arXiv:2404.13986v1 [econ.EM] 22 Apr 2024

Stochastic Volatility in Mean: Efficient Analysis by a Generalized Mixture Sampler

Daichi Hiraki

Graduate School of Economics, University of Tokyo, Tokyo 113-0033, Japan

hdaichi397@gmail.com

Siddhartha Chib

Olin School of Business, Washington University, St Louis, USA

chib@wustl.edu

Yasuhiro Omori

Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan

omori@e.u-tokyo.ac.jp

April 2024

Abstract

In this paper we consider the simulation-based Bayesian analysis of stochastic volatility in mean (SVM) models. Extending the highly efficient Markov chain Monte Carlo mixture sampler for the SV model proposed in Kim et al. (1998) and Omori et al. (2007), we develop an accurate approximation of the non-central chi-squared distribution as a mixture of thirty normal distributions. Under this mixture representation, we sample the parameters and latent volatilities in one block. We also detail a correction of the small approximation error by using additional Metropolis-Hastings steps. The proposed method is extended to the SVM model with leverage. The methodology and models are applied to excess holding yields in empirical studies, and the SVM model with leverage is shown to outperform competing volatility models based on marginal likelihoods.

JEL classification: C11, C15, C32, C58

Keywords: Excess Holding Yield; GARCH in Mean; EGARCH-in Mean; Markov chain Monte Carlo; Mixture Sampler; Risk Premium; Stochastic Volatility in Mean

1 Introduction

In financial time series, volatility clustering, the phenomenon of persistent volatility, is well known to exist. One way to model time-varying volatility, or volatility clustering, is by using the stochastic volatility (SV) model of Taylor (2008). In the simplest version of this model, the standard deviation of the outcome is given by an exponential transformation of an unobserved log-variance variable h_t, where h_t in turn is modeled by a stationary first-order autoregressive process. The model in this basic form can be viewed as a state-space model in which the measurement equation is nonlinear in the latent variance h_t. A significant variant of this standard SV model is one in which the standard deviation of the outcome $\exp(h_t/2)$ appears as a predictor variable in the mean of the measurement equation. This is called the SV in mean model (SVM) and is similar in spirit to the ARCH in mean (ARCH-M) model introduced by Engle et al. (1987). Like the standard SV model, the SVM model has also been used in various fields, including macroeconomics and finance (see Koopman and Hol Uspensky (2002), Berument et al. (2009), Mumtaz and Zanetti (2013), Cross et al. (2023)).

In the Markov chain Monte Carlo (MCMC) estimation of model parameters for the SV-type models, it is often observed that sampling one latent variable conditional on all the other latent variables and parameters, which is referred to as the single-move sampler, is inefficient in the sense that MCMC draws are highly autocorrelated. To address this issue, Kim et al. (1998) introduced the mixture sampler as a highly efficient Bayesian estimation method for the standard SV model. This approach was extended to SV models with jumps and fat-tailed errors in Chib et al. (2002) and to SV models with leverage in Omori et al. (2007).

In the existing literature, the SVM model without leverage has been estimated by the multi-move (block) sampler (Shephard and Pitt (1997) and Omori and Watanabe (2008)), for example, Abanto-Valle et al. (2011), Abanto-Valle et al. (2012), and Leão et al. (2017), and by other approaches, for example, Chan (2017), Abanto-Valle et al. (2021), and Abanto-Valle et al. (2023). There is no known mixture sampler approach for SVM models with leverage. In this paper, we develop efficient MCMC based algorithms for SVM models, with and without leverage, that are based on accurate representations of these models in terms of mixtures of conditionally Gaussian linear state-space models, just as in the approach of Kim et al. (1998). However, due to the $\beta \exp(h_t/2)$ term in the mean equation, instead of characterizing

the distribution of a central chi-squared distribution with one degrees of freedom in terms of mixtures of normal distributions, a mixture representation to the distribution of a non-central chi-squared distribution with one degrees of freedom is needed, in which the non-centrality parameter is β^2. We show that the latter distribution has an infinite series expansion. Our estimation approach uses a truncated version of this series expansion to develop a highly efficient fitting algorithm. It can be viewed as a generalized mixture sampler. Furthermore, the small approximation error due to the truncation can be corrected by a data augmentation method by incorporating a pseudo-target probability density whose marginal probability density is the exact conditional posterior density.

We apply the proposed method to excess holding yields data, where we also consider alternative ways of introducing the log-variance in the mean function. We show that our proposed SVM model outperforms other competing volatility models such as the basic SV model, alternative SVM models, GARCH models, GARCH in mean models, EGARCH models and EGARCH-in mean models based on marginal likelihoods.

The rest of this paper is organized as follows. Section 2 introduces the SVM model and describes the novel mixture sampler as an efficient sampling method for such models. Section 3 illustrates the performance of this sampling method using the simulated data for several cases. Section 4 further extends the SVM model to incorporate the leverage effect. Finally, in Section 5, we apply our proposed SVM model to financial data and perform a comprehensive evaluation of the model. Conclusion and remarks are given in Section 6.

2 Stochastic volatility in mean model

2.1 SVM model

We define the stochastic volatility in mean (SVM) model as follows.

$$y_t = \beta \exp(h_t/2) + \epsilon_t \exp(h_t/2), \quad t = 1, ..., n, \tag{1}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 1, ..., n-1, \tag{2}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 \end{pmatrix}, \tag{3}$$

$$h_1 \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \tag{4}$$

where $N(m, S)$ denotes normal distribution with mean vector m and covariance matrix S, $\theta = (\mu, \phi, \sigma^2, \beta)$ is a model parameter vector of interest, and $h = (h_1, \ldots, h_n)'$ is the logarithm of the latent volatility vector. We use $\beta \exp(h_t/2)$ rather than $\beta \exp(h_t)$ in the mean equation since β in our specification is not affected by the change of measurement units. For $\beta \neq 0$, we denote it as the stochastic volatility in mean (SVM) model. The standard stochastic volatility (SV) model is obtained as a special case with $\beta = 0$. For θ, we assume the prior distributions

$$\mu \sim N(\mu_0, \sigma_0^2), \quad \frac{\phi + 1}{2} \sim Beta(a, b),$$
$$\sigma^2 \sim IG\left(\frac{n_0}{2}, \frac{s_0}{2}\right), \quad \beta \sim N(b_0, B_0).$$

where $Beta(a, b)$ denotes beta distribution with parameters (a, b) and $IG(a, b)$ denotes inverse gamma distribution with parameters (a, b) whose probability density function is

$$\pi(x|a, b) \propto x^{-(a+1)} \exp(-bx), \quad x > 0, \quad a, b > 0.$$

We let $f(y, h|\theta)$ and $\pi(\theta)$ denote the probability density function of (y, h) given θ and the prior probability density function of θ where $y \equiv (y_1, \ldots, y_n)'$. The posterior density function of (h, θ) is given in Appendix A.

Remark. It is straightforward to include the constant term in the measurement equation. However, noting that $\beta \times \exp(h_t/2) \approx \beta \times (1 + h_t/2)$, it is often confounded with β and therefore omitted in this paper.

