HIAS-E-148

Bayesian Analysis of Business Cycles in Japan by Extending the Markov Switching Model

Toshiaki Watanabe

Hitotsubashi University,

August 24, 2025

Hitotsubashi Institute for Advanced Study, Hitotsubashi University
2-1, Naka, Kunitachi, Tokyo 186-8601, Japan
tel:+81 42 580 8668   http://hias.hit-u.ac.jp/

Bayesian Analysis of Business Cycles in Japan by Extending the Markov Switching Model

Toshiaki Watanabe*†

Abstract

This paper analyzes business cycles in Japan by applying Markov switching (MS) models to monthly data on the coincident indicator of composite index (CI) during the period of 1985/01–2025/05 calculated by Economic and Social Research Institute (ESRI), Cabinet Office, the Government of Japan. During the latter half of the sample period, the Japanese economy experienced major shocks such as the global financial crisis in 2008, the Great East Japan Earthquake in 2011 and the COVID-19 pandemic in 2020. CI fell sharply during these periods, which make it difficult to estimate business cycle turning points using the simple MS model. In this paper, the MS model is extended by incorporating Student's t-error and stochastic volatility (SV). Since it is difficult to evaluate the likelihood once SV is introduced, a Bayesian method via Markov chain Monte Carlo (MCMC) is employed. The MS model with t-error or SV is shown to provide the estimates of the business cycle turning points close to those published by ESRI. A new method for evaluating marginal likelihood is evaluated. Bayesian model comparison based on marginal likelihood provides evidence that t-error is not needed once SV is introduced. Using the MS model with normal error and SV, structural changes in CI's mean growth rates during booms and recessions are also analyezed and two break points are found in the both mean growth rates. One is 2008/10 and the other is 2010/02, during which the mean growth rate during recession falls and that during boom rises due to the global financial crisis.

JEL classification: C11, C22, C51, C52, E32.

*Graduate School of Social Data Science, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601 JAPAN. E-mail: t.watanabe@r.hit-u.ac.jp

†This paper was presented at the JER-HIAS-Kakenhi Conference "Macroeconomics and Japan's Reality: Effects of Monetary and Fiscal Policies." I would like to thank Etsuro Shioji and Jouchi Nakajima for organizing this conference. I would also like to thank thank the chair, Yasuhiro Omori, discussant, Joichi Nakajima, and the participants in this conference for their invaluable comments and suggestions. Financial support from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government through Grant-in-Aid for Scientific Research (23H00048, 24H00142) and Hitotsubashi Institute of Adavanced Study (HIAS) is gratefully acknowledged. All remaining errors are solely my own responsibility.

1 Introduction

The Markov switching (MS) model proposed by Hamilton (1989) has frequently been applied to the empirical analysis of business cycles (Kim and Nelson (1998, 1999a,b), Watanabe (2014) and Ishihara and Watanabe (2015)). This model produces the posterior probability of boom and recession for each period, which can be used for dating the business cycle turning points.

This paper analyzes business cycles in Japan by applying Markov switching (MS) models to monthly data on the coincident indicator of composite index (CI) during the period of 1985/01–2025/05 calculated by Economic and Social Research Institute (ESRI), Cabinet Office, the Government of Japan. During the latter half of the sample period, the Japanese economy experienced major shocks such as the global financial crisis in 2008, the Great East Japan Earthquake in 2011 and the COVID-19 pandemic in 2020. Figures 1 and 2 plot CI and the growth rate of CI calculated as the percentage log difference of CIs in two consecutive months where the shadow areas are the recession periods published by ESRI. As shown in these figures, CI fell sharply during these periods. The introduction of a consumption tax of 3% in 1989/04, the increase to 5% in 1997/04, the increase to 8% in 2014/04, and the increase to 10% in 2019/10 also caused CI to fall.[1]

[Insert Figure 1]

[Insert Figure 2]

In this paper, we first show that these large shocks make it difficult to estimate business cycle turning points using the simple MS model. The simple MS model estimates only the periods of the global financial crisis in 2008, the Great East Japan Earthquake in 2011 and the COVID-19 pandemic in 2020 as well as the consumption tax hike in 2019 as recessions and estimates all other periods as booms. Therefore, in order to capture these large-scale shocks using the error term rather than switching the value of the average growth rate of the CI, we make the distribution of the error term a heavy-tailed distribution and the variance of the error term time-varying. We use the Student's t-distribution for a heavy-tailed distribution and the stochastic volatility (SV) model for time-varying variance. We show that the extended models estimate the business cycle turning points close to those published by the ESRI.

Once SV model is introduced, it is difficult to evaluate the likelihood. We employ a Bayesian estimation where the parameters and latent variables are sampled from the posterior distribution using Markov Chain Monte Carlo (MCMC) techniques and the obtained draws are used for estimating the parameters. We use the multimove sampler proposed by Kim and Nelson (1998, 1999a,b) to sample the state variable representing the boom or recession in the MS model and the block sampler proposed by Watanabe and Omori (2004) to sample the latent volatility in the SV model. We also use the method proposed by Watanabe (2001) to sample the degree-of-freedom of the Student's t-distribution.

Watanabe (2014) and Ishihara and Watanabe (2015) perform similar analyses using the CI for 1985/01–2013/11 and 1985/01–2014/05, respectively. This paper extends the sample period to 1985/01–2025/05, and the extended period includes the period of the COVID-19 pandemic. The difference between this paper and Watanabe (2014) and Ishihara and Watanabe (2015) is

[1] When the consumption tax was introduced in 1989/04, CI rose in the previous month due to last-minute demand.

not only the sample period. There is a technical problem in Watanabe (2014) and Ishihara and Watanabe (2015). They conduct model comparison based on the marginal likelihood calculated using the modified harmonic mean method proposed by Geweke (1999) with the complete-data likelihood, i.e., the joint density of the data and latent variables given the parameters. Chan and Grant (2015) provide evidence that the modified harmonic mean method with the complete-data likelihood has a substantial bias and tends to select the wrong model in practice. For MS models without SV, it is straightforward to evaluate the observed-data likelihood, i.e., the density of the data without the latent variables, using the Hamilton (1989) filter, so that it suffices to use the modified harmonic mean method with the obseved-data likelihood instead of the complete-data likelihood. For MS models with SV, it is not straightforward to evaluate the observed likelihood, so that we employ the methods proposed by Chib (1995) and Chib and Jeliazkov (2001). These methods also require the evaluation of the observed-data-likelihood, but whereas the modified harmonic mean method requires multiple evaluations, these methods require only one evaluation. Shibata and Watanabe (2005) propose a method for evaluating the observed-data likelihood of the Markov switching stochastic volatility (MSSV) model. Modifying this method, we develop a new method for evaluating the observed-data likelihood of MS models with SV. Model comparison based on the marginal likelihood calculated in this way provides evidence that t-error is not needed once the SV is introduced.

Using the MS model with normal error and SV selected based on the marginal likelihood, structural changes in CI's mean growth rates during booms and recessions are also analyezed and two break points are found in the both mean growth rates. One is 2008/10 and the other is 2010/02, during which the mean growth rate during recession falls and that during boom rises due to the global financial crisis.

The rest of this article is organized as follows. Next section reviews the simple MS model and extends it by incorporating with t-error and SV. Section 3 explains the Bayesian method using MCMC for the analysis of MS models. The simple and extended MS models are fitted to the CI in Japan in Section 4. Section 5 concludes.

2 Markov Switching Models

Let y_t denote the growth rate of CI and S_t denote a dummy variable that takes 0 when the economy is in the recession regime and 1 when the economy is in the boom regime. We assume that the mean of y_t, denoted by μ_t, is $\mu^{(0)}$ in the recession regime ($S_t = 0$) and $\mu^{(1)}$ in the boom regime ($S_t = 1$). Then, the MS model is represented as follows.

$$y_t = \mu_t + \phi(y_{t-1} - \mu_{t-1}) + e_t, \tag{1}$$

$$\mu_t = \mu^{(0)}(1 - S_t) + \mu^{(1)} S_t, \quad \mu^{(0)} < \mu^{(1)}, \tag{2}$$

$$S_t = \begin{cases} 1 & \text{boom}, \\ 0 & \text{recession}, \end{cases} \tag{3}$$

where S_t is assumed to follow a Markov process with transition probabilities:

$$\begin{aligned} \pi(S_t = 1 \mid S_{t-1} = 1) = p_{11}, \quad \pi(S_t = 0 \mid S_{t-1} = 1) = 1 - p_{11}, \\ \pi(S_t = 0 \mid S_{t-1} = 0) = p_{00}, \quad \pi(S_t = 1 \mid S_{t-1} = 0) = 1 - p_{00}. \end{aligned} \tag{4}$$

e_t in equation (1) is the error term. We consider the following four models by varying the assumptions about the distribution and variance of this error term.

Model 1: MS model with normal error and constant volatility

Model 1 is the simplest one where we assume that the distribution of the error term e_t in equation (1) is normal with the 0 mean and the constant variance σ^2.

$$e_t = \sigma\epsilon_t, \quad \epsilon_t \sim \text{i.i.d.N}(0,1), \tag{5}$$

where "i.i.d.N(0,1)" represents the identically and independently distributed standard normal distribution.

As will be shown in Section 4, this model cannot estimate the business cycle turning points properly. Therefore, we extend this model by making the distribution of ϵ_t a heavy-tailed distribution or/and by making the variance σ^2 time-varying.

Model 2: MS model with *t*-error and constant volatility

We first extend Model 1 by incorporating a heavy-tailed distribution. Specifically, we assume that ϵ_t in (5) follows the Student's t-distribution standardized such that the variance is 1.

$$e_t = \sigma\epsilon_t, \quad \epsilon_t \sim \text{i.i.d. standardized } t(\nu), \tag{6}$$

where "standardized $t(\nu)$" represents the standardized Student's t-distribution with the degree-of-freedom ν. We assume that $\nu > 2$ because the Student's t-distribution would not have a finite variance otherwise.

Model 3: MS model with normal error and SV

In this model, we introduce SV as follows.

$$e_t = \sigma_t\epsilon_t, \quad \epsilon_t \sim \text{i.i.d. N}(0,1), \tag{7}$$

$$\ln\sigma_t^2 = \omega + \psi(\ln\sigma_{t-1}^2 - \omega) + \eta_t \quad \eta_t \sim \text{i.i.d. N}(0,\sigma_\eta^2), \tag{8}$$

where the parameter ψ in equation (8) captures the autocorrelation in the log-variance.

Model 4: MS model with *t*-error and SV

We also estimate the MS model with the both t-error and SV by combing Models 2 and 3.

$$e_t = \sigma_t\epsilon_t, \quad \epsilon_t \sim \text{i.i.d. standardized } t(\nu), \tag{9}$$

$$\ln\sigma_t^2 = \omega + \psi(\ln\sigma_{t-1}^2 - \omega) + \eta_t, \quad \eta_t \sim \text{i.i.d. N}(0,\sigma_\eta^2). \tag{10}$$

3 Methodology

3.1 Bayesian Estimation

Once SV is introduced, it is difficult to evaluate the likelihood. Thus, we employ a Bayesian method using MCMC for Models 3 and 4, which include SV. It is straightforward to evaluate the likelihood of Models 1 and 2, which do not include SV, using the Hamilton (1989) filter, so that we can estimate the parameters in those models using the maximum likelihood method. We, however, apply a Bayesian method using MCMC to Models 1 and 2 as well for comparing them with Models 3 and 4. In this method, we first set the priors of all parameters. Under these priors, we calculate the full conditional posterior distribution of each parameter and latent variable, which is the distribution conditional on the other parameters and latent variables as well as data. We sample parameters and latent variables from thier jount posterior distributions using the Gibbs sampler. The Gibbs sampler enable us to sample parameters and latent variables from thier jount posterior distribution by sampling sequentially from the full conditional posterior distributions. See Appendix A for details on this method.

Model 1

Let $\tilde{\mu} = (\mu^{(0)}, \mu^{(1)})'$, $\tilde{p} = (p_{00}, p_{11})'$, $\boldsymbol{S}_T = (S_1, \ldots, S_T)$ and $\boldsymbol{y}_T = (y_1, \ldots, y_T)$ where T is the sample size. Then, the parameters in Model 1 are $\boldsymbol{\theta} = (\tilde{\mu}, \phi, \sigma^2, \tilde{p})$ and the latent variables are $\boldsymbol{S}_T$. We sample these parameters and latent variables sequentially from the joint posterior distribution by sampling sequentially from the following full conditional posterior densities.

$$f(\tilde{\mu}|\boldsymbol{\theta}_{/\tilde{\mu}}, \boldsymbol{S}_T, \boldsymbol{y}_T), \quad f(\phi|\boldsymbol{\theta}_{/\phi}, \boldsymbol{S}_T, \boldsymbol{y}_T), \quad f(\sigma^2|\boldsymbol{\theta}_{/\sigma^2}, \boldsymbol{S}_T, \boldsymbol{y}_T), \quad f(\tilde{p}|\boldsymbol{\theta}_{/\tilde{p}}, \boldsymbol{S}_T, \boldsymbol{y}_T), \quad f(\boldsymbol{S}_T|\boldsymbol{\theta}, \boldsymbol{y}_T),$$

where $\boldsymbol{\theta}_{/x}$ is the set of all parameters except x.

