$$p(\boldsymbol{\theta}_1^*|\boldsymbol{y}_T) = \int p(\boldsymbol{\theta}_1^*|\boldsymbol{y}_T, \boldsymbol{\theta}_T, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_B, S_T)dp(\boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_B, S_T|\boldsymbol{y}) \tag{C.2}$$

where the integral is taken with respect to parameters other than $\boldsymbol{\theta}_1$. The set $\{(\boldsymbol{\theta}_2^{(g)}, \cdots, \boldsymbol{\theta}_B^{(g)}, S_T^{(g)})|g = 1, \cdots, G\}$ is obtained as a generated sample from G times Gibbs sampling of the joint posterior of $(\boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_B, S_T)$, $p(\boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_B, S_T|\boldsymbol{y})$. Therefore, the integral (C.2) is approximately computed as

$$\hat{p}(\boldsymbol{\theta}_1^*|\boldsymbol{y}_T) = G^{-1}\sum_{g=1}^{G} p(\boldsymbol{\theta}_1^*|\boldsymbol{y}_T, \boldsymbol{\theta}_2^{(g)}, \cdots, \boldsymbol{\theta}_B^{(g)}, S_T^{(g)})).$$

Now for $r = 2, \cdots, B$ we calculate consecutively $p(\boldsymbol{\theta}_r^*|\boldsymbol{y}_T, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*)$,

$$\begin{aligned} &p(\boldsymbol{\theta}_r^*|\boldsymbol{y}_T, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*) \\ &\quad = \int p(\boldsymbol{\theta}_r^*|\boldsymbol{y}_T, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*, \boldsymbol{\theta}_{r+1}, \cdots, \boldsymbol{\theta}_B, S_T)dp(\boldsymbol{\theta}_{r+1}, \cdots, \boldsymbol{\theta}_B, S_T|\boldsymbol{y}), \end{aligned}$$

so that we can get

$$p(\boldsymbol{\theta}_r^*|\boldsymbol{y}_T, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*) = G^{-1}\sum_{g=1}^{G} p(\boldsymbol{\theta}_r^*|\boldsymbol{y}_T, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*, \boldsymbol{\theta}_{r+1}^{(g)}, \cdots, \boldsymbol{\theta}_B^{(g)}, S_T^{(g)}).$$

Here, $(\boldsymbol{\theta}_{r+1}^{(g)}, \cdots, \boldsymbol{\theta}_B^{(g)}, S_T^{(g)})$ can be obtained from the previously applied Gibbs sampler by fixing $(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_{r-1}) = (\boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*)$ and by executing Gibbs sampling again. That is, we perform Gibbs sampling consecutively for a conditional posterior as in the following

$$
\begin{aligned}
&\boldsymbol{\theta}_r^{(g)} \leftarrow p(\boldsymbol{\theta}_r|\boldsymbol{y}, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*, \boldsymbol{\theta}_r^{(g-1)}, \cdots, \boldsymbol{\theta}_B^{(g-1)}, S_T^{(g-1)}) \\
&\quad \boldsymbol{\theta}_{r+1}^{(g)} \leftarrow p(\boldsymbol{\theta}_{r+1}|\boldsymbol{y}, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*, \boldsymbol{\theta}_r^{(g)}, \boldsymbol{\theta}_{r+1}^{(g-1)}, \cdots, \boldsymbol{\theta}_B^{(g-1)}, S_T^{(g-1)}) \\
&\quad \vdots \\
&\quad \boldsymbol{\theta}_B^{(g)} \leftarrow p(\boldsymbol{\theta}_B|\boldsymbol{y}, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*, \boldsymbol{\theta}_r^{(g)}, \cdots, \boldsymbol{\theta}_{B-1}^{(g)}, S_T^{(g-1)}) \\
&\quad S_T^{(g)} \leftarrow p(S_T|\boldsymbol{y}, \boldsymbol{\theta}_1^*, \cdots, \boldsymbol{\theta}_{r-1}^*, \boldsymbol{\theta}_r^{(g)}, \cdots, \boldsymbol{\theta}_B^{(g)}).
\end{aligned}
$$

The value of $\boldsymbol{\theta}_r$ generated along with values of the other parameters in Gibbs sampling is used for posterior instead of the fixed value $\boldsymbol{\theta}_r^*$. After computing posterior of (C.1) we get the marginal likelihood

$$\ln \hat{p}(\boldsymbol{y}_T) = \ln l(\boldsymbol{y}_T|\boldsymbol{\theta}) + \ln p(\boldsymbol{\theta}) - \ln \hat{p}(\boldsymbol{\theta}|\boldsymbol{y}_T).$$

We can easily calculate $l(\boldsymbol{y}_T|\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$. Numerical standard deviations of a posterior density and a marginal likelihood can be calculated as is explained in Chib(1995).

For the model of Markov switching regime change we can partition parameters as $\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \boldsymbol{P}$ and apply the above algorithm.