2.2 Transformation of the measurement equation

To sample h from its conditional distribution, we transform Equation (1) as below:

$$y_t^* = h_t + \epsilon_t^*, \quad y_t^* = \log(y_t^2), \quad \epsilon_t^* = \log(\beta + \epsilon_t)^2, \tag{5}$$

Since $(\beta + \epsilon_t) \sim N(\beta, 1)$, its square $(\beta + \epsilon_t)^2 \sim \chi_1^2(\beta^2)$ where $\chi_1^2(\beta^2)$ denotes the non-central chi-square distribution with the non-centrality parameter β^2 and one degrees-of-freedom. The special case with $\beta = 0$ is considered in Kim et al. (1998) and Omori et al. (2007), who introduced the idea of accurately approximating the probability of the logarithm of the central chi-square distribution with one degrees of freedom, $\log \chi_1^2(0)$,

$$f(\epsilon_t^*) = \frac{1}{\sqrt{2\pi}} \left(\frac{\epsilon_t^* - \exp(\epsilon_t^*)}{2}\right), \quad -\infty < \epsilon_t^* < \infty,$$

by a mixture of normal distributions. Below, we elaborate a highly accurate approximation of the distribution of ϵ_t^* given $\beta \neq 0$ by the mixture of normal distributions. Let $p(x;\nu,\lambda)$ be the probability density function of $\chi^2_\nu(\lambda)$. It can be expressed as an infinite mixture of central χ^2 probability density functions (see, e.g. Johnson et al. (1995)):

$$\begin{aligned} p(x;\nu,\lambda) &= \sum_{j=0}^{\infty} \left\{ \frac{\left(\frac{\lambda}{2}\right)^j}{j!} \exp\left(-\frac{\lambda}{2}\right) \right\} p(x;\nu+2j,0), \\ & \qquad p(x;\nu+2j,0) = \frac{x^{\frac{\nu}{2}+j-1}}{2^{\frac{\nu}{2}+j}\Gamma(\frac{\nu}{2}+j)} \exp\left(-\frac{x}{2}\right). \end{aligned}$$

Setting $\nu = 1$ and noting that

$$p(x;1+2j,0) = \frac{x^j \Gamma\left(\frac{1}{2}\right)}{2^j \Gamma\left(\frac{1}{2}+j\right)} \times p(x;1,0),$$

we obtain the expression

$$p(x;1,\lambda) = \sum_{j=0}^{\infty} \left\{ \frac{\left(\frac{\lambda}{2}\right)^j}{j!} \exp\left(-\frac{\lambda}{2}\right) \right\} \frac{x^j \Gamma\left(\frac{1}{2}\right)}{2^j \Gamma\left(\frac{1}{2}+j\right)} \times p(x;1,0). \tag{6}$$

Let $f(u;\lambda)$ denote the probability density function of $U \sim \log \chi^2_1(\lambda)$. Using (6), it follows that

$$f(u;\lambda) = \sum_{j=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^j}{j!} \exp\left(-\frac{\lambda}{2}\right) \frac{\exp(uj)\Gamma\left(\frac{1}{2}\right)}{2^j \Gamma\left(\frac{1}{2}+j\right)} \times f(u;0). \tag{7}$$

As in Omori et al. (2007), we consider the mixture of ten normal distributions to approximate $f(u;0)$, the probability density function of $\log \chi^2_1(0)$,

$$f(u;0) \approx \sum_{i=1}^{K} p_i v_i^{-1} \phi\left(\frac{u-m_i}{v_i}\right), \quad K = 10, \tag{8}$$

where $\phi(\cdot)$ denotes the probability density function of the standard normal distribution. The values of (p_i, m_i, v_i^2) are taken from Omori et al. (2007) and are reproduced in Table 1. The columns labeled a_i and b_i will be used when we consider the model with leverage in Section 5.

i	p_i	m_i	v_i^2	a_i	b_i
1	0.00609	1.92677	0.11265	1.01418	0.50710
2	0.04775	1.34744	0.17788	1.02248	0.51124
3	0.13057	0.73504	0.26768	1.03403	0.51701
4	0.20674	0.02266	0.40611	1.05207	0.52604
5	0.22715	-0.85173	0.62699	1.08153	0.54076
6	0.18842	-1.97278	0.98583	1.13114	0.56557
7	0.12047	-3.46788	1.57469	1.21754	0.60877
8	0.05591	-5.55246	2.54498	1.37454	0.68728
9	0.01575	-8.68384	4.16591	1.68327	0.84163
10	0.00115	-14.65000	7.33342	2.50097	1.25049

Table 1: Selection of $(p_i, m_i, v_i^2, a_i, b_i)$ introduced in Omori et al. (2007).

By substituting Equation (8) to Equation (7), we obtain

$$\begin{aligned}
f(u;\lambda) &\approx \sum_{j=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^j}{j!} \exp\left(-\frac{\lambda}{2}\right) \frac{\exp(uj)\Gamma\left(\frac{1}{2}\right)}{2^j\Gamma\left(\frac{1}{2}+j\right)} \sum_{i=1}^{K} p_i v_i^{-1} \phi\left(\frac{u-m_i}{v_i}\right) \\
&= \sum_{i=1}^{K}\sum_{j=0}^{\infty} p_i \exp\left(-\frac{\lambda}{2}\right) \frac{\Gamma\left(\frac{1}{2}\right)}{2^j j!\Gamma\left(\frac{1}{2}+j\right)} \left(\frac{\lambda}{2}\right)^j \exp\left(m_i j + \frac{j^2 v_i^2}{2}\right) \\
&\qquad \times \frac{1}{\sqrt{2\pi v_i^2}} \exp\left\{-\frac{(u-(m_i+jv_i^2))^2}{2v_i^2}\right\} \\
&= \sum_{i=1}^{K}\sum_{j=0}^{\infty} p_i \exp\left(-\frac{\lambda}{2} + m_i j + \frac{j^2 v_i^2}{2}\right) \frac{\Gamma\left(\frac{1}{2}\right)}{2^j j!\Gamma\left(\frac{1}{2}+j\right)} \left(\frac{\lambda}{2}\right)^j \\
&\qquad \times v_i^{-1} \phi\left(\frac{u-(m_i+jv_i^2)}{v_i}\right) \\
&\approx \sum_{i=1}^{K}\sum_{j=0}^{J} \tilde{p}_{i,j} v_i^{-1} \phi\left(\frac{u-\tilde{m}_{i,j}}{v_i}\right), \qquad (9)
\end{aligned}$$

where

$$\tilde{p}_{i,j} = \frac{w_{i,j}}{\sum_{i=1}^{K}\sum_{j=0}^{J} w_{i,j}}, \quad w_{i,j} = p_i \exp\left(-\frac{\lambda}{2} + m_i j + \frac{j^2 v_i^2}{2}\right) \frac{\Gamma\left(\frac{1}{2}\right)}{2^j j!\Gamma\left(\frac{1}{2}+j\right)} \left(\frac{\lambda}{2}\right)^j,$$

and $\tilde{m}_{i,j} = m_i + jv_i^2$. In the last equality, we truncate the summation at $j = J$ and normalize $\tilde{p}_{i,j}$ to ensure that $\sum_{i=1}^{K}\sum_{j=0}^{J} \tilde{p}_{i,j} = 1$.