We set the prior for each parameter as follows.

$$\begin{aligned} &\tilde{\mu} \sim N\left(M_{\mu 0}, \Sigma_{\mu 0}\right) \boldsymbol{I}\left[\mu^{(0)} < \mu^{(1)}\right], \quad \frac{\phi+1}{2} \sim \text{Beta}\left(\alpha_{\phi 0}, \beta_{\phi 0}\right), \quad \sigma^2 \sim \text{IG}\left(\frac{\alpha_{\sigma^2 0}}{2}, \frac{\beta_{\sigma^2 0}}{2}\right), \\ &p_{00} \sim \text{Beta}\left(u_{00}, u_{01}\right), \quad p_{11} \sim \text{Beta}\left(u_{11}, u_{10}\right), \end{aligned} \tag{11}$$

where $\boldsymbol{I}[\,\cdot\,]$ is the indicator function that takes 1 when the inequality in the parenthesis is satisfied and 0 otherwise.

The prior of $\tilde{\mu}$ is the bivariate normal distribution truncated such that $\mu^{(0)} < \mu^{(1)}$ is satisfied. Then, the full conditional posterior distribution of $\tilde{\mu}$ is also the truncated bivariate normal. The prior of σ^2 is the inverted gamma, so that $\sigma^2 > 0$. Then, the full conditional posterior distribution of σ^2 is also the inverted gamma. It is straightforward to sample from these distributions.

For the prior of ϕ, we assume that $(\phi + 1)/2$ follows the beta distributions for stationarity, i.e., $|\phi| < 1$. The priors of $\tilde{p}$ are the independent beta distributions, so that $0 < p_{00} < 1$ and $0 < p_{11} < 1$. Since the full conditional posterior distributions of ϕ and $\tilde{p}$ are non-standard, we sample them using the Metropolis-Hastings (MH) algorithm.

We sample the latent variables $\boldsymbol{S}_T$ from the full conditional posteior distribution using the multimove sampler proposed by Kim and Nelson (1998) (see also Chapter 9 in Kim and Nelson

(1999b)).

Model 2

Equation (6) may be represented as

$$e_t = \sigma\sqrt{\lambda_t} z_t, \quad \frac{\nu - 2}{\lambda_t} \sim \text{i.i.d.}\chi^2(\nu), \quad z_t \sim N(0,1). \tag{12}$$

We treat $\boldsymbol{\lambda}_T = (\lambda_1, \ldots, \lambda_T)$ as latent variables.

The parameters are $\tilde{\theta} = (\tilde{\mu}, \phi, \sigma^2, \tilde{p}, \nu)$ and the latent variables are $\boldsymbol{S}_T$ and $\boldsymbol{\lambda}_T$. We sample the parameters and latent variables from the joint posterior distribution by sampling sequentially from the following full conditional posterior densities.

$$f(\tilde{\mu}|\tilde{\theta}_{/\tilde{\mu}}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{y}_T), \quad f(\phi|\tilde{\theta}_{/\phi}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{y}_T), \quad f(\sigma^2|\tilde{\theta}_{/\sigma^2}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{y}_T), \quad f(\tilde{p}|\tilde{\theta}_{/\tilde{p}}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{y}_T),$$
$$f(\nu|\tilde{\theta}_{/\nu}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{y}_T) \quad f(\boldsymbol{S}|\tilde{\theta}, \boldsymbol{\lambda}_T, \boldsymbol{y}_T), \quad f(\boldsymbol{\lambda}|\tilde{\theta}, \boldsymbol{S}_T, \boldsymbol{y}_T).$$

For parameters common to Model 1, we set the same prior distributions as (11). We set the prior of a new parameter ν as the gamma distribution truncated such that $\nu > 2$ is satisfied.

$$\nu \sim \text{Gamma}(\alpha_{\nu 0}, \beta\nu 0)\boldsymbol{I}[\nu > 2]. \tag{13}$$

Since the full conditional posterior distribution of ν is non-standard, we sample ν using the MH algorithm where the proposal density is selected following Watanabe (2001). For this model, we must also sample $\boldsymbol{\lambda}_T$ from the full conditional posterior distributions. It is straightforward to sample $\boldsymbol{\lambda}_T$ because their full conditional posterior distributions are mutually independent and given by

$$(\epsilon_t^2 + \nu - 2)/\lambda_t \sim \chi^2(\nu + 1), \quad (t = 1, \ldots, T). \tag{14}$$

Model 3

Let $h_t = \ln \sigma_t^2$. Then, equation (8) may be represented as

$$e_t = \exp(h_t/2) z_t, \quad z_t \sim \text{i.i.d.N}(0,1). \tag{15}$$

We treat $\boldsymbol{h}_T = (h_1, \ldots, h_t)$ as latent variables.

Then, the parameters in Model 3 are $\boldsymbol{\theta} = (\tilde{\mu}, \phi, \sigma^2, \tilde{p}, \omega, \psi, \sigma_\eta^2)$ and the latent variables are $\boldsymbol{S}_T$ and $\boldsymbol{h}_T$. We sample the parameters and the latent variables from the joint posterior distribution by sampling sequentially from the following full conditional posterior densities.

$$f(\tilde{\mu}|\tilde{\theta}_{/\tilde{\mu}}, \boldsymbol{S}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\phi|\tilde{\theta}_{/\phi}, \boldsymbol{S}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\tilde{p}|\tilde{\theta}_{/\tilde{p}}, \boldsymbol{S}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\omega|\tilde{\theta}_{/\omega}, \boldsymbol{S}_T, \boldsymbol{h}_T, \boldsymbol{y}_T),$$
$$f(\psi|\tilde{\theta}_{/\psi}, \boldsymbol{S}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\sigma_\eta^2|\tilde{\theta}_{/\sigma_\eta^2}, \boldsymbol{S}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\boldsymbol{S}_T|\boldsymbol{\theta}, \boldsymbol{\lambda}_T, \boldsymbol{y}_T), \quad f(\boldsymbol{h}_T|\boldsymbol{\theta}, \boldsymbol{S}_T, \boldsymbol{y}_T).$$

For parameters common to Model 1, we set the same prior distributions as (11). The priors for new parameters $(\omega, \psi, \sigma_\eta^2)$ are set as follows.

$$\omega \sim N\left(m_{\omega 0}, v_{\omega 0}\right), \quad \frac{\psi+1}{2} \sim \mathrm{Beta}\left(\alpha_{\psi 0}, \beta_{\psi 0}\right), \quad \sigma_\eta^2 \sim \mathrm{IG}\left(\frac{\alpha_{\sigma_\eta^2 0}}{2}, \frac{\beta_{\sigma_\eta^2 0}}{2}\right), \tag{16}$$

We set the prior of ω as a normal. Then, the full conditional posterior distribution is also a normal. We assume the inverted gamma distribution for the prior distribution of σ_η^2. Then, the full conditional posterior distribution is also the inverse gamma distribution. For the prior of ψ, we assume that $(\psi+1)/2$ follow the beta distributions for stationarity, i.e., $|\psi| < 1$. Like ϕ, the full conditional posterior distribution of ψ is non-standard, but it can be sampled using the same MH method as for ϕ.

For this model, we must also sample $\boldsymbol{h}_T$ from the full conditional posterior distributions. We sample $\boldsymbol{h}$ using the block sampler proposed by Watanabe and Omori (2004).

Model 4

Equation (10) may be represented as

$$e_t = \exp(h_t/2)\sqrt{\lambda_t} z_t, \quad \frac{\nu-2}{\lambda}_t \sim \text{i.i.d.}\chi^2(\nu), \quad z_t \sim N(0,1). \tag{17}$$

The parameters are $\boldsymbol{\theta} = (\tilde{\mu}, \phi, \sigma^2, \tilde{p}, \nu, \omega, \psi, \sigma_\eta^2)$ and the latent variables are $\boldsymbol{S}_T$, $\boldsymbol{\lambda}_T$ and $\boldsymbol{h}_T$. We sample the parameters and latent variables from the joint posterior distribution by sampling sequentially from the following full conditional posterior densities.

$$\begin{aligned}
&f(\tilde{\mu}|\boldsymbol{\theta}_{/\tilde{\mu}}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\phi|\boldsymbol{\theta}_{/\phi}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\tilde{p}|\boldsymbol{\theta}_{/\tilde{p}}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{h}_T, \boldsymbol{y}_T),\\
&f(\nu|\boldsymbol{\theta}_{/\nu}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\boldsymbol{S}_T|\boldsymbol{\theta}, \boldsymbol{\lambda}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\omega|\boldsymbol{\theta}_{/\omega}, \boldsymbol{S}_T, \boldsymbol{h}_T, \boldsymbol{y}_T),\\
&f(\psi|\boldsymbol{\theta}_{/\psi}, \boldsymbol{S}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\sigma_\eta^2|\boldsymbol{\theta}_{/\sigma_\eta^2}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{h}_T, \boldsymbol{y}_T), \quad f(\boldsymbol{S}|\boldsymbol{\theta}, \boldsymbol{\lambda}_T, \boldsymbol{h}_T, \boldsymbol{y}_T),\\
&f(\boldsymbol{h}|\boldsymbol{\theta}, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{y}_T).
\end{aligned}$$

The prior distributions of the parameters are set in the same way as in Models 1–3, and the parameters and latent variables are sampled from the full conditional posterior distributions in the same way as in Models 1–3.

3.2 Model Comparion

Model comparison in a Bayesian framework can be performed using the posterior odds ratio. The posterior odds ratio between model i, M_i, and model j, M_j, is given by

$$\mathrm{POR} = \frac{f(M_i|\boldsymbol{y}_T)}{f(M_j|\boldsymbol{y}_T)} = \frac{f(\boldsymbol{y}_T|M_i)}{f(\boldsymbol{y}_T|M_j)}\frac{f(M_i)}{f(M_j)},$$

where $f(\boldsymbol{y}_T|M_i)/f(\boldsymbol{y}_T|M_j)$ and $f(M_i)/f(M_j)$ are called Bayes factor and prior odds ratio respectively. If POR is greater than one, M_i is favored over M_j. The prior odds ratio is usually set to be one, so that the posterior odds ratio is equal to the Bayes factor. To evaluate the Bayes factor, we must calculate $f(\boldsymbol{y}_T|M_i)$ and $f(\boldsymbol{y}_T|M_j)$ called marginal likelihood.

A widely used method for calculating marginal likelihood is the modified harmonic mean method proposed by Geweke (1999). In what follows, we omit M_i and write $\boldsymbol{\theta}_i$ as $\boldsymbol{\theta}$ for simplicity. Let $g(\boldsymbol{\theta})$ be a probability density function. Then, marginal likelihood can be estimated as follows.

$$f(\boldsymbol{y}_T) = \frac{1}{E\left[\frac{g(\boldsymbol{\theta})}{f(\boldsymbol{y}_T|\boldsymbol{\theta})f(\boldsymbol{\theta})}\right]} \approx \left[\frac{1}{M}\sum_{j=1}^{M}\frac{g(\boldsymbol{\theta}_j)}{f(\boldsymbol{y}_T|\boldsymbol{\theta}_j)f(\boldsymbol{\theta}_j)}\right]^{-1} \tag{18}$$

Geweke (1999) proposes to make $g(\boldsymbol{\theta})$ the truncated normal denstity as follows.

$$\begin{aligned} g(\boldsymbol{\theta}) =& \tau^{-1}(2\pi)^{-k/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})\right] \\ &\times \boldsymbol{I}\left[(\boldsymbol{\theta}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu}) \le F^{-1}_{\chi^2_k}(\tau)\right] \end{aligned} \tag{19}$$

where k is the number of parameters and $F^{-1}_{\chi^2_k}(\tau)$ is the inverse function of the χ^2 cdf with the degree-of-freedom k and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the sample mean and covariance matrix of $\boldsymbol{\theta}$ sampled from the posterior distribution using MCMC and $\boldsymbol{I}[\ \cdot\ ]$is the indicator function that takes 1 if the condition in the bracket is satisfied and 0 otherwise.