This expression implies that the probability density function of $\log \chi_1^2(\lambda)$ is approximated by the mixture of $K(J+1)$ normal distributions. Especially, when $\lambda = 0$ and $J = 0$,

the approximation (9) reduces to (8). The extent of the impact on the approximation is based on the value $w_{i,j}$ which includes $\lambda = \beta^2$. The coefficient β of volatility $\exp(h_t/2)$ is often estimated to be $\hat{\beta} = 0.1 \sim 0.2$ in past empirical studies for GARCH-M (generalized autoregressive conditional heteroscedasticity in mean) models. When $\lambda = \beta^2$ is less than 1.0, $J = 1$ or 2 makes $\sum_{i=1}^{K} \sum_{j=0}^{J} w_{i,j}$ more than 0.9. For $J = 2$, Figures 1 and 2 show the true and approximate densities of $\log \chi_1^2(\beta^2)$ for $\beta = 0.3, 0.5$ and 0.7 (equivalently, $\beta = -0.3, -0.5$ and -0.7) and the difference between the two densities, respectively. Since these differences are quite small and the approximation almost overlaps the true probability density of $\log \chi_1^2(\beta^2)$ even for $\beta = 0.7$, we employ $J = 2$ in this paper. That is, we approximate the probability density of $\epsilon_t^* | \beta \sim \log \chi_1^2(\beta^2)$ by

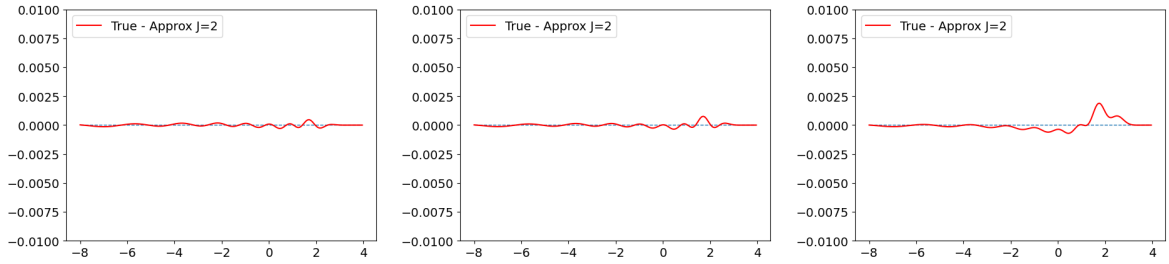
$$f(\epsilon_t^* | \beta) \approx \sum_{i=1}^{10} \sum_{j=0}^{2} \tilde{p}_{i,j} v_i^{-1} \phi \left(\frac{u - \tilde{m}_{i,j}}{v_i} \right), \tag{10}$$

where

$$\tilde{p}_{i,j} = \frac{p_i \exp\left(m_i j + \frac{j^2 v_i^2}{2} \right) \frac{1}{2^j j! \Gamma(1/2+j)} \left(\frac{\beta^2}{2} \right)^j}{\sum_{i=1}^{10} \sum_{j=0}^{2} p_i \exp\left(m_i j + \frac{j^2 v_i^2}{2} \right) \frac{1}{2^j j! \Gamma(1/2+j)} \left(\frac{\beta^2}{2} \right)^j}, \quad \tilde{m}_{i,j} = m_i + j v_i^2.$$

(a) $\beta = 0.3$ (b) $\beta = 0.5$ (c) $\beta = 0.7$

Figure 1: True and approximation densities of $\log \chi_1^2(\beta^2)$ for $\beta = 0.3, 0.5$ and 0.7.



(a) $\beta = 0.3$ (b) $\beta = 0.5$ (c) $\beta = 0.7$

Figure 2: Differences between the true density of $\log \chi_1^2(\beta^2)$ and the approximate densities for $\beta = 0.3, 0.5$ and 0.7.

Let $s_t = (s_{1t}, s_{2t}) \in \{(i,j)|i = 1, ..., K, j = 0, ..., J\}$ denote the component of the mixture of the normal densities in (10) at time t. Given $s_t = (i,j)$, we have $\epsilon_t^*|s_t = (i,j) \sim N(\tilde{m}_{i,j}, v_i^2)$ and we see that the SVM model can be approximated by the linear Gaussian state space form

$$y_t^* = \tilde{m}_{s_{1t},s_{2t}} + h_t + (v_{s_{1t}}, 0)z_t, \tag{11}$$

$$\begin{aligned} h_{t+1} &= \mu(1-\phi) + \phi h_t + (0,\sigma)z_t, \\ t &= 1, ..., n-1, \quad h_1 \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \quad |\phi| < 1, \\ z_t &= (z_{1t}, z_{2t})' \sim N(0, I_2), \end{aligned} \tag{12}$$

where $y_t^* = \log(y_t^2)$. In the following sections, $\tilde{m}_{s_{1t},s_{2t}}$ and $\tilde{p}_{s_{1t},s_{2t}}$ are abbreviated as $\tilde{m}_{s_t}$ and $\tilde{p}_{s_t}$, and we write $v_{s_t}^2$ instead of $v_{s_{1t}}^2$.

3 MCMC simulation and associated particle filter

3.1 MCMC algorithm

Algorithm 1. Let us denote $\theta = (\alpha, \beta)$ where $\alpha = (\mu, \phi, \sigma^2)$. The Markov chain Monte Carlo simulation is implemented in four blocks:

1. Initialize h and $\theta = (\alpha, \beta)$.
2. Generate $\beta|\alpha, h, y \sim \pi(\beta|\alpha, h, y)$.
3. Generate $(\alpha, h)|\beta, y \sim \pi(\alpha, h|\beta, y)$.
4. Go to Step 2.

Step 2. Generation of $\beta|\alpha, h, y$

The conditional posterior distribution of β is normal with mean b_1 and variance B_1 where

$$b_1 = B_1\left(X'\Omega^{-1}y + B_0^{-1}b_0\right), \quad B_1^{-1} = X'\Omega^{-1}X + B_0^{-1},$$

and

$$X = \begin{pmatrix} \exp(h_1/2) \\ \exp(h_2/2) \\ \vdots \\ \exp(h_n/2) \end{pmatrix}, \quad \Omega = \mathrm{diag}\left(\exp(h_1), \exp(h_2), \cdots, \exp(h_n)\right).$$

Thus we generate $\beta \sim N(b_1, B_1)$.

Step 3. Generation of $(\alpha, h)|\beta, y$

As discussed in the previous section, we sample h using the mixture sampler using the mixture of normal distributions. Since our approximation is highly accurate, we incorporate this mixture approximation in Step 3 without correcting the approximation error as in Step 2 of Algorithm 1 in Chib et al. (2002). The additional step to correct for the approximation error is described in Algorithm 2 in the Appendix A. Let $f_N(\cdot|m, s^2)$ denote the probability density of $N(m, s^2)$, and let $\pi(\alpha)$ denote the prior density of α. Define our target density in Step 3 as

$$\begin{aligned} \pi^*(\alpha, h, s|\beta, y) &= \pi^*(\alpha, h|\beta, s, y) \times q(s), \\ &= \pi^*(h|\alpha, \beta, s, y)\pi^*(\alpha|\beta, s, y) \times q(s), \\ q(s) &= \prod_{t=1}^{n} \tilde{p}_{s_t}, \quad y_t^* = \log(y_t^2), \end{aligned}$$

where

$$\pi^*(h|\alpha, s, \beta, y^*) = \frac{\prod_{t=1}^n g(y_t^*|h_t, \alpha, \beta, s_t)}{m(y^*|\alpha, s, \beta)} \times \prod_{t=1}^{n-1} f_N(h_{t+1}|\mu(1-\phi) + \phi h_t, \sigma^2) \times f_N\left(h_1\middle|\mu, \frac{\sigma^2}{1-\phi^2}\right).$$

$$\pi^*(\alpha|s, \beta, y^*) \propto m(y^*|\alpha, s, \beta)\pi(\alpha),$$

$$g(y_t^*|h_t, \alpha, \beta, s_t) = f_N(y_t^*|\tilde{m}_{s_t} + h_t, v_{s_t}^2), \quad t = 1, \ldots, n,$$

and $\tilde{m}_{s_t} = \tilde{m}_{s_{1t}, s_{2t}}$ and $\tilde{p}_{s_t} = \tilde{p}_{s_{1t}, s_{2t}}$ are defined in (10) and $(p_{s_{1t}}, m_{s_{1t}}, v_{s_{1t}}^2)$ are given in Table 1. Note that $m(y^*|\alpha, s, \beta)$ is a normalizing constant for $\pi^*(h|\alpha, s, \beta, y^*)$ and is evaluated using

the Kalman filter algorithm. Only $\tilde{p}_{s_t}$, which depends on β, needs to be updated according to the formula in (10) before sampling. Note that the target density $\pi^*(\alpha, h|\beta, y)$ approximates the true conditional density $\pi(\alpha, h|\beta, y)$ accurately. We generate the sample (α, h, s) in two steps.

(a) Generate $s \sim q(s|h, \alpha, \beta, y^*)$ where

$$q(s|h, \alpha, \beta, y^*) = \prod_{t=1}^{n} \frac{\tilde{p}_{s_t} g(y_t^*|h_t, \alpha, \beta, s_t)}{\sum_{i=1}^{10} \sum_{j=0}^{2} \tilde{p}_{i,j} g(y_t^*|h_t, \alpha, \beta, s_t = (i, j))}.$$

(b) Generate $(\alpha, h) \sim \pi^*(\alpha, h|s, \beta, y)$

(i) Generate $\alpha \sim \pi^*(\alpha|s, \beta, y^*)$. We first transform α to $\vartheta = (\mu, \log\{(1+\phi)/(1-\phi)\}, \log\sigma^2)$ to remove parameter constraints and perform the Metropolis-Hastings (MH) algorithm (Chib and Greenberg, 1995) to sample from the conditional posterior distribution with density $\pi^*(\vartheta|s, \beta, y) = \pi^*(\alpha|s, \beta, y) \times |d\alpha/d\vartheta|$ where $|d\alpha/d\vartheta|$ is the Jacobian of the transformation. Compute the posterior mode $\hat{\vartheta}$ and define ϑ_* and Σ_* as

$$\vartheta_* = \hat{\vartheta}, \quad \Sigma_*^{-1} = -\frac{\partial^2 \log \pi^*(\vartheta|s, \beta, y)}{\partial\vartheta\partial\vartheta'}\bigg|_{\vartheta=\hat{\vartheta}}.$$

Given the current value ϑ, generate a candidate $\vartheta^\dagger$ from the distribution $N(\vartheta_*, \Sigma_*)$ and accept it with probability

$$\alpha(\vartheta, \vartheta^\dagger|s, \beta, y) = \min\left\{1, \frac{\pi^*(\vartheta^\dagger|s, \beta, y) f_N(\vartheta|\vartheta_*, \Sigma_*)}{\pi^*(\vartheta|s, \beta, y) f_N(\vartheta^\dagger|\vartheta_*, \Sigma_*)}\right\},$$

where $f_N(\cdot|\vartheta_*, \Sigma_*)$ is the probability density of $N(\vartheta_*, \Sigma_*)$. If the candidate $\vartheta^\dagger$ is rejected, we take the current value ϑ as the next draw. When the Hessian matrix is not negative definite, we may take a flat normal proposal $N(\vartheta_*, c_0 I)$ using some large constant c_0.

(ii) Generate $h|\alpha, s, \beta, y \sim \pi^*(h|\alpha, s, \beta, y)$. We generate $h = (h_1, ..., h_n)$ using a simulation smoother introduced by de Jong and Shephard (1995) and Durbin and Koopman (2002) for the linear Gaussian state space model as in (14)-(16).

3.2 Associated particle filter

We describe how to compute the likelihood $f(y|\theta)$

$$f(y|\theta) = \int f(y, h|\theta) dh,$$

numerically as it is necessary to obtain the marginal likelihood, $f(y) = \int f(y|\theta)\pi(\theta)d\theta$ and Bayes factor for the model comparison. The filtering and the associated particle computations are carried out by the auxiliary particle filter (see e.g. Pitt and Shephard (1999), Omori et al. (2007)). Let us denote $Y_t = (y_1, \ldots, y_n)$, and

$$\begin{aligned} f(y_t|h_t, \theta) &= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}h_t - \frac{1}{2}\{y_t - \beta\exp(h_t/2)\}^2\exp(-h_t)\right] \\ f(h_{t+1}|h_t, y_t, \theta) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma}\exp\left\{-\frac{(h_{t+1}-\mu_{t+1})^2}{2\sigma^2}\right\}, \\ \mu_{t+1} &= \mu + \phi(h_t - \mu), \end{aligned}$$

and consider the importance function for the auxiliary particle filter

$$\begin{aligned} q(h_{t+1}, h_t^i|Y_{t+1}, \theta) &\propto f(y_{t+1}|\mu_{t+1}^i, \theta) f(h_{t+1}|h_t^i, y_t, \theta)\hat{f}(h_t^i|Y_t, \theta) \\ &\propto f(h_{t+1}|h_t^i, y_t, \theta) q(h_t^i|Y_{t+1}, \theta) \end{aligned}$$

where

$$\begin{aligned} q(h_t^i|Y_{t+1}, \theta) &= \frac{f(y_{t+1}|\mu_{t+1}^i, \theta)\hat{f}(h_t^i|Y_t, \theta)}{\sum_{j=1}^I f(y_{t+1}|\mu_{t+1}^j, \theta)\hat{f}(h_t^j|Y_t, \theta)}, \\ f(y_{t+1}|\mu_{t+1}^i, \theta) &= \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\mu_{t+1}^i - \frac{1}{2}\{y_t - \beta\exp(h_t^i/2)\}^2\exp(-\mu_{t+1}^i)\right], \\ \mu_{t+1}^i &= \mu + \phi(h_t^i - \mu). \end{aligned}$$

This leads to the following particle filtering.

1. Compute $\hat{f}(y_1|\theta)$ and $\hat{f}(h_1^i|Y_1, \theta) = \pi_1^i$ for $i = 1, \ldots, I$.

   (a) Generate $h_1^i \sim f(h_1|\theta)$ $(= N(\mu, \sigma^2/(1-\phi^2)))$ for $i = 1, \ldots, I$.

   (b) Compute

   $$\begin{aligned} \pi_1^i &= \frac{w_i}{\sum_{i=1}^I w_i}, \quad w_i = f(y_1|h_1, \theta), \quad W_i = F(y_1|h_1, \theta), \\ \hat{f}(y_1|\theta) &= \overline{w}_1 = \frac{1}{I}\sum_{i=1}^I w_i, \quad \hat{F}(y_1|\theta) = \overline{W}_1 = \frac{1}{I}\sum_{i=1}^I W_i, \end{aligned}$$

   where $f(y_1|\theta)$ and $F(y_1|\theta)$ are the marginal density function and the marginal distribution function of y_1 given θ. Let $t = 1$.

2. Compute $\hat{f}(y_{t+1}|\theta)$ and $\hat{f}(h_{t+1}^i|Y_{t+1}, \theta) = \pi_{t+1}^i$ for $i = 1, \ldots, I$.

(a) Sample $h_t^i \sim q(h_t|Y_t,\theta)$, $i=1,\ldots,I$.