Watanabe (2014) and Ishihara and Watanabe (2015) use the complete-data likelihood, i.e., $f(\boldsymbol{y}_T, \boldsymbol{S}_T, \boldsymbol{\lambda}_T, \boldsymbol{h}_T|\boldsymbol{\theta})$ for Model 4 for example, instead of the observed-data likelihood, i.e., $f(\boldsymbol{y}_T|\boldsymbol{\theta})$. Chan and Grant (2015) provide evidence that the modified harmonic mean method with the complete-data likelihood has a substantial bias and tends to select the wrong model in practice. For Models 1 and 2, which does not include SV, it is straightforward to evaluate the observed-data likelihood using the Hamilton (1989) filter, so that it suffices to use the modified harmonic mean method with the observed-data likelihood. Once SV is introduced, it is not straightforward to evaluate the observed-data likelihood, so that it is difficult to use the modified harmonic mean method with the observed-data likelihood for Models 3 and 4. For these models, we employ the methods proposed by Chib (1995) and Chib and Jeliazkov (2001). These methods also require the evaluation of the observed-data likelihood, but whereas the modified harmonic mean method requires multiple evaluations, these methods require only one evaluation. Shibata and Watanabe (2005) propose a method for evaluating the observed-data likelihood of the Markov switching stochastic volatility (MSSV) model. Modifying this method, we develop a new method for evaluating the observed-data likelihood of MS models with SV. See Appendix B for details on this method.

From the Bayes theorem, the marginal likelihood can be written as

$$f(\boldsymbol{y}_T) = \frac{f(\boldsymbol{y}_T|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{\theta}|\boldsymbol{y}_T)},$$

where $f(\boldsymbol{y}_T|\boldsymbol{\theta})$ is likelihood, $f(\boldsymbol{\theta})$ is prior density and $f(\boldsymbol{\theta}|\boldsymbol{y}_T)$ is posterior density. The above identity holds for any value of $\boldsymbol{\theta}$, but following Chib (1995), we set $\boldsymbol{\theta}$ equal to its posterior mean $\hat{\boldsymbol{\theta}}$.

It is straightforward to evaluate the prior density $f(\hat{\boldsymbol{\theta}})$. Chib (1995) proposes a method for evaluating the posterior density $f(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_T)$ using the Gibbs sampler. The log of the posterior density can be calculated using the following equation.

$$\ln f(\hat{\boldsymbol{\theta}}|\boldsymbol{y}_T) = \ln f(\hat{\theta}_1|\boldsymbol{y}_T) + \sum_{i=2}^{k} \ln f(\hat{\theta}_i|\hat{\theta}_1, \ldots, \hat{\theta}_{i-1}, \boldsymbol{y}_T). \tag{20}$$

We evalute $f(\hat{\theta}_1|\boldsymbol{y}_T)$ using the draws $(\theta_1^{(j)}, \ldots, \theta_k^{(j)})$ $(j = 1, \ldots, M)$ sampled from $f(\theta_1, \ldots, \theta_k|\boldsymbol{y})$ via the Gibbs sampler as follows.

$$f(\hat{\theta}_1|\boldsymbol{y}_T) = \frac{1}{M}\sum_{j=1}^{M} f(\hat{\theta}_1|\theta_2^{(j)}, \ldots, \theta_k^{(j)}, \boldsymbol{y}_T), \tag{21}$$

and $f(\hat{\theta}_i|\hat{\theta}_1, \ldots, \hat{\theta}_{i-1}, \boldsymbol{y}_T)$ $(i = 2, \ldots, k)$ using the draws $(\theta_i^{(j)}, \ldots, \theta_k^{(j)})$ $(j = 1, \ldots, M)$ sampled from $f(\theta_i, \ldots, \theta_k|\hat{\theta}_1, \ldots, \hat{\theta}_{i-1}, \boldsymbol{y}_T)$ via the Gibbs sampler as follows.

$$f(\hat{\theta}_i|\hat{\theta}_1, \ldots, \hat{\theta}_{i-1}, \boldsymbol{y}_T) = \frac{1}{M}\sum_{j=1}^{M} f(\hat{\theta}_i|\hat{\theta}_1, \ldots, \hat{\theta}_{i-1}, \theta_{i+1}^{(j)}, \ldots, \theta_k^{(j)}, \boldsymbol{y}_T). \tag{22}$$

This method requires the posterior density to be known up to the normalizing constant. The normalizing constant of the posterior densities of ϕ, ψ and ν are not known, but they can be sampled using the MH algorithm. In such a case, we can evaluate the posterior density using the method proposed by Chib and Jeliazkov (2001).

3.3 Structural Changes

Following Ishihara and Watanabe (2015), we also analyze structural changes in the mean growth rates $\mu^{(0)}$ and $\mu^{(1)}$ in equation (2). Assuming that only $\mu^{(0)}$ and $\mu^{(1)}$ are subject to structural changes, we give the subscript t to μ, $\mu^{(0)}$ and $\mu^{(1)}$ as follows.

$$\mu_t = \mu_t^{(0)}(1 - S_t) + \mu_t^{(1)} S_t, \quad \mu_t^{(0)} < \mu_t^{(1)}. \tag{B3'}$$

Let D_t denote the number of structural changes up to time t and n denote the total number of structural changes during the sample period. Then, we can define $\mu_t^{(0)}$ and $\mu_t^{(1)}$ as follows.

$$\mu_t^{(0)} = \begin{cases} \mu^{(00)}, & D_t = 0 \\ \mu^{(01)}, & D_t = 1 \\ \vdots & \\ \mu^{(0i)}, & D_t = i \\ \vdots & \\ \mu^{(0,n-1)}, & D_t = n-1 \\ \mu^{(0n)}, & D_t = n \end{cases}, \quad \mu_t^{(1)} = \begin{cases} \mu^{(10)}, & D_t = 0 \\ \mu^{(11)}, & D_t = 1 \\ \vdots & \\ \mu^{(1i)}, & D_t = i \\ \vdots & \\ \mu^{(1,n-1)}, & D_t = n-1 \\ \mu^{(1n)}, & D_t = n \end{cases} \tag{23}$$

Assuming that D_t follows an irreversible Markov process, we express the transition probabilities as follows.

$$\begin{aligned}
&p(D_t = 0 \mid D_{t-1} = 0) = q_{00},\\
&p(D_t = 1 \mid D_{t-1} = 0) = 1 - q_{00},\\
&p(D_t = 1 \mid D_{t-1} = 1) = q_{11},\\
&p(D_t = 1 \mid D_{t-1} = 0) = 1 - q_{11},\\
&\qquad\vdots\\
&p(D_t = n-1 \mid D_{t-1} = n-1) = q_{n-1,n-1},\\
&p(D_t = n \mid D_{t-1} = n-1) = 1 - q_{n-1,n-1},\\
&p(D_t = n \mid D_{t-1} = n) = 1
\end{aligned} \tag{24}$$

Like $\tilde{\mu}$ in Models 1–4, we set the prior of $\tilde{\mu}_i = [\mu_{i0}, \mu_{i1}]$ $(i = 1, \ldots, n)$ as the truncated normal such that $\mu_{i0} < \mu_{i1}$. Then, the full conditional posterior distribution is also the truncated normal. Like $\tilde{p}$, we set the prior of q_{ii} $(i = 1, \ldots, n-1)$ as the Beta distribution. The full conditional posterior distribution is non-standard, but we can sample them using the acceptance-rejection algorithm in the same way as $\tilde{p}$. Like $\boldsymbol{S}$, we can sample the latent variables $\boldsymbol{D} = (D_1, \ldots, D_T)$ using the multimive sampler.

We choose the total number of structural changes n using the marginal likelihood calculated as mentioned above.

4 Empirical Analysis

4.1 Data

We use the monthly data on the coincident indicator of the composite index (CI) in Japan. ESRI calculates two types of CI with and without outlier replacement. We use CI with outlier replacement. This data are plotted in Figure 1 where the shadow areas are the recession periods published by ESRI. We use the growth rate of CI for y_t. We calculate the growth rate of CI as the percentage log difference of CIs in two consecutive months, which is plotted in Figure 2.

[Insert Figure 1]

[Insert Figure 2]

[Insert Table 1]

The descriptive statistics of the growth rate of CI are summarized in Table 1. The mean is not significantly different from 0. The kurtosis is significantly larger than 3, showing that the growth rate of CI is more leptokurtic than the normal distribution. The skewness is significantly negative, indicating that the growth rate of CI is negatively skewed. We do not, however, take account of the skewness in this paper. The Jarque-Bera (JB) statistic is so large that it rejects the null hypothesis of normality strongly. LB(10) is the Ljung-Box statistics adjusted for heteroskedasticity following Diebold (1988) to test the null hypothesis of no autocorrelations up to 10 lags. According to the values of LB(10), the null hypothesis is rejected at the 1% significance level. The autocorrelation in y_t may fully or partly be explained by the switch in its mean.

4.2 Estimation results for the simple MS model

We first estimate the simple MS model with normal error and constant volatility (Model 1). The parameters in Model 1 are $(\mu^{(0)}, \mu^{(1)}, \phi, \sigma^2, p_{00}, p_{11})$, whose prior distributions are set as follows.

$$(\mu^{(0)}, \mu^{(1)})' \sim N\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}\right) \boldsymbol{I}\left[\mu^{(0)} < \mu^{(1)}\right],$$

$$\frac{\phi+1}{2} \sim \text{Beta}(1,1), \quad \sigma^2 \sim \text{IG}(6,4),$$

$$p_{00} \sim \text{Beta}(9,1), \quad p_{11} \sim \text{Beta}(9,1).$$

We throw away the first $5,000$ draws of all parameters as burn-in and use the next $10,000$ draws for the parameter estimation.

To analyze the impact of large shocks, we estimate the model using the subsample 1985/02–2008/08, which is prior to large shocks, as well as the full sample 1985/02–2025/05.

[Insert Figure 3]

Figures 3 depicts the posterior probabilities of recession for each period estimated using the full sample and the subsample respectively. The posterior probabilities of recession for the subsample almost coincide with the recession periods published by ESRI. Those for the full sample are close to 0 except the periods of the global financial crisis in 2008, the Great East Japan Earthquake in 2011, the consumption tax hike in 2019, and the COVID-19 pandemic in 2020 when they are close to 1. The negative impact of these events is so large that the simple MS model leads to the result that only those five periods are in the recession regime and all other periods are in the boom regime.

The estimation result of each parameter in Model 1 is summarized in Table 2. The mean and standard deviation (SD) are calculated as the sample mean and standard deviation of the $10,000$ draws of each parameter after burn-in. The standard error (SE) of sample mean is calculated using the Parzen window because the draws sampled using MCMC are autocorrelated. The 95% Bayesian credible interval is obtained as the 2.5th and 97.5th percentiles of the $10,000$ draws of each parameter. CD is the convergence diagnostic statistic proposed by Geweke (1992), whose asymptotic distribution is the standard normal if the draws have converged to the ones from the posterior distribution. The standard error of CD statistic is also calculated using the Parzen window. IF is the inefficiency factor proposed by Chib (2001). If this is equal to 2, it implies that the number of draws must be twice as much as that of random sampling to make the both standard errors the same. The inefficiency factor increases with the autocorrelation in draws.

[Insert Table 2]

According to the CD values, the null hypothesis that the $10,000$ draws used for estimation have converged to the ones from the posterior distribution is accepted for all parameters in the both subsample and full sample. The IF values for all parameters are small, indicating that the sampling method is efficient. The mean and 95% interval of $\mu^{(0)}$ in the full sample are much smaller than those in the subsample, showing that the negative impact of the global financial crisis in 2008, the Great East Japan Earthquake in 2011, the consumption tax hike in 2019, and the COVID-19 pandemic in 2020 on the growth rate of CI is large. The mean and 95% interval of ϕ are negative in the subsample while they are positive in the full sample.

4.3 Estimation results for the extended MS model

Next, we estimate the extended models. The new parameter in the MS model with t-error and constant volatility (Model 2) is ν, which is the degree-of-freedom of the Student's t-distribution. Its prior is set as follows.

$$\nu \sim \text{Gamma}(1, 0.1)\boldsymbol{I}[\nu > 2].$$

The priors for all other parameters are set as the same as those in Model 1. The new parameters in the MS model with normal error and SV (Model 3) are $(\omega, \phi, \sigma_\eta^2)$. Their priors are set as follows.

$$\omega \sim N(0, 10), \quad \frac{\psi + 1}{2} \sim \text{Beta}(2, 1), \quad \sigma_\eta^2 \sim \text{IG}(6, 4).$$

The priors for all other parameters are set as the same as those in Model 1. The priors of the parameters in the MS model with t-error and SV (Model 4) are set as the same as before. We throw away the first $10,000$ draws of all parameters as burn-in and use the next $10,000$ draws for the parameter estimation.

Tables 3–5 summarize the estimation results for the extended MS models using the full sample. According to the CD values, the null hypothesis that the $10,000$ draws used for estimation have converged to the ones from the posterior distribution is accepted at the 5% significance level for all parameters in all models. The IF values for all parameters are larger than those in Model 1, but they are still not so large indicating that the sampling method we use is efficient. The mean and 95% interval of $\mu^{(0)}$ and p_{00} are small in Model 1 for the full sample, but in all extended models, they recover close to those in Model 1 for the subsample prior to large shocks. In all extended models, the mean of ϕ is negative. The mean of ν in Model 4 is 30.2283 while that in Model 2 is 2.8527. If volatility changes, the distribution of $e_t = \sigma_t \epsilon_t$ becomes leptokurtic even if the distribution of ϵ_t is normal. The leptokurtosis may be captured by changes in volatility. This is the reason why the mean of ν increases once SV is introduced. The means of ψ are 0.8640 and 0.8540 in Models 3 and 4 respectively, indicating that shocks to volatility are persistent although the persistence is small compared with financial volatility.

[Insert Table 3]

[Insert Table 4]

[Insert Table 5]

For model comparison, we calculate the log marginal likelihoods for all models using the methods explained in Section 3. Table 6 shows the result. We do not report their standard errors because they are close to 0. The log marginal likelihood in Model 3 is significantly larger than those of other models. The conclusion must be that the MS model with normal error and SV (Model 3) fits the data best.

[Insert Table 6]

[Insert Figure 4]

[Insert Figure 5]

[Insert Figure 6]

Figures 4–6 depicts the posterior probabilities of recession for each period estimated by Models 2–4. All models provide the similar posterior probabilities. As is shown in Figure 3, the simple MS model with normal error and constant volatility predicts that the periods of large shocks are in the recession regime, which does not hold true once t-error or SV is introduced. Using the posterior probabilities of recession, we estimate the business cycle turning points as follows. Let $\hat{p}(S_t = 0|\boldsymbol{y}_T)$ be the estimated probability of recession for period t. We define t as peak if $\hat{p}(S_{t-1} = 0|\boldsymbol{y}_T) < 0.5$ and $\hat{p}(S_t = 0|\boldsymbol{y}_T) > 0.5$ and as trough if $\hat{p}(S_{t-1} = 0|\boldsymbol{y}_T) > 0.5$ and $\hat{p}(S_t = 0|\boldsymbol{y}) < 0.5$. Table 7 shows the result. The turning points estimated by Models 2–4 are close to those by ESRI except that Model 4 estimates 2015/08 as peak and 2016/03 as trough.

[Insert Table 7]

[Insert Figure 7]

Figure 7 plots the posterior mean of volatility σ_t^2 estimated using Model 3 (solid line) and Model 4 (dotted line). During the periods of large shocks, the posterior means of volatility estimated by the both models jumps up. The difference in the posterior mean of volatility between Models 3 and 4 is negligible because the degree of freedom ν of the t-distribution in Model 4 is so large that it is no different from using the standard normal distribution.

Since Model 3 (normal error + SV) fits the data best, using this model, we analyze structural changes in in the mean growth rates $\mu^{(0)}$ and $\mu^{(1)}$. We set the priors of new parameters $\tilde{\mu}_i = [\mu^{(0i)}, \mu^{(1i)}]'$ $(i = 0, 1, \ldots, n)$, q_{jj} $(j = 0, 1, \ldots, n-1)$ as follows.

$$\tilde{\mu}_i \sim N\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}\right) \boldsymbol{I}\left[\mu^{(0)} < \mu^{(1)}\right],\ i = 0, 1, \ldots, n$$

$$q_{jj} \sim \text{Beta}(9, 0.1),\ j = 0, 1, \ldots, n-1$$

Table 8 shows the log marginal likelihood for each number of break points between 0 and 4. According to the table, the log marginal likelihood of the model with two break points is the highest. Hereinafter, the model with two structural change points added to model 4 will be called model 5.

[Insert Table 8]

[Insert Figure 8]

Figure 8 shows the posterior distribution of each structural change point estimated from Model 5 with two structural change points. These are the 10,000 samples after the burn-in of $\tilde{D}_T$, and since there are two ts where $D_{t-1} < D_t$, these were extracted and their respective histograms were drawn. The peak of the posterior distribution of the first structural change point is in 2008/10 after the Lehman Shock, and the peak of the posterior distribution of the second structural change point is in 2010/02.

Table 9 shows the parameter estimation results for model 5 with two structural change points. Comparing the posterior means of $\mu^{(00)}$ and $\mu^{(01)}$ and those of $\mu^{(10)}$ and $\mu^{(11)}$, we can see that at the first structural change point, the mean growth rate during recessions declines and that during booms rises. Comparing the posterior means of $\mu^{(01)}$ and $\mu^{(02)}$ and those of $\mu^{(11)}$ and $\mu^{(12)}$, we can see that at the second structural change point, the mean growth rate during recessions rises and that during expansions falls. These results show that this structural change captures the instability of economic fluctuations during financial crises and the subsequent recovery periods. $\mu^{(01)}$ and $\mu^{(02)}$ have large standard errors and standard deviations, but this is thought to be because the period of the economic downturn between the first and second structural turning points and the period of the economic downturn after the second structural turning point are short, and the number of samples that can be used to estimate $\mu^{(01)}$ and $\mu^{(02)}$ is small.

[Insert Table 9]

[Insert Figure 10]

Table 9 shows the business cycle turning points estimated by Model 5 (normal error error + stochastic volatility + 2 change points). This model produces the estimates of peak and trough similar to those published by ESRI and estimated by other models except that this model estimates 2015/02 as peak and 2016/06 as trough.

5 Conclusion

We analyze the business cycles in Japan by applying MS models to the growth rate of the coincident indicator of CI during the period of 1985/01–2025/05 calculated by ESRI. We first show that the impact of the global financial crisis in 2008, the Great East Japan Earthquake in 2011, the consumption tax hikes in 2019, and the COVID-19 pandemic in 2020 on this index is so large that the simple MS model with the normal error and constant volatility cannot estimare the business cycle turning points properly. We extend the MS model by incorporating t-error and SV and employ a Bayesian method via MCMC for the analysis of the extended models. We show that the MS model with t-error or SV provides the estimates of the business cycle turning points close to those published by ESRI. The marginal likelihoods provide evidence that the MS model with normal error and SV fits the data best. Using the MS model with normal error and SV, we also analyze structural changes in CI's mean growth rates during booms and recessions are also analyezed and two break points are found in the both mean growth rates. One is 2008/10 and the other is 2010/02, during which the mean growth rate during recession falls and that during boom rises due to the global financial crisis.

The ESRI's business cycle turning points are determined based on the historical DIs created from each adopted series of the CI coincident index, taking into account discussions at the Committee for Business Cycle Indicators. Although they may not necessarily be the correct turning points, it is important that the MS models extended in this paper estimate the turning points close to those determined carefully by the ESRI.

The SV model has been used to analyze stochastic changes in financial volatility, but this paper shows that it is also important in analyzing business cycles. In recent years, the SV model has become increasingly important in macroeconometric analysis. For example, the SV model is used in time-varying VAR models (Primiceri 2005 and Nakajima et al. 2011)

The models developed in this paper may also be applicable to macro and financial stress tests.

Appendix A: Prior and full conditional posterior distributions

In this appendix, we explain the prior distribution of each parameter, the full conditional posterior distribution and the sampling method from there.

Model 1

(1) $\tilde{\mu} = (\mu^0, \mu_1^0)'$

We set the prior of $\tilde{\mu}$ as the following truncated normal distribution.

$$\tilde{\mu} \sim \mathrm{N}(M_0, \Sigma_0) I(\mu^{(0)} < \mu^{(1)}), \tag{A1}$$

where $I(\mu^{(0)} < \mu^{(1)})$ is the indication function that takes 1 if the inequality in the parenthesis is satisfied and 0 otherwise. Therefore, this prior distribution is a bivariate normal distribution with mean $M_{\mu 0}$ and covariance matrix $\Sigma_{\mu 0}$, truncated so that only the range $\mu^{(0)} < \mu^{(1)}$ remains.

Under this prior distribution, the full conditional posterior distribution of $\tilde{\mu}$ is also the bivariate truncated normal distribution as follows:

$$\tilde{\mu} \mid \phi, \sigma^2, \tilde{p}, \tilde{S}_T, \tilde{y}_T \sim \mathrm{N}(M_{\mu 0}, \Sigma_{\mu 0}) I(\mu^{(0)} < \mu^{(1)}) \tag{A2}$$

where

$$\Sigma_{\mu 1} = (\Sigma_{\mu 0}^{-1} + X^{'} X)^{-1}, \quad M_{\mu 1} = \Sigma_{\mu 1}(\Sigma_{\mu 0}^{-1} M_0 + X'Y) \tag{A3}$$

$$Y = \left[\sqrt{1-\phi^2} y_1/\sigma, (y_2 - \phi y_1)/\sigma, \ldots, (y_{\tau_1 - 1} - \phi y_{\tau_1 - 2})/\sigma\right]' \tag{A4}$$

$$X = \begin{bmatrix} \sqrt{1-\phi^2}(1-S_1)/\sigma, & \sqrt{1-\phi^2} S_1/\sigma \\ \{(1-S_2) - \phi(1-S_1)\}/\sigma, & (S_2 - \phi S_1)/\sigma \\ \vdots & \vdots \\ \{(1-S_T) - \phi(1-S_{T-1})\}/\sigma, & (S_T - \phi S_{T-1})/\sigma \end{bmatrix} \tag{A5}$$

To sample from this distribution, we first sample from a bivariate normal distribution with mean $M_{\mu 1}$ and variance $\Sigma_{\mu 1}$. If the sampled value does not satisfy $\mu^{(0)} < \mu^{(1)}$, we discard it and take a new sample.

(2) ϕ

If the prior distribution of ϕ is $\pi(\phi)I(-1 < \phi < 1)$, then the full conditional posterior probability density function of ϕ can be expressed as follows:

$$\pi(\phi|\tilde{\mu}, \sigma^2, \tilde{p}, \tilde{S}_T, \tilde{y}_T) \propto \pi(\phi)\sqrt{1-\phi^2}\exp\left\{-\frac{(\phi - m_\phi)^2}{2\sigma_\phi^2}\right\} I(-1 < \phi < 1) \tag{A6}$$

where

$$m_\phi = \frac{\sum_{t=2}^{T}(y_t - \mu_t)(y_{t-1} - \mu_{t-1})}{\sum_{t=2}^{T-1}(y_t - \mu_t)^2}, \qquad \sigma_\phi^2 = \frac{\sigma^2}{\sum_{t=2}^{T-1}(y_t - \mu_t)^2}.$$

We use the MH algorithm to sample from this full conditional posterior probability density function as follows. If we ignore $\pi(\phi)\sqrt{1-\phi^2}$ in equation (A6), we obtain the kernel of a trucated normal distribution $\mathrm{N}(m_\phi, \sigma_\phi^2)I(-1 < \phi < 1)$. Therefore, we sample a proposal ϕ^* from this truncated normal distribution and accept this proposal with the following probability:

$$\min\left\{\frac{\pi(\phi^*)\sqrt{1-\phi^{*2}}}{\pi(\phi)\sqrt{1-\phi^2}}, 1\right\}.$$

where ϕ is the value sampled in the previous sampling. If ϕ^* is rejected, ϕ is sampled again. In this paper, we assume that $(1+\phi)/2$ follows a beta distribution, so that $-1 < \phi < 1$.

(3) σ^2

We assume the following inverse gamma distribution for the prior distribution of σ^2:

$$\sigma^2 \sim IG\left(\frac{\alpha_{\sigma^2 0}}{2}, \frac{\beta_{\sigma^2 0}}{2}\right) \tag{A7}$$

Under this prior distribution, the full conditional posterior distribution is also an inverse gamma distribution.

$$\sigma^2|\tilde{\mu}, \phi, \tilde{p}, \tilde{S}_T, \tilde{y}_T \sim IG\left(\frac{\alpha_{\sigma^2 1}}{2}, \frac{\beta_{\sigma^2 1}}{2}\right) \tag{A8}$$

where

$$\alpha_{\sigma^2 1} = \alpha_{\sigma^2 0} + T, \quad \beta_{\sigma^2 1} = \beta_{\sigma^2 0} + (1-\phi^2)(y_1 - \mu_1)^2 + \sum_{t=2}^{T}\{y_t - \mu_t - \phi(y_{t-1} - \mu_{t-1})\}^2$$

To sample from this distribution, we simply sample from the gamma distribution and take the inverse.