(b) Generate $h_{t+1}^i|h_t^i,y_t,\theta \sim f(h_{t+1}|h_t^i,y_t,\theta)$ $(=N(\mu_{t+1}^i,\sigma^2))$ for $i=1,\ldots,I$.

(c) Compute

$$\pi_{t+1}^i = \frac{w_i}{\sum_{i=1}^I w_i}, \quad w_i = \frac{f(y_{t+1}|h_{t+1}^i,\theta)f(h_{t+1}^i|h_t^i,y_t,\theta)\hat{f}(h_t^i|Y_t,\theta)}{f(h_{t+1}^i|h_t^i,y_t,\theta)q(h_t^i|Y_{t+1},\theta)} = \frac{f(y_{t+1}|h_{t+1}^i,\theta)\hat{f}(h_t^i|Y_t,\theta)}{q(h_t^i|Y_{t+1},\theta)},$$

$$W_i = \frac{F(y_{t+1}|h_{t+1}^i,\theta)\hat{f}(h_t^i|Y_t,\theta)}{q(h_t^i|Y_{t+1},\theta)},$$

$$\hat{f}(y_{t+1}|Y_t,\theta) = \overline{w}_{t+1} = \frac{1}{I}\sum_{i=1}^I w_i, \quad \hat{F}(y_{t+1}|\theta) = \overline{W}_{t+1} = \frac{1}{I}\sum_{i=1}^I W_i.$$

3. Increment t and go to 2.

It can be shown that as $I\to\infty$, $\overline{w}_{t+1} \xrightarrow{p} f(y_{t+1}|Y_t,\theta)$, and $\overline{W}_{t+1} \xrightarrow{p} F(y_{t+1}|Y_t,\theta)$. Therefore, it follows that

$$\sum_{t=1}^n \log \overline{w}_t \xrightarrow{p} \sum_{t=1}^n \log f(y_t|y_1,\ldots,y_{t-1},\theta),$$

is a consistent estimate of the conditional log-likelihood and can be used as an input in the calculation of the marginal likelihood by the method of Chib (1995).

4 Illustrative numerical examples

This section illustrates our proposed estimation method using the simulated data. We generate y_t $(t=1,\ldots,1000)$ by setting

$$\phi = 0.97, \quad \mu = 0, \quad \sigma = 0.3.$$

To avoid the case $y_t=0$ which leads to $\log(y_t^2)=-\infty$, we introduce very small value c and use $y_t^* = \log(y_t^2+c)$. We set c equal to 1.0×10^{-7}. For β, we consider three cases $\beta = 0.3, 0.5$ and 0.7 to investigate the effect of the approximation error. The common random numbers are used to generate y_t's for three cases. In these simulation studies, we specify the prior as

$$\mu \sim N(0,1000^2), \quad \frac{\phi+1}{2} \sim Beta(1,1),$$
$$\sigma^2 \sim IG\left(\frac{0.001}{2}, \quad \frac{0.001}{2}\right), \quad \beta \sim N(0,1).$$

The prior on $(\phi+1)/2$ is set to ensure the stationarity of the latent volatility process. We iterated MCMC simulation 50,000 times after discarding initial 10,000 MCMC draws as burn-in period.

(i) Case $\beta = 0.3$. The acceptance rates of the MH algorithms for α is 73.6%. The sample paths are given in Figure 3, and the MCMC chain mixes quite well.

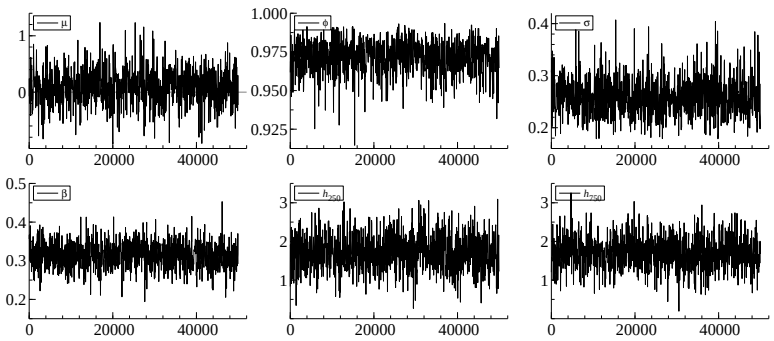


Figure 3: Sample paths for θ, h_{250}, and h_{750}. $\beta = 0.3$.

The sample autocorrelation functions are shown in Figure 4 and they decay very quickly.

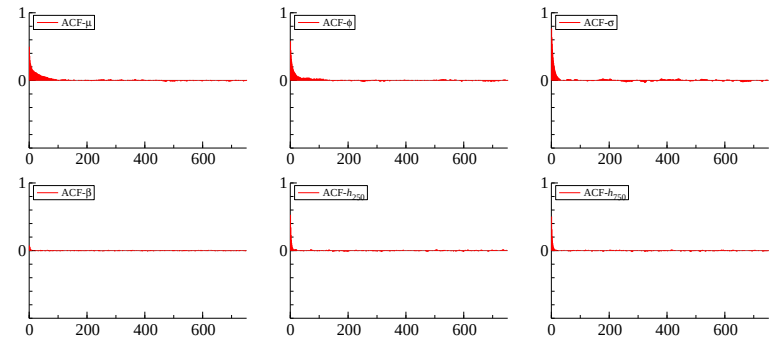


Figure 4: Sample autocorrelation functions for θ, h_{250}, and h_{750}. $\beta = 0.3$.

Par	True	Mean	Std Dev	95% interval	IF
μ	0	0.090	0.319	(-0.575, 0.725)	17
ϕ	0.97	0.971	0.011	(0.948, 0.989)	15
σ	0.3	0.259	0.037	(0.195, 0.338)	13
β	0.3	0.315	0.033	(0.251, 0.380)	1
h_{250}	2.310	1.735	0.460	(0.843, 2.647)	5
h_{750}	2.077	1.674	0.412	(0.900, 2.520)	5

Table 2: True values, posterior means, posterior standard deviations, 95% credible intervals, and inefficiency factors. $\beta = 0.3$.

Table 2 shows the posterior mean, 95% credible intervals and inefficiency factors (IF). The estimated parameters are close to true values. IF is calculated by $1 + 2\sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s. This is interpreted as the ratio of the numerical variance of the posterior mean from the chain to the variance of the posterior mean from hypothetical uncorrelated draws. They are overall small as expected, which means that the MCMC sampling is close to the uncorrelated sampling. Note that those IF's for h_{250} and h_{750} are quite small, which suggests the use of mixture sampler for the MH algorithm for h is highly efficient. Finally Figure 5 shows true values, 95% credible intervals and the posterior medians or volatilities. The estimated smoothed values follow the true values values that are almost covered by 95% intervals, indicating that MCMC estimations works well.

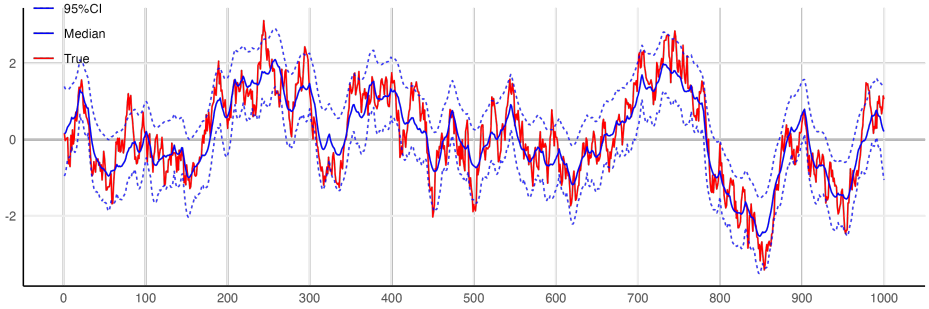


Figure 5: Log volatilities: True values, 95% credible intervals and posterior median.