(4) $\boldsymbol{S}_T = (S_1, \ldots, S_T)$

To sample $\boldsymbol{S}_T$, we use the multi-move sampler proposed by Kim and Nelson (1998, 1999b). Define $\tilde{S}^{t+1} = (S_{t+1}, \cdots, S_T)'$. Then, the full conditional posterior probability of $\tilde{S}_T$ can be represented as follows.

$$\pi(\boldsymbol{S}_T \mid \tilde{y}_T, \boldsymbol{\theta}) = \pi(S_T \mid \tilde{y}_T, \boldsymbol{\theta}) \cdots \pi(S_t \mid \tilde{S}^{t+1}, \tilde{y}_T, \boldsymbol{\theta}) \cdots \pi(S_1 \mid \tilde{y}_T, \boldsymbol{\theta}) \tag{A9}$$

To sample $\tilde{S}_T$ from this distribution, we first sample S_T from the first term on the right-hand side, $\pi(S_T \mid \boldsymbol{y}, \boldsymbol{\theta})$. Based on the sampled value of S_T, we then sample S_{T-1} from the second term on the right-hand side, $\pi(S_{T-1} \mid S^T, \boldsymbol{y}, \boldsymbol{\theta})$. In this way, we can sample from each term on the right-hand side in reverse order up to S_1. Each term on the right side can be expressed as follows:

$$\pi(S_t = i \mid S^{t+1}, \tilde{y}_t, \boldsymbol{\theta}) = \frac{\pi(S_t = i \mid \tilde{y}_t, \boldsymbol{\theta})\pi(S_{t+1} \mid S_t = i, \tilde{p})}{\sum_{j=0}^{1} \pi(S_t = j \mid \tilde{y}_t, \tilde{\theta})\pi(S_{t+1} \mid S_t = j, \tilde{p})}, \quad i = 0, 1 \tag{A10}$$

The second term in the numerator and denominator of the right-hand side of this equation is the transition probability (4). The first term is the stationary probability. Starting from the stationary probabilities

$$\pi(S_0 = 1 \mid \tilde{\theta}) = \pi(S_0 = 1 \mid \tilde{p}) = \frac{1 - p_{00}}{2 - p_{00} - p_{11}} \tag{A11}$$

$$\pi(S_0 = 0 \mid \tilde{\theta}) = \pi(S_0 = 0 \mid \tilde{p}) = \frac{1 - p_{11}}{2 - p_{00} - p_{11}} \tag{A12}$$

we proceed in time order until $t = T$ using the Hamilton (1989) filter.

To sample $\tilde{S}_T$ from (A9), the following two steps are sufficient.

(i) First, starting from the stationary probabilities (A11) and (A12), run the Hamilton (1989) filter in time order from $t = 1$ to $t = T$, and calculate $\pi(S_t \mid \tilde{y}_t, \theta)$ $(t = 1, \ldots, T)$.

(ii) Next, sample S_T using $\pi(S_T \mid \tilde{y}_T, \theta)$, calculated last in (i). Specifically, we sample $S_T = 1$, with probability $\pi(S_T = 1 \mid \tilde{y}_T, \tilde{\theta})$ and $S_T = 0$ with the remaining probability. Then, we use the sampled S_T to calculate the following probability.

$$\begin{aligned} &\pi(S_{T-1} = 1 \mid S^T, \tilde{y}_T, \tilde{\theta}) \\ &\quad = \frac{\pi(S_{T-1} = 1 \mid \tilde{y}_{T-1}, \tilde{\theta})\pi(S_T \mid S_{T-1} = 1, \tilde{p})}{\sum_{j=0}^{1} \pi(S_{T-1} = j \mid \tilde{y}_t, \tilde{\theta})\pi(S_T \mid S_{T-1} = j, \tilde{p})} \end{aligned} \tag{A13}$$

where $\pi(S_{T-1} = 1 \mid \tilde{y}_{T-1}, \tilde{\theta})$ is calculated using the Hamilton (1989) filter in (i), and $\pi(S_T \mid S_{T-1} = 1, \tilde{p})$ is the transition probability (4). Sample $S_{T-1} = 1$ with this probability, and $S_{T-1} = 0$ with the remaining probability. Repeat this procedure backwards in time until $t = 1$.

(5) $\tilde{p} = [p_{00}, p_{11}]'$

We set a prior for transition probabilities $\tilde{p} = [p_{00}, p_{11}]'$ as the following mutually independent beta distribution.

$$p_{00} \sim Beta(u_{00}, u_{01}), \quad p_{11} \sim Beta(u_{11}, u_{10}) \tag{A14}$$

By using the beta distribution, p_{00} and p_{11} are always between 0 and 1.

Under this prior distribution, the full conditional posterior distribution is:

$$\pi(\tilde{p} \mid \tilde{S}_T) \propto \frac{(1-p_{00})^{S_1}(1-p_{11})^{1-S_1}}{2-p_{00}-p_{11}} \times p_{00}^{u_{00}+n_{00}}(1-p_{00})^{u_{01}+n_{01}} p_{11}^{u_{11}+n_{11}}(1-p_{11})^{u_{10}+n_{10}} \tag{A15}$$

where n_{00} is the number of ts for which $S_t = 0$ and $S_{t+1} = 0$ in the condition $\tilde{S}_T$, n_{01} is the number of ts for which $S_t = 0$ and $S_{t+1} = 1$, and similarly for n_{11} and n_{10}. These can be obtained from the samples of $\tilde{S}_T$. Since $0 \leq \frac{(1-p_{00})^{S_1}(1-p_{11})^{1-S_1}}{2-p_{00}-p_{11}} \leq 1$, we use the accept-reject algorithm. If we omit $\frac{(1-p_{00})^{S_1}(1-p_{11})^{1-S_1}}{2-p_{00}-p_{11}}$, equation (A15) is the mutually independent beta distribution as follows,

$$p_{00} \sim Beta(u_{00}+n_{00}, u_{01}+n_{01}), \quad p_{11} \sim Beta(u_{11}+n_{11}, u_{10}+n_{10}) \tag{A16}$$

Therefore, we sample from these mutually independent beta distributions and accept them with probability $\frac{(1-p_{00})^{S_1}(1-p_{11})^{1-S_1}}{2-p_{00}-p_{11}}$. If they are rejected, we repeat sampling until they are accepted.

Model 2

We set the same priors for $\tilde{\mu} = (\mu^{(0)}, \mu^{(1)})'$, ϕ, σ^2 and $\tilde{p} = (p_{00}, p_{11})'$ as those in Model 1. Using latent variables $\boldsymbol{\lambda}_T = (\lambda_1, \ldots, \lambda_T)$ defined by equation (**??**), equation (1) may be represented as

$$y_t = \mu_t + \phi(y_{t-1} - \mu_{t-1}) + \sigma\sqrt{\lambda_t} z_t, \quad (\nu-2)/\lambda_t \sim \chi^2(\nu), \quad z_t \sim \text{i.i.d.} N(0,1). \tag{A17}$$

Thus, the full conditional posterior distributions of these parameters are those of Model 1 with σ replaced by $\sigma\sqrt{\lambda_t}$. The full conditional posterior distribution of S_t is similar, and these can be sampled in the same way as Model 1.

We set the prior distribution of ν as the gamma distribution truncated such that $\nu > 2$ is satisfied. Since the full conditional posterior distribution of ν is non-standard. we sample it using the MH algorithm where the proposal density is selected following Watanabe (2001).

For this model, we must also sample $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)$ and ν from their full conditional posterior distributions. It is straightforward to sample $\boldsymbol{\lambda}$ because the full conditional posterior distribution of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)$ are mutually independent and given as

$$(\epsilon_t^2 + \nu - 2)/\lambda_t \sim \chi^2(\nu+1), \quad (t = 1, \ldots, T). \tag{A18}$$

Model 3

We set the same priors for $\tilde{\mu}$, ϕ, σ^2 and $\tilde{p}$ as those in Models 1 and 2. Using latent variables $\boldsymbol{h}$, equation (1) may be represented as

$$y_t = \mu_t + \phi(y_{t-1} - \mu_{t-1}) + \exp(h_t/2) z_t, \quad z_t \sim \text{i.i.d.N}(0,1). \tag{A19}$$

Thus, the full conditional posterior distributions of these parameters are those of Model 1 with σ replaced by $\exp(h_t/2)$. The full conditional posterior distribution of S_t is similar, and these can be sampled in the same way as Model 1.

For this model, we must also sample $\boldsymbol{h}_T$, ω, ψ and σ_η^2 from their full conditional posterior distributions. We sample $\boldsymbol{h}_T$ using the block sampler proposed by Watanabe and Omori (2004). We set the prior of ω as a normal. Then, the full conditional posterior distribution is also a normal. For the prior of ψ, we assume that $(\psi+1)/2$ follow the beta distributions for stationarity, i.e., $|\psi| < 1$. Like ϕ, the full conditional posterior distribution of ψ is non-standard, but it can be sampled using the same MH method as for ϕ. We assume the inverted gamma distribution for the prior distribution of σ_η^2. Then, the full conditional posterior distribution is also the inverse gamma distribution.

Model 4

The parameters in Model 4 are $\boldsymbol{\theta} = (\tilde{\mu}, \phi, \tilde{p}, \nu, \omega, \psi, \sigma_\eta^2)$ and the latent variables are $\boldsymbol{S}_T$, $\boldsymbol{\lambda}_T$ and $\boldsymbol{h}_T$. We set the same priors for $\tilde{\mu}$, ϕ, $\tilde{p}$, ν, ω, ψ and σ_η^2 as those in Models 2 and 3. Using latent variables $\boldsymbol{\lambda}_T$ and $\boldsymbol{h}_T$, equation (1) may be represented as

$$y_t = \mu_t + \phi(y_{t-1} - \mu_{t-1}) + \exp(h_t/2)\sqrt{\lambda_t} z_t, \quad (\nu-2)/\lambda_t \sim \chi^2(\nu), \quad z_t \sim \text{i.i.d.N}(0,1). \tag{A20}$$

Thus, the full conditional posterior distributions of these parameters are those of Models 3 with $\exp(h_t/2)$ replaced by $\exp(h_t/2)\sqrt{\lambda_t}$.

Appendix B: Likekihood

We evaluate the likelihood of Models 1 and 2, whichi do not include SV, using the Hamilton (2009) filter. It is difficult to evaluate the likelihood of Models 3 and 4, which include SV. Shibata and Wtanabe (2005) propose a method for calculating the likelihood of the Markov switching SV model as follows.

$$\begin{aligned} y_t =& \sigma_t \epsilon_t, \quad \epsilon_t \sim \text{i.i.d. N}(0,1), & \text{(B1)} \\ \ln \sigma_t^2 =& \omega_t + \psi(\ln \sigma_{t-1}^2 - \omega_{t-1}) + \eta_t, \quad \eta_t \sim \text{i.i.d. N}(0, \sigma_\eta^2), & \text{(B2)} \end{aligned}$$

where y_t is financial return and

$$\omega_t = \omega^{(0)}(1 - S_t) + \omega^{(1)} S_t, \quad \omega^{(0)} < \omega^{(1)}, \tag{B3}$$

$$S_t = \begin{cases} 1 & \text{high volatility} \\ 0 & \text{low volatility} \end{cases}. \tag{B4}$$

S_t is assumed to follow a Markov process with transition probabilities defined by equation (4). We modify this mothod for calculating the likelihood of models 4 and 5 in this paper.

Let $\boldsymbol{y}_t$ denote $(y_1, \ldots, y_t)$. Then, the likelihood can be expressed as

$$f(\boldsymbol{y}_T) = f(y_1) \prod_{t=1}^{T-1} f(y_{t+1}|\boldsymbol{y}_t).$$

As for model 3, $f(y_{t+1}|\boldsymbol{y}_t)$ is written as[2]

$$f(y_{t+1}|\boldsymbol{y}_t) = \int f(y_{t+1}|S_{t+1}, S_t, h_{t+1}, y_t) f(h_{t+1}|h_t) \times \pi(S_{t+1}|S_t) f(S_t, h_t|\boldsymbol{y}_t) dh_{t+1} dh_t dS_{t+1} dS_t. \quad \text{(B5)}$$

Suppose that we have M draws $(h_{t|t}^{(m)}, S_{t|t}^{(m)})$ $(m = 1, \ldots, M)$ sampled from the filtering density $f(h_t, S_t|\boldsymbol{y}_t)$. Then, we can sample $S_{t+1|t}^{(m)}$ $(m = 1, \ldots, M)$ using the transition probability $\pi(S_{t+1}|S_{t|t}^{(m)})$ and $h_{t+1|t}^{(m)}$ $(m = 1, \ldots, M)$ from $f(h_{t+1}|h_{t|t}^{(m)})$, which is the normal density with mean $\omega + \psi(h_{t|t}^{(m)} - \omega)$ and variance σ_η^2. Using $h_{t+1|t}^{(m)}$ $(m = 1, \ldots, M)$, equation (B5) can be evaluated as

$$f(y_{t+1}|\boldsymbol{y}_t) \approx \frac{1}{M} \sum_{m=1}^{M} f(y_{t+1}|S_{t+1|t}^{(m)}, S_{t|t}^{(m)}, h_{t+1|t}^{(m)}, y_t).$$

Define

$$A_{t+1|t}^{(m)} \equiv y_t - \mu_0 S_{t+1|t}^{(m)} - \mu_1(1 - S_{t+1|t}^{(m)}) - \phi \left\{ y_{t-1} - \mu_0 S_{t|t}^{(m)} - \mu_1(1 - S_{t|t}^{(m)}) \right\}.$$

Then,

$$f(y_{t+1}|S_{t+1|t}^{(m)}, S_{t|t}^{(m)}, h_{t+1|t}^{(m)}, y_t) = \frac{1}{\sqrt{2\pi \exp(h_{t+1|t}^{(m)})}} \exp\left[-\frac{A_{t+1|t}^{(m)2}}{2\exp(h_{t+1|t}^{(m)})} \right].$$