(ii) Case $\beta = 0.5$. The acceptance rates of the MH algorithms for α is 73.6%. The plot of the sample paths and log volatilities are similar to those in (i) and hence omitted to save space. Table 3 shows the posterior means, 95% credible intervals and inefficiency factors. The estimated parameters are close to true values, and inefficiency factors (IF) are overall small as in (i). The IF's for h_{250} and h_{750} are sufficiently small, and indicates that the algorithm is still highly efficient.

Par	True	Mean	Std Dev	95% interval	IF
μ	0	0.102	0.319	(-0.562, 0.754)	18
ϕ	0.97	0.971	0.011	(0.947, 0.988)	13
σ	0.3	0.263	0.037	(0.198, 0.342)	13
β	0.5	0.511	0.035	(0.443, 0.579)	3
h_{250}	2.310	1.811	0.459	(0.928, 2.719)	4
h_{750}	2.077	1.641	0.410	(0.866, 2.476)	4

Table 3: True values, posterior means, posterior standard deviations, 95% credible intervals, and inefficiency factors. $\beta = 0.5$.

(iii) Case $\beta = 0.7$. The acceptance rates of the MH algorithms for α is 73.5%. The convergence seems to become slightly slower, but the chain mixes well. The plot of the sample paths and log volatilities are similar to those in (i) and hence omitted to save space. Table 9 shows the posterior means, 95% credible intervals and inefficiency factors. The estimated parameters are close to true values, and inefficiency factors (IF) are overall relatively small. The IF's for h_{250} and h_{750} are small, and indicates that the algorithm is still works well.

Par	True	Mean	Std Dev	95% interval	IF
μ	0	0.109	0.326	(-0.616, 0.756)	31
ϕ	0.97	0.971	0.010	(0.949, 0.989)	23
σ	0.3	0.265	0.035	(0.202, 0.339)	15
β	0.7	0.704	0.037	(0.631, 0.776)	5
h_{250}	2.310	1.881	0.454	(1.007, 2.792)	4
h_{750}	2.077	1.662	0.404	(0.901, 2.493)	5

Table 4: True values, posterior means, posterior standard deviations, 95% credible intervals, and inefficiency factors. $\beta = 0.7$.

These simulation results show that our proposed sampling method works well for those β's found in the past empirical studies.

5 Extension to SVM model with Leverage (SVML model)

In this section, we consider the SVM model with leverage which we call SVML model. The leverage effect implies the decrease in the return at time t followed by the increase in the volatility at time $t+1$. Thus we incorporate the correlation ρ between y_t and h_{t+1} and replace (3) by

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \overset{\text{i.i.d.}}{\sim} N(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}. \tag{13}$$

The negative correlation, $\rho < 0$, indicates the existence of the leverage effect. Next, we construct the linear and Gaussian state space model that approximates the SVM model with leverage using the mixture of the normal densities given in (10). We first let $d_t = I(y_t \geq 0) - I(y_t < 0)$ where $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. Noting that

$$\begin{aligned} y_t =& d_t \exp(y_t^*/2), \quad \epsilon_t = d_t \exp(\epsilon_t^*/2) - \beta \\ & \eta_t|\epsilon_t \sim N(\rho\sigma\epsilon_t, \sigma^2(1\text{–}\rho^2)), \end{aligned}$$

we rewrite the conditional distribution as

$$\eta_t|\epsilon_t \sim N\left(\rho\sigma\{d_t \exp(\epsilon_t^*/2) - \beta\}, \sigma^2(1\text{–}\rho^2)\right).$$

Let $s_t = (s_{1t}, s_{2t}) \in \{(i,j)|i = 1, ..., K, j = 0, ..., J\}$ denote the component of the mixture of normal densities in (10) at time t. Given $s_t = (i,j)$, we have $\epsilon_t^*|s_t = (i,j) \sim N(\tilde{m}_{i,j}, v_i^2)$. Furthermore, we approximate $\exp(\epsilon_t^*/2)$ by $\exp(\tilde{m}_{i,j}/2)\{a_i + b_i(\epsilon_t^* - \tilde{m}_{i,j})\}$ with $a_i = \exp(v_i^2/8)$, $b_i = \frac{1}{2}\exp(v_i^2/8)$, as in Table 1, which minimize the mean square norm

$$E[\exp(\epsilon_t^*/2) - \exp(\tilde{m}_{i,j}/2)\{a_i + b_i(\epsilon_t^* - \tilde{m}_{i,j})\}]^2.$$

Thus, the approximate conditional distribution is

$$\eta_t|\epsilon_t \sim N\left(\rho\sigma[d_t \exp(\tilde{m}_{i,j}/2)\{a_i + b_i(\epsilon_t^* - \tilde{m}_{i,j})\} - \beta], \sigma^2(1\text{–}\rho^2)\right).$$

Given $s = (s_1, ..., s_n)$, we find that the SVM model can be approximated by the linear Gaussian state space form

$$y_t^* = \tilde{m}_{s_{1t},s_{2t}} + h_t + (v_{s_{1t}}, 0)z_t, \tag{14}$$

$$h_{t+1} = \mu(1-\phi) + \rho\sigma\{d_t a_{s_{1t}} \exp(\tilde{m}_{s_{1t},s_{2t}}) - \beta\} + \phi h_t \tag{15}$$
$$+ (d_t\rho\sigma b_{s_{1t}} v_{s_{1t}} \exp(\tilde{m}_{s_{1t},s_{2t}}/2), \sigma\sqrt{1-\rho^2})z_t,$$

$$t = 1, ..., n-1, \quad h_1 \sim N\left(\mu, \frac{\sigma^2}{1-\phi^2}\right), \quad |\phi| < 1, \tag{16}$$

$$z_t = (z_{1t}, z_{2t})' \sim N(0, I_2),$$

where $y_t^* = \log(y_t^2)$ and $d_t = I(y_t \geq 0) - I(y_t < 0)$. In the following subsections, $\tilde{m}_{s_{1t},s_{2t}}$ and $\tilde{p}_{s_{1t},s_{2t}}$ are abbreviated as $\tilde{m}_{s_t}$ and $\tilde{p}_{s_t}$, and we write $v_{s_t}^2, a_{s_t}$, and b_{s_t} instead of $v_{s_{1t}}^2, a_{s_{1t}}$, and $b_{s_{1t}}$, respectively. The MCMC algorithm and the particle filter are detailed in Appendix B.

6 Empirical studies of excess holding yield data

6.1 Data

This section applies the SVM model with leverage and several alternative models to three excess holding yields data. Further, we conduct a comprehensive model comparison including the GARCH, GARCH in mean, EGARCH and EGARCH in mean models. The descriptions of the data (labeled as TB, DGS and AAA) are given below[1].

(1) TB: the excess holding yield using 3 and 6 months treasury bills with 258 observations from the forth quarter of 1958 to the first of 2023. It is defined as

$$y_t = \left\{ \frac{\left(1 + \frac{R_t}{100}\right)^2}{1 + \frac{r_{t+1}}{100}} - \left(1 + \frac{r_t}{100}\right) \right\} \times 100,$$

at annual rate where R_t and r_t are secondary market rates of the 6-month and 3-month Treasury bill (discount basis, percent, daily, not seasonally adjusted), measured at the beginning of the quarter.

(2) DGS: the excess holding yield using 1 and 3 month market yields on U.S. treasury securities with 266 observations from August 2001 to September 2023. The excess

[1]The data are obtained from the website of Federal Reserve Bank of St. Louis.

holding yield, y_t is defined as

$$y_t = \left\{ \frac{\left(1 + \frac{R_t}{100}\right)^3}{\left(1 + \frac{r_{t+1}}{100}\right)\left(1 + \frac{r_{t+2}}{100}\right)} - \left(1 + \frac{r_t}{100}\right) \right\} \times 100,$$

at annual rate where R_t and r_t are market yields of 3 and 1 month on US Treasury securities at constant maturity of 3 months (quoted on an investment basis, percent, daily, not seasonally adjusted), measured at the beginning of the month.

(3) AAA: the excess holding yield using 3 months and 20 years Moody's Seasoned AAA corporate bond yield with 359 observations from the fourth quarter of 1933 to the second of 2023. As discussed in Engle et al. (1987), assuming that the bonds are effectively infinitely lived, we define

$$y_t = \left\{ \frac{R_t}{100} - \frac{r_t}{100} - 1 + \frac{R_t}{R_{t+1}} \right\} \times 100,$$

at annual rate where R_t is the Moody's seasoned AAA corporate bond yield (percent, monthly, not seasonally adjusted) and r_t is the 3-month treasury bill secondary market rate (percent, discount basis, monthly, not seasonally adjusted), measured at the beginning of the quarter.

The time series plots of three datasets are shown in Figures 6, 7 and 8. Volatility clustering phenomena is observed for all three series, suggesting that the stochastic volatility models are appropriate to describe these excess holding yield data.

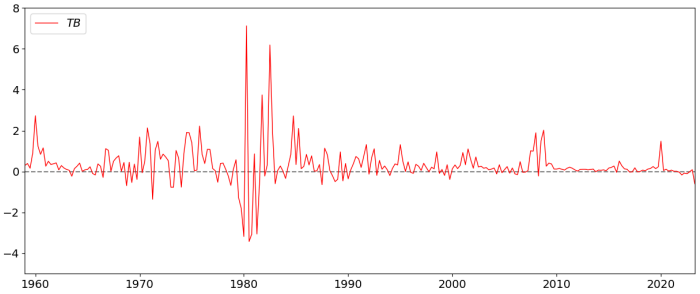


Figure 6: Time series plot. TB data: 1958Q4–2023Q1.

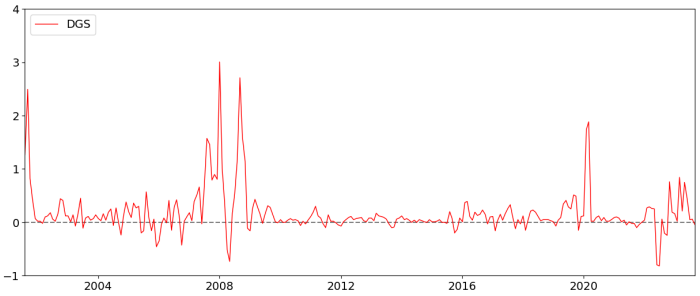


Figure 7: Time series plot. DGS data: 2001/8–2023/9.

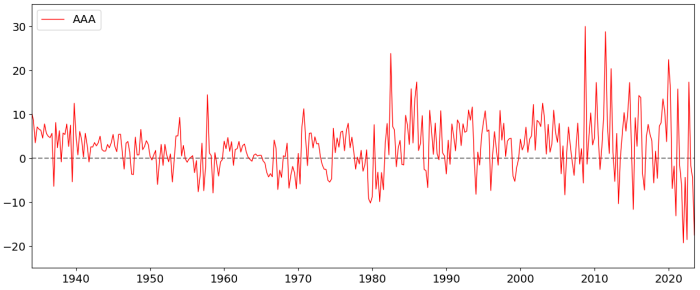


Figure 8: Time series plot. AAA data: 1933Q4–2023Q2.

6.2 Estimation results for SVM model

Using the same prior distributions as in illustrative examples in Section 4, the proposed SVM models with leverage are fitted to TB, DGS and AAA data. We iterated MCMC simulation 150,000 times after discarding initial 30,000 MCMC draws as burn-in period using Algorithm 3 in Appendix B. The acceptance rates of the MH algorithms for α and (α, h) are 70.2% and 20.7% with TB data, 69.5% and 13.2% with DGS data, and 51.1% and 36.6% with AAA data, respectively.

Par	Mean	Std Dev	95% interval	IF	Pr(+)
μ	-1.882	0.623	(-3.163, -0.656)	109	0.003
	-3.824	0.641	(-5.053, -2.319)	130	0.000
	3.699	0.882	(2.136, 5.877)	86	1.000
ϕ	0.920	0.024	(0.867, 0.961)	72	1.000
	0.901	0.040	(0.815, 0.965)	156	1.000
	0.982	0.015	(0.943, 0.999)	65	1.000
σ	0.690	0.099	(0.519, 0.911)	105	1.000
	0.923	0.130	(0.696, 1.2)	175	1.000
	0.196	0.055	(0.103, 0.32)	79	1.000
ρ	-0.544	0.139	(-0.79, -0.253)	116	0.000
	0.062	0.132	(-0.203, 0.313)	151	0.672
	-0.305	0.130	(-0.549, -0.038)	41	0.012
β	0.651	0.074	(0.507, 0.797)	45	1.000
	0.735	0.075	(0.589, 0.881)	52	1.000
	0.440	0.057	(0.328, 0.552)	12	1.000
h_{100}	-0.680	0.957	(-2.692, 1.076)	56	0.242
	-5.493	1.007	(-7.454, -3.467)	72	0.000
	2.873	0.358	(2.208, 3.634)	17	1.000

Table 5: Posterior mean, standard deviation, 95% credible interval, inefficient factor and the posterior probability that the parameter is positive. TB (top row), DGS (middle row), and AAA (bottom row).

Table 5 shows posterior means, standard deviations, 95% credible intervals, inefficiency factors for parameters (IF), and the posterior probability that the parameter is positive for three datasets. Further, the IF's for log volatilities, h_t's are found to be less than 80, which implies that our mixture sampler is highly efficient as shown in Section 4. The coefficient β is estimated to be greater than 0.4, and we find a strong evidence of the positive risk premium since the posterior probability that β is positive is almost one, $Pr(\beta > 0|y) \approx 1.000$. The autoregressive parameter ϕ for the log volatility process is estimated to be more than 0.9, suggesting the high persistence in the volatility as found in the various empirical studies in the previous literature. The correlation parameter ρ is estimated to be negative for TB data and AAA data, implying a strong evidence of the leverage effect since the posterior probability that ρ is negative is almost one, $Pr(\rho < 0|y) \approx 1.000$. On the other hand, there

is no evidence that ρ is negative for DGS data. Finally, Figures 9, 10 and 11 show 95% credible intervals and posterior medians for h in the case of TB, DGS and AAA data. Noting that

$$\log(y_t^2) = h_t + \log \chi_1^2(\beta^2),$$

we further plotted moving average $\sum_{s=-10}^{10} z_{t+s}/21$ for reference where $z_t = \log(y_t^2) - E(\log \chi_1^2(\beta^2))$ is evaluated at the posterior means of β. The expected values of $\log \chi_1^2(\beta^2)$ are computed numerically as $-0.88, -0.78$, and -1.08 for TB, DGS, and AAA using Monte Carlo integration. The traceplot of the estimated log volatilities is similar to that of the moving average series taking account of 95% credible intervals. The large volatilities around years 1980, 2008 and 2010-2023 are well captured by the proposed model as shown in Figures 9, 10 and 11 respectively.