Suppose that we have M draws $(h_{t-1|t-1}^{(m)}, S_{t-1|t-1}^{(m)})$ $(m = 1, \ldots, M)$ sampled from $f(h_{t-1}, S_{t-1}|\boldsymbol{y}_{t-1})$. Then, the filtering density of model 3 can be written as

$$\begin{aligned} & f(h_t, S_t|\boldsymbol{y}_t) \\ & \propto \int f(y_t|h_t, S_t, S_{t-1}, y_{t-1}) f(h_t|h_{t-1}) \pi(S_t|S_{t-1}) \\ & \qquad \times f(h_{t-1}, S_{t-1}|\boldsymbol{y}_{t-1}) dh_{t-1} dS_{t-1}, \\ & \approx \frac{1}{M} \sum_{m=1}^{M} f(y_t|h_t, S_t, S_{t-1|t-1}^{(m)}, y_{t-1}) f(h_t|h_{t-1|t-1}^{(m)}) \pi(S_t|S_{t-1|t-1}^{(m)}). \end{aligned} \quad \text{(B6)}$$

Define

$$B_{t|t-1}^{(m)} \equiv y_t - \mu_0 S_t - \mu_1(1 - S_t) - \phi \left\{ y_{t-1} - \mu_0 S_{t-1|t-1}^{(m)} - \mu_1(1 - S_{t-1|t-1}^{(m)}) \right\}$$

Then, for Model 3 with normal error, we have

$$\ln f(y_t|h_t, S_t, S_{t-1|t-1}^{(m)}, y_{t-1}) = \text{const} - \frac{1}{2} h_t - \frac{B_{t|t-1}^{(m)2}}{2} \exp(-h_t). \quad \text{(B7)}$$

[2]Strictly speaking, integration with respect to S_{t+1} and S_t must be replaced by summation because they are discrete variables that take 0 or 1, but we use integration for simplicity.

Define

$$\ln f^*(h_t, m) = -\frac{1}{2}h_t - \frac{B^{(m)2}_{t|t-1}}{2}\exp(-h_t). \tag{B8}$$

Since this function is concave, applying the first-order Taylor expansion to $\ln f^*(h_t)$ around $h_t = \hat{h}_t$, which we set as the posterior mean of h_t, yields the following inequality.

$$\begin{aligned} \ln f^*(h_t, m) &\leq -\frac{1}{2}h_t - \frac{B^{(m)2}_{t|t-1}}{2}\exp(-\hat{h}_t)(1 + \hat{h}_t - h_t), \\ &\equiv \ln g^*(h_t, m). \end{aligned} \tag{B9}$$

The product of $g^*(h_t, m)$ and $f(h_t|h^{(m)}_{t-1|t-1})\pi(S_t|S^{(m)}_{t-1|t-1})$ appeared in equation (B6) can be expressed as

$$g^*(h_t, m)f(h_t|h^{(m)}_{t-1|t-1})\pi(S_t|S^{(m)}_{t-1|t-1}) \propto \pi(S_t, m)f_N(h_t|h^{*(m)}_t, \sigma^2_\eta), \tag{B10}$$

where $f_N(h_t|h^{*(m)}_t, \sigma^2_\eta)$ is the normal density with mean $h^{*(m)}_t$ and variance σ^2_η, and

$$h^{*(m)}_t = \omega + \psi h^{(m)}_{t-1|t-1} + \frac{\sigma^2_\eta}{2}\left\{B^{(m)2}_{t|t-1}\exp(-\hat{h}_t) - 1\right\}, \tag{B11}$$

$$\pi(S_t, m) = \exp\left[-\frac{\{\omega + \phi(h^{(m)}_{t-1|t-1} - \omega)\}^2 - h^{*(m)2}_t}{2\sigma^2_\eta}\right]\pi(S_t|S^{(m)}_{t-1|t-1}). \tag{B12}$$

Thus, equation (B6) may be written as

$$\begin{aligned} f(h_t, S_t|\boldsymbol{y}_t) &\propto \frac{1}{M}\sum_{m=1}^{M} f^*(h_t, m)f(h_t|h^{(m)}_{t-1|t-1})\pi(S_t|S^{(m)}_{t-1|t-1}), \\ &\leq \frac{1}{M}\sum_{m=1}^{M} g^*(h_t, m)f(h_t|h^{(m)}_{t-1|t-1})\pi(S_t|S^{(m)}_{t-1|t-1}), \\ &\propto \frac{1}{M}\sum_{m=1}^{M} \pi(S_t, m)f_N(h_t|h^{*(m)}_t, \sigma^2_\eta). \end{aligned}$$

Therefore, we can sample from the filtering density $f(h_t|\boldsymbol{y}_t)$ using the accept-reject algorithm. Fisrt, we draw a proposal (h_t, S_t) from the mixture of M normal densities

$$\sum_{m=1}^{M} \pi^*(S_t, m)f_N(h_t|h^{*(m)}_t, \sigma^2_\eta),$$

where $\pi^*(S_t, m) = \pi(S_t, m)/\sum_{m=1}^{M}\pi(S_t, m)$. We can sample from this mixture distribution by first selecting the indices (S_t, m) with probability

$$\frac{\pi(S_t, m)}{\sum_{S_t=0}^{1}\sum_{m=1}^{M}\pi(S_t, m)},$$

and then sampling from $f_N(h_t|h_t^{*(m)},\sigma_\eta^2)$. Second, we accept it with probability $f^*(h_t)/g^*(h_t)$. If rejected, we return to the first step and draw a new proposal.

For Model 4 with $t-$error, equations (B7)–(B9) are replaced by

$$\begin{aligned}
&\ln f(y_t|h_t, S_t, S_{t-1|t-1}^{(m)}, y_{t-1}) \\
&= \text{const} - \frac{1}{2}h_t - \frac{\nu+1}{2}\ln\left[1+\frac{B_{t|t-1}^{(m)2}}{\nu-2}\exp(-h_t)\right]. \qquad \text{(B13)}\\
&\ln f^*(h_t, m) \\
&= -\frac{1}{2}h_t - \frac{\nu+1}{2}\left[1+\frac{B_{t|t-1}^{(m)2}}{\nu-2}\exp(-h_t)\right]^{-1}\frac{B_{t|t-1}^{(m)2}}{\nu-2}\exp(-h_t). \qquad \text{(B14)}\\
&\ln f^*(h_t, m) \\
&\leq -\frac{1}{2}h_t - \frac{\nu+1}{2}\left[1+\frac{B_{t|t-1}^{(m)2}}{\nu-2}\exp(-\hat{h}_t)\right]^{-1}\frac{B_{t|t-1}^{(m)2}}{\nu-2} \\
&\quad \times \exp(-\hat{h}_t)(1+\hat{h}_t-h_t), \\
&\equiv \ln g^*(h_t, m). \qquad \text{(B15)}
\end{aligned}$$

and equation (B11) is replaced by

$$\begin{aligned}
h_t^{*(m)} =& \omega + \psi h_{t-1|t-1}^{(m)} + \frac{\sigma_\eta^2}{2}\frac{\nu+1}{2}\left[1+\frac{B_{t|t-1}^{(m)2}}{\nu-2}\exp(-\hat{h}_t)\right]^{-1} \\
&\times\left\{B_{t|t-1}^{(m)2}\exp(-\hat{h}_t)-1\right\}. \qquad \text{(B16)}
\end{aligned}$$

References

Chan, J. C. C. and Grant, A. L. (2015), "Pitfalls of estimating the marginal likelihood using the modified harmonic mean," *Economics Letters*, 131(1), 29–33.

Chib, S. (1995), "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, 90(432), 1313–1321.

Chib, S. (2001), "Markov Chain Monte Carlo Methods: Computation and Inference," in J. J. Hechman and E. Leeper (eds.), *Handbook of Econometrics*, Vol.5, Chapter 57, Elsevier, pp.3569–3649.

Chib S., and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*, 96(453), 270–291.

Diebold, F. (1988), *Empirical Modeling of Exchange Rate Dynamics*, Springer-Verlag.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior moments," (with discussion), in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics* 4, Oxford University Press, pp.169–191.

Geweke, J. (1999), “Using simulation methods for Bayesian econometric models: inference, developments, and communications,” *Econometric Reviews*, 18(1), 1–127.

Hamilton, J. D. (1989), “A new approach to the economic analysis of nonstationary time series and business cycle,” *Econometrica*, 57(2), 357–384.

Ishihara, T. and Watanabe, T. (2015), “Econometric analysis of business cycles: A survey with the application to the composite index in Japan,” (in Japanese), *Economic Review*, Institute of Economic Research, Hitotsubashi University, 66(2), 145–168.

Kim, C.-J. and Nelson, C. R. (1998), “Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching,” *Review of Economics and Statistics*, 80(2), 188–201.

Kim, C.-J. and Nelson, C. R. (1999a), “Has the U.S. economy become more stable? A Bayesian approach based on a Markov-Switching model of the business cycle,” *Review of Economics and Statistics*, 81(2), 608–616.

Kim, C.-J. and Nelson, C. R. (1999b), *State-Space Models with Regime Switching*, MIT Press.

Nakajima, J., Kasuya, M. and Watanabe, T. (2011). ”Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy,” *Journal of the Japanese and International Economies*, 25(3), 225–245.

Primiceri, G. E., (2005), “ Time varying structural vector autoregressions and monetary policy, ” *Review of Economic Studies*, 72 (3), 821-852.

Shibata, M. and Watanabe, T. (2005), “Bayesian analysis of a Markov switching stochastic volatility model,” *Journal of the Japan Statistical Society*, 35(2), 205–219.

Watanabe, T. (2001), “On sampling the degree-of-freedom of student’s t–disturbances,” *Statistics and Probability Letters*, 52(2), 177–181.

Watanabe, T. (2014), “Bayesian analysis of business cycle in Japan using Markov switching model with stochastic volatility and fat-tail distribution,” *Economic Review*, Institute of Economic Research, Hitotsubashi University, 65(2), 156–167.

Watanabe, T. and Omori, Y. (2004), “A multi-move sampler for estimating non-Gaussian time series model: Comments on Shephard & Pitt (1997),” *Biometrika*, 91(1), .246–248.

Table 1: Descriptive statistics of the growth rate of CI

Mean	SD	Skewness	Kurtosis	JB	LB(10)
0.0511	1.4889	-2.2838	16.9305	4334.24	22.38
(0.0677)		(0.1113)	(0.2227)		

The sample period is 1985/2–2025/05 and the sample size is 484. The numbers in parentheses are standard errors. SD is the standard deviation. JB is the Jarque-Bera statistic to test the null hypothesis of normality. LB(10) is the Ljung-Box statistics adjusted for heteroskedasticity following Diebold (1988) to test the null hypothesis of no autocorrelations up to 10 lags.

Table 2: Estimation result for Model 1 (normal error + constant volatility)

Full Sample (1985/02–2025/05)

	Mean	SE	SD	95% Interval	CD	IF
$\mu^{(0)}$	-6.7753	0.0070	0.4863	[-7.7374, -5.8170]	-0.56	2.28
$\mu^{(1)}$	0.1741	0.0004	0.0605	[0.0535, 0.2931]	-0.20	1.11
ϕ	0.1257	0.0004	0.0472	[0.0318, 0.2165]	-0.52	1.77
σ^2	1.3114	0.0010	0.0846	[1.1554, 1.4870]	1.82	1.30
p_{00}	0.7337	0.0009	0.1024	[0.5147, 0.9064]	0.54	1.46
p_{11}	0.9900	0.0000	0.0046	[0.9793, 0.9969]	-0.43	1.63

Subsample (1985/02–2008/08)

	Mean	SE	SD	95% Interval	CD	IF
$\mu^{(0)}$	-0.7481	0.0020	0.1132	[-0.9841, -0.5381]	-0.73	2.05
$\mu^{(1)}$	0.3830	0.0009	0.0550	[0.2768, 0.4928]	-0.63	2.55
ϕ	-0.2941	0.0007	0.0610	[-0.4109, -0.1714]	1.71	1.55
σ^2	0.6954	0.0007	0.0617	[0.5821, 0.8267]	-1.70	1.46
p_{00}	0.9323	0.0005	0.0335	[0.8517, 0.9806]	-1.36	1.12
p_{11}	0.9698	0.0003	0.0139	[0.9368, 0.9903]	0.45	3.67

SE is the standard error of mean and SD is the standard deviation. CD is the convergence diagnostic statistics proposed by Geweke (1992). IF is the inefficiency factor proposed by Chib (2001).

Table 3: Estimation result for Model 2 (t-error + constant volatility)

	Mean	SE	SD	95% Interval	CD	IF
$\mu^{(0)}$	-0.5870	0.0061	0.1536	[-0.9150, -0.3114]	-0.79	14.60
$\mu^{(1)}$	0.4486	0.0028	0.0740	[0.3108, 0.6022]	-0.65	12.54
ϕ	-0.0526	0.0013	0.0523	[-0.1546, 0.0501]	-0.79	7.55
σ^2	1.9650	0.0025	0.1467	[1.6980, 2.2728]	-1.24	2.09
p_{00}	0.9192	0.0005	0.0296	[0.8483, 0.9653]	0.11	3.84
p_{11}	0.9652	0.0004	0.0154	[0.9288, 0.9881]	1.16	8.25
ν	2.8527	0.0103	0.1558	[2.6013, 3.2064]	0.28	34.16

SE is the standard error of mean and SD is the standard deviation. CD is the convergence diagnostic statistics proposed by Geweke (1992). IF is the inefficiency factor proposed by Chib (2001).

Table 4: Estimation result for Model 3 (normal error + SV)

	Mean	SE	SD	95% Interval	CD	IF
$\mu^{(0)}$	-0.4597	0.0080	0.1884	[-0.8770, -0.1455]	-0.29	31.75
$\mu^{(1)}$	0.3892	0.0022	0.0646	[0.2710, 0.5219]	-0.34	20.52
ϕ	-0.1368	0.0011	0.0584	[-0.2520, -0.0240]	0.79	7.34
p_{00}	0.9175	0.0006	0.0325	[0.8399, 0.9670]	0.67	5.86
p_{11}	0.9617	0.0007	0.0190	[0.9158, 0.9892]	1.46	13.93
ω	-0.0113	0.0020	0.2183	[-0.4472, 0.4223]	0.28	1.11
ψ	0.8640	0.0017	0.0394	[0.7791, 0.9334]	-0.51	25.13
σ_η^2	0.3640	0.0057	0.0930	[0.2111, 0.5832]	0.0082	54.69

SE is the standard error of mean and SD is the standard deviation. CD is the convergence diagnostic statistics proposed by Geweke (1992). IF is the inefficiency factor proposed by Chib (2001).

Table 5: Estimation result for Model 4 (t-error+SV)

	Mean	SE	SD	95% Interval	CD	IF
$\mu^{(0)}$	-0.4353	0.0082	0.1512	[-0.7875, -0.1837]	0.60	32.71
$\mu^{(1)}$	0.4026	0.0027	0.0614	[0.2883, 0.5314]	-0.01	23.26
ϕ	-0.1396	0.0014	0.0571	[-0.2486, -0.0256]	-1.52	5.84
p_{00}	0.9177	0.0005	0.0305	[0.8466, 0.9654]	0.99	4.99
p_{11}	0.9597	0.0007	0.0187	[0.9147, 0.9872]	-0.20	16.45
ω	-0.1821	0.0069	0.2254	[-0.6232, 0.2723]	-1.06	6.42
ψ	0.8540	0.0027	0.0443	[0.7522, 0.9274]	-0.76	33.22
σ_η^2	0.3737	0.0082	0.0996	[0.2220, 0.6025]	0.74	55.69
ν	30.2283	2.7244	18.7448	[7.3251, 80.6351]	0.39	116.53

SE is the standard error of mean and SD is the standard deviation. CD is the convergence diagnostic statistics proposed by Geweke (1992). IF is the inefficiency factor proposed by Chib (2001).

Table 6: Log marginal likelihood

Model 1	Model 2	Model 3	Model 4
-795.51	-783.90	-703.27	-710.98
(0.03)	(0.03)	(0.05)	(0.09)

Numbers in parentheses are standard errors.

Table 7: Business cycle turning points: Models 2–4

Peak				Trough			
ESRI	Model 2	Model 3	Model 4	ESRI	Model 2	Model 3	Model 4
85/06	85/10	85/11	85/08	86/11	86/11	86/09	86/09
91/02	90/11	90/11	90/11	93/10	94/01	94/01	94/01
97/05	97/07	97/07	97/07	99/01	98/11	98/11	98/12
00/10	01/01	01/01	01/01	02/01	02/01	02/01	02/01
08/02	07/09	07/09	07/07	09/03	09/04	09/03	09/03
12/03	12/04	12/04	12/04	12/11	12/11	12/11	12/12
–	–	–	15/08	–	–	–	16/03
18/10	18/06	18/06	18/05	20/05	20/06	20/05	20/05

Table 8: Business cycle turning points: Models 2–4

ESRI	Model 2	Model 3	Model 4
Peak			
85/06	85/10	85/11	85/08
91/02	90/11	90/11	90/11
97/05	97/07	97/07	97/07
00/10	01/01	01/01	01/01
08/02	07/09	07/09	07/07
12/03	12/04	12/04	12/04
–	–	–	15/08
18/10	18/06	18/06	18/05
Trough			
86/11	86/11	86/09	86/09
93/10	94/01	94/01	94/01
99/01	98/11	98/11	98/12
02/01	02/01	02/01	02/01
09/03	09/04	09/03	09/03
12/11	12/11	12/11	12/12
–	–	–	16/03
20/05	20/06	20/05	20/05

Table 9: Number of break points and log marginal likelihood

0	1	2	3	4
-795.51	-750.43	-690..85	-710.79	-712.33
(0.05)	(0.09)	(0.11)	(0.19)	(0.28)

Numbers in parentheses are standard errors.

Table 10: Estimation result for Model 5 (normal error + SV + 2 change points)

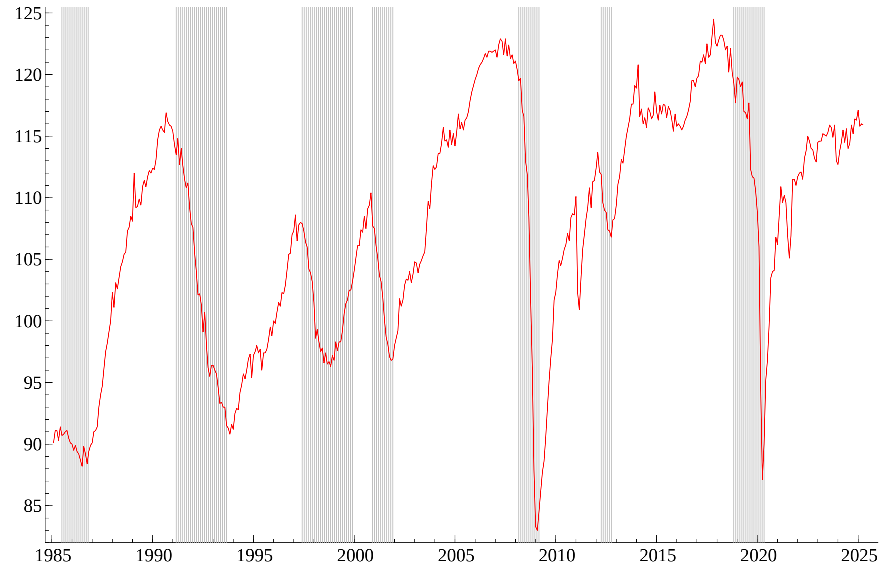
	Mean	SE	SD	95% Interval	CD	IF
$\mu^{(00)}$	-0.5372	0.0091	0.1516	[-0.8781, -0.2819]	-1.83	32.11
$\mu^{(01)}$	-5.5976	0.0338	1.2258	[-7.3781, -2.4430]	0.64	2.94
$\mu^{(02)}$	-0.2869	0.0359	0.9460	[-2.7238, 0.1630]	-0.66	3.46
$\mu^{(10)}$	0.4098	0.0021	0.0574	[0.2976, 0.5236]	0.62	9.04
$\mu^{(11)}$	1.8311	0.0146	0.4554	[1.2562, 2.3585]	0.86	7.90
$\mu^{(12)}$	0.4528	0.0181	0.2381	[0.1287, 0.8646]	-0.60	61.55
ϕ	-0.2050	0.0013	0.0575	[-0.3168, -0.0912]	-0.024	5.79
p_{00}	0.9366	0.0009	0.0284	[0.8697, 0.9788]	-1.73	13.77
p_{11}	0.9622	0.0005	0.0168	[0.9232, 0.9882]	-1.34	2.75
q_{00}	0.9962	0.0000	0.0036	[0.9864, 0.9999]	-0.48	1.40
q_{11}	0.9575	0.0006	0.0395	[0.8528, 0.9985]	1.17	1.15
ω	-0.3102	0.0030	0.1793	[-0.6625, 0.0449]	-0.89	3.42
ψ	0.8007	0.0035	0.0575	[0.6722, 0.8966]	-0.89	46.15
σ_η^2	0.4173	0.0088	0.1211	[0.2270, 0.6983]	0.98	57.44

SE is the standard error of mean and SD is the standard deviation. CD is the convergence diagnostic statistics proposed by Geweke (1992). IF is the inefficiency factor proposed by Chib (2001).

Table 11: Business cycle turning points: Model 5 (normal error + SV + 2 change points)

Peak		Trough	
ESRI	Model 5	ESRI	Model 5
85/06	85/11	86/11	86/09
91/02	90/11	93/10	94/01
97/05	97/07	99/01	98/11
00/10	01/01	02/01	02/01
08/02	07/11	09/03	09/03
12/03	12/04	12/11	12/10
–	15/02	–	16/06
18/10	18/02	20/05	20/06

Figure 1: CI



The coincident indicator of composite index published by Economic and Social Research Institute (ESRI), Cabinet Office, the Government of Japan. Shadow areas are the recession periods published by ESRI.

Figure 2: Growth rate of CI (%)

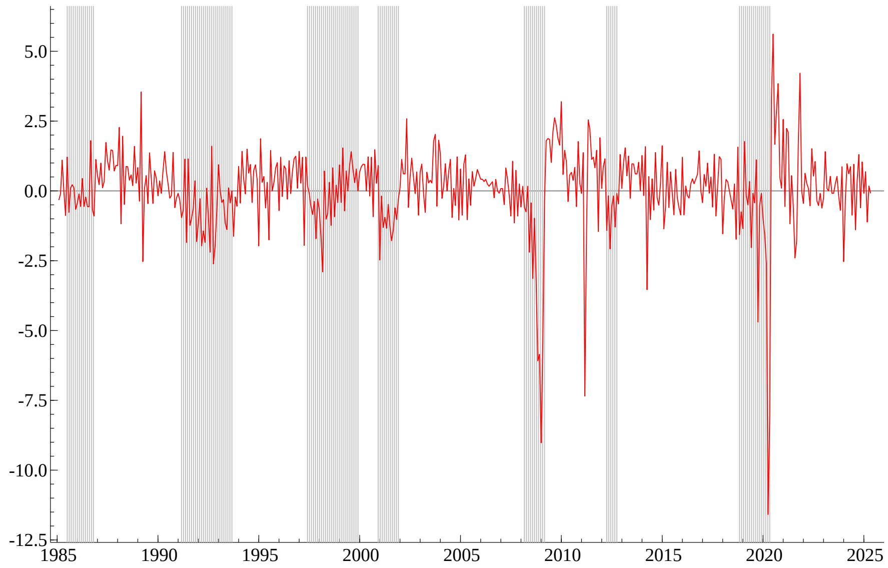
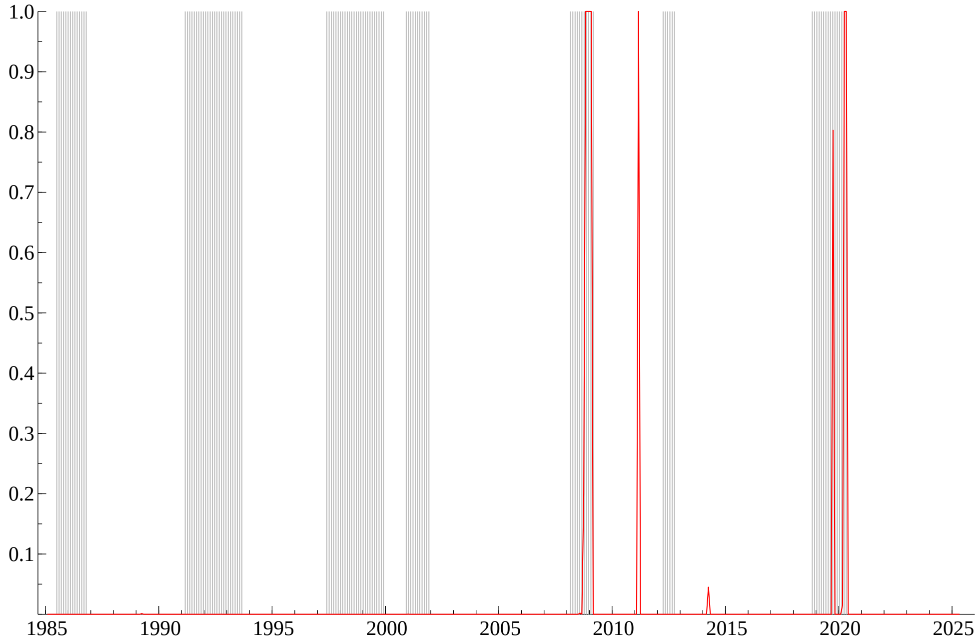


Figure 3: Posterior probabilities of recession: Model 1 (normal error + constant volatility)

Full Sample (1985/02–2025/05)



Subsample (1985/02–2008/08)

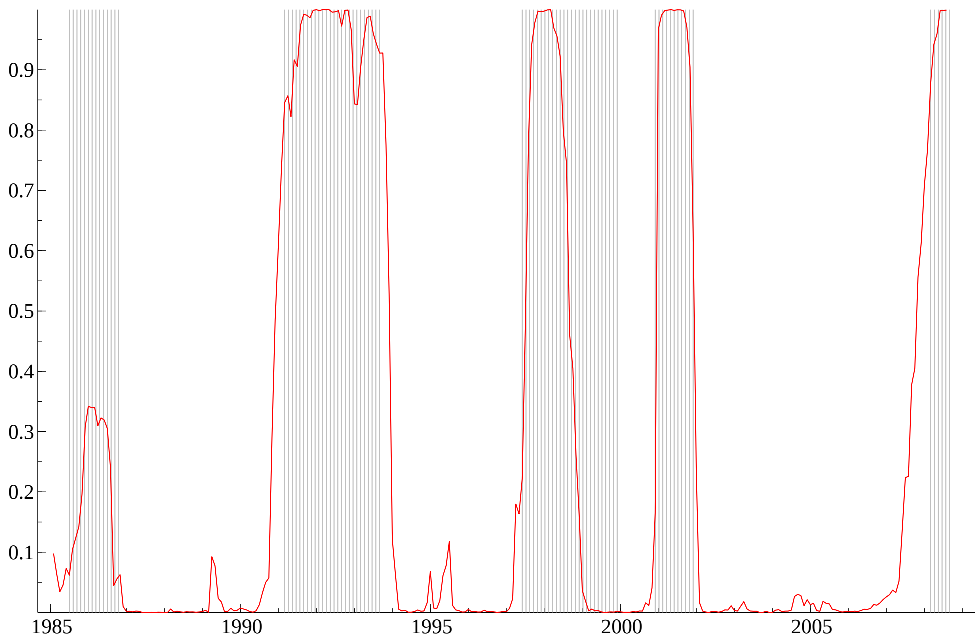


Figure 4: Posterior probabilities of recession: Model 2 (t-error + constant volatility)

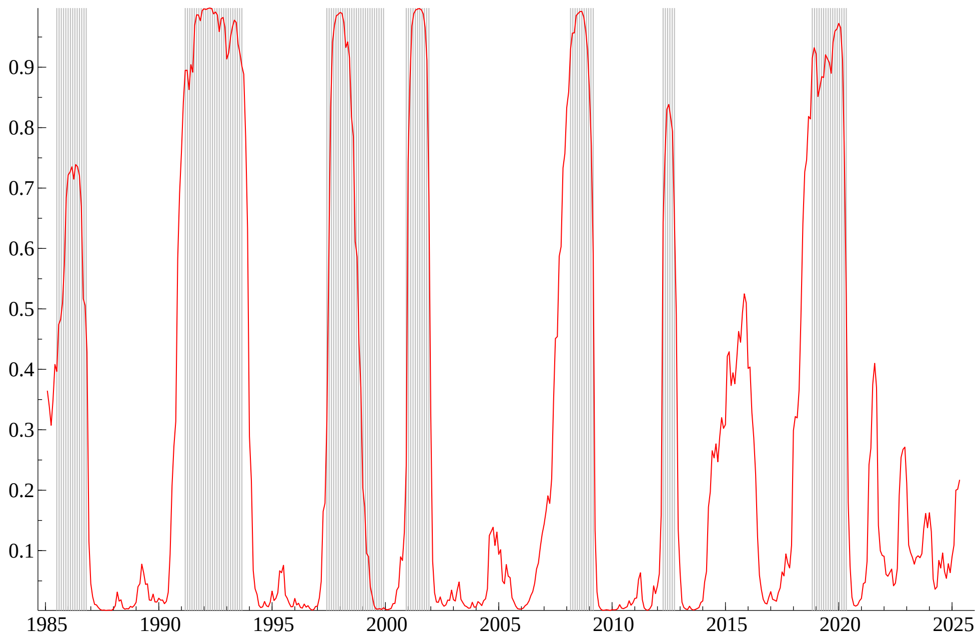


Figure 5: Posterior probabilities of recession: Model 3 (normal error + SV)

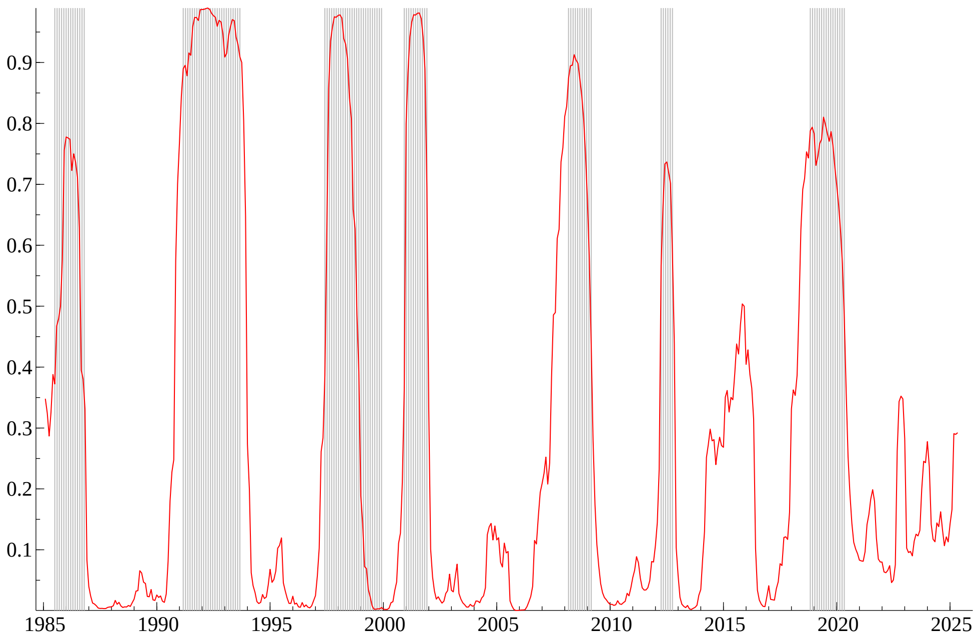


Figure 6: Posterior probabilities of recession: Model 4 (t-error + SV)

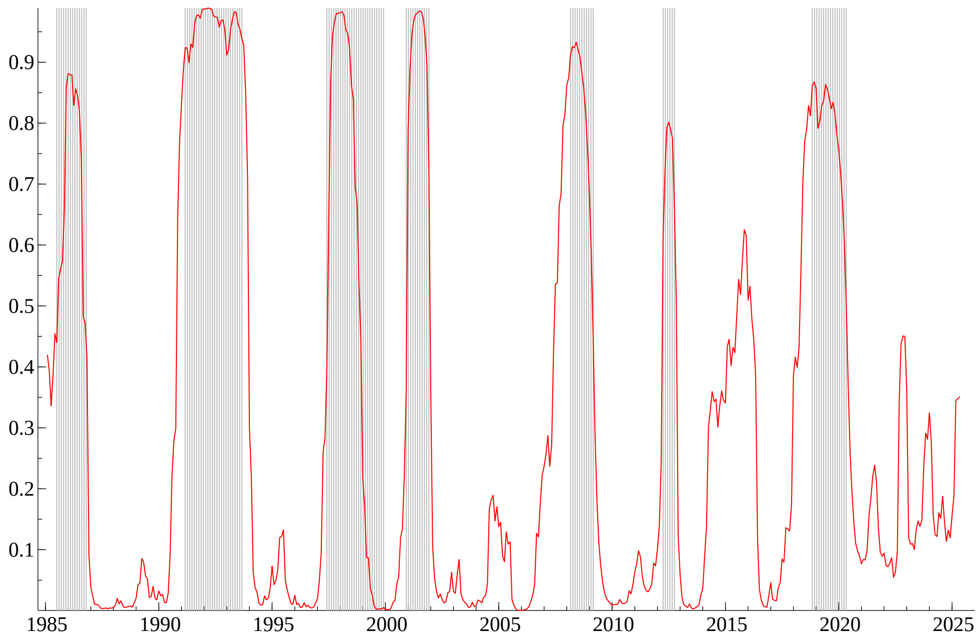
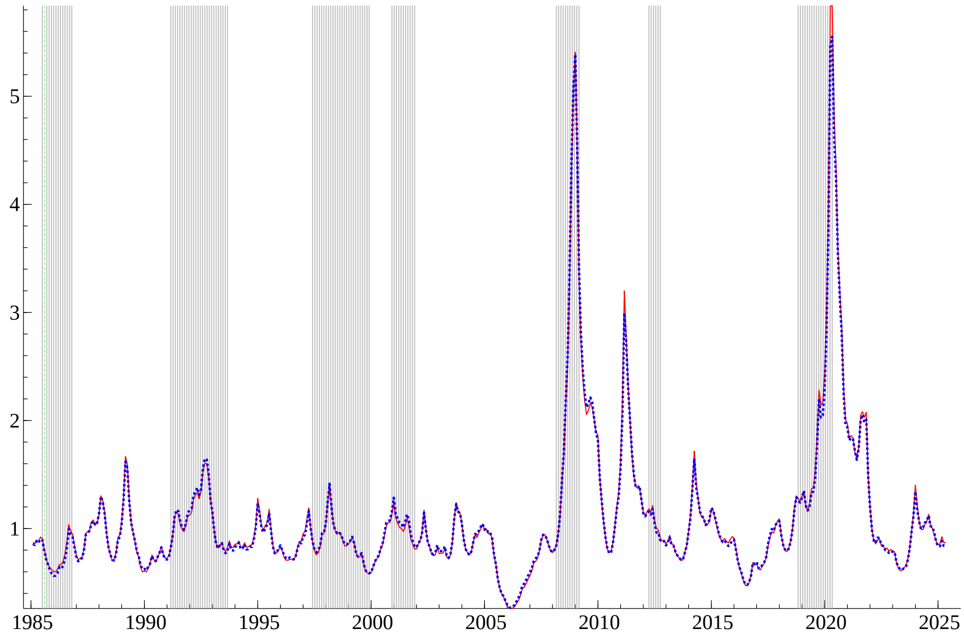
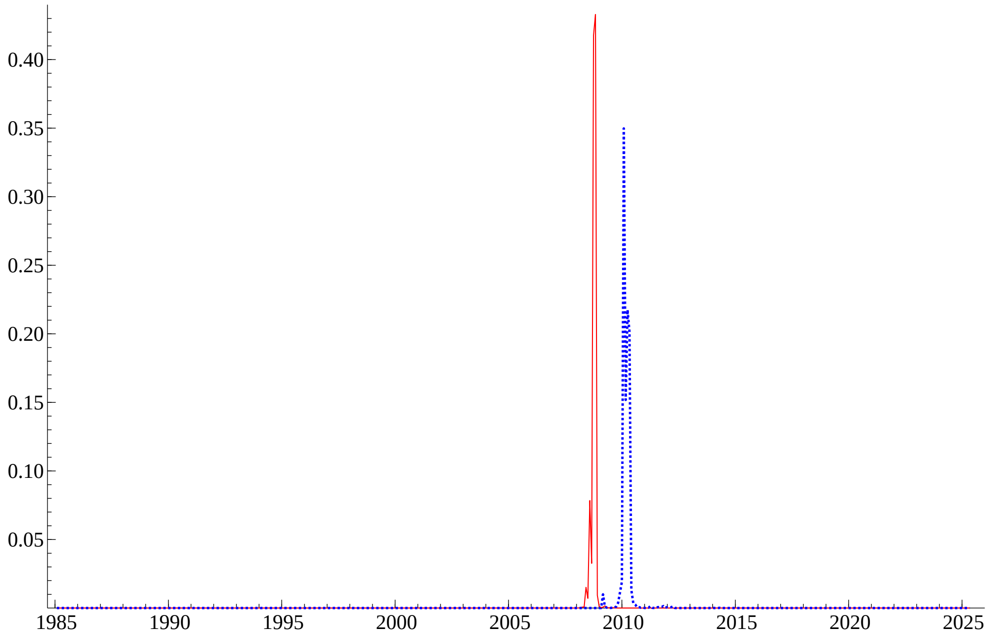


Figure 7: Posterior mean of volatility: Models 3 (normal error + SV) and Model 4 (t-error + SV)



The solid line is the posterior mean of volatility σ_t^2 from Model 3, and the dotted line is that from Model 4.

Figure 8: Posterior probabilities of change point: Model 5 (t-error+stochastic volatility+2 change points)



The solid line is the first change point and the dotted line is the second change point.

Figure 9: Posterior probabilities of recession: Model 5 (t-error + stochastic volatility+2 change points)