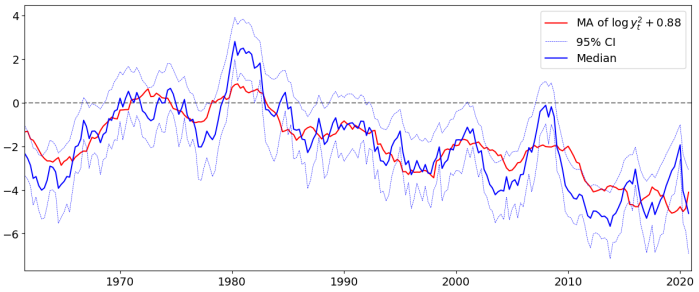


Figure 9: Log volatilities: Moving average of $\log(y_t^2) - E(\log \chi_1^2(\beta^2))$, 95% credible intervals and posterior median. TB data.

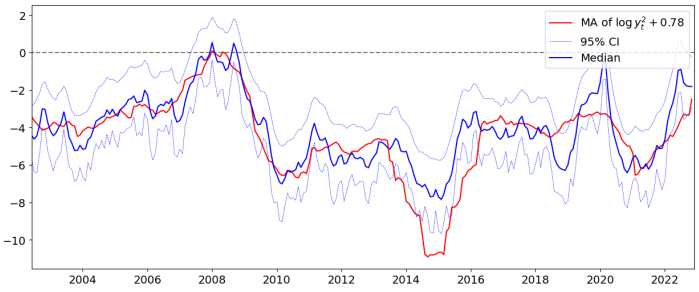


Figure 10: Log volatilities: Moving average of $\log(y_t^2) - E(\log \chi_1^2(\beta^2))$, 95% credible intervals and posterior median. DGS data.

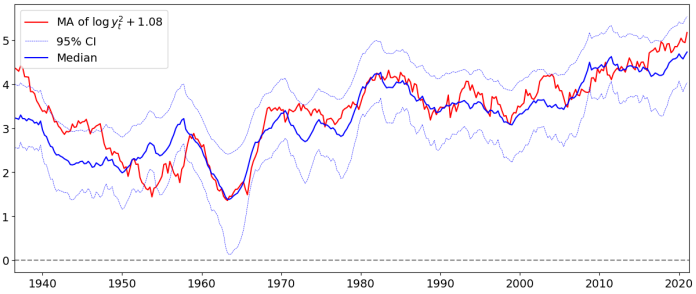


Figure 11: Log volatilities: Moving average of $\log(y_t^2) - E(\log \chi_1^2(\beta^2))$, 95% credible intervals and posterior median. AAA data.

6.3 Model comparison

In this section we use Bayesian marginal likelihoods to conduct a comprehensive comparisons of various volatility models that includes GARCH models. We calculate the marginal likelihood using the method of Chib (1995). Suppressing the model index, this method is based on an identity introduced in that paper:

$$\log m(y) = \log f(y|\theta) + \log \pi(\theta) - \log \pi(\theta|y),$$

where the first term on the right side is the log of the likelihood. the second term is the prior, and the third is the posterior density. We evaluate each of these terms with the posterior mean of θ. For each model, we calculate the first term using the particle filter method given in Section 3.2, setting $I = 80,000$. To compute the posterior density ordinate, we apply Chib and Jeliazkov (2001) to the MCMC draws from Algorithm 4 in Appendix C. As Engle et al. (1987) considered three volatility in mean models in the context of the autoregressive conditional heteroscedasticity in mean (ARCH-M) model. The GARCH-M-type models are defined as

$$y_t = \beta\varphi(h_t) + u_t, \quad u_t = \sqrt{h_t}\nu_t, \tag{17}$$

$$h_t = \omega + \gamma h_{t-1} + \alpha u_{t-1}^2, \quad h_1 = \frac{\omega}{1-\alpha-\gamma}, \tag{18}$$

where $\nu_t \overset{\text{i.i.d.}}{\sim} N(0,1)$, $\alpha > 0$, $\gamma > 0$, $\alpha+\gamma < 1$, and $\omega > 0$. Note that the h_t in the GARCH-M models represents the standard deviation of the error term u_t, and this is different from the h_t used in SVM models which represents the log volatility. We consider

- GARCH-M model: $\varphi(h_t) = \sqrt{h_t}$.

- GARCH-M Ⅱ model: $\varphi(h_t) = h_t$.
- GARCH-M Ⅲ model: $\varphi(h_t) = \log h_t$.

and GARCH model (GARCH, $\beta = 0$). The EGARCH-M models are also defined by replacing (18) with

$$\log h_t = \omega + \gamma \log h_{t-1} + \alpha\nu_{t-1} + \delta(|\nu_{t-1}| - E[|\nu_{t-1}|]), \quad \log h_1 = \frac{\omega}{1-\gamma},$$

where $|\gamma| < 1$. As in GARCH-M models, we consider EGARCH-M, EGARCH-M Ⅱ, EGARCH-M Ⅲ, EGARCH models for the comparison. The prior distributions are chosen to reflect no prior information regarding of model parameters. Correspondingly, we additionaly consider

- SVM Ⅱ model: $y_t = \beta \exp(h_t) + \epsilon_t \exp(h_t/2)$.
- SVM Ⅲ model: $y_t = \beta h_t + \epsilon_t \exp(h_t/2)$.

Table 6 presents the logarithm of the marginal likelihood for TB, DGS, and AAA data. As for TB and DGS data, the SVML model yields the largest log marginal likelihood among those competing models, which implies our proposed model best describes the risk premium and the time varying volatility among competing models. However, taking account of the standard error, the SVM model performs as well as SVML model for DGS data. Note that our proposed models (SVM and SVML) outperform other SV and SVM models. These results are also consistent with high posterior probabilities of $Pr(\beta > 0|y)$ for TB and DGS data, and of $Pr(\rho < 0|y)$ for TB data as given in Table 5 of Section 6.2. In most cases, the class of SV models outperforms that of the GARCH and EGARCH models. This comparison shows that the SVM-type models fit better than the GARCH-M and EGARCH-M-type models in the analysis for TB and DGS data and that our proposed SVM model outperform other comprehensive volatility in mean models.

On the other hand, for AAA data, we found the GARCH-M Ⅲ model attains the largest marginal likelihood. This is somewhat surprising, but it may be because the length of the bond holding period is much longer than those of TB and AAA data. Moreover, the volatility clustering in Figure 8 looks different from those in Figures 6 and 7 for TB and DGS data. Also, the level of the volatilities in Figure 11 changes more smootherly than those in Figures 9 and 10. However, we note that the SVML model still outperform other SV models.

Model	TB	DGS	AAA
SVM	-185.636(0.156)	**66.390(0.109)**	-1131.540(0.051)
SVML	**-178.409(0.111)**	**66.616(0.122)**	-1126.976(0.061)
SVM Ⅱ	-211.910(0.072)	52.084(0.046)	-1136.338(0.047)
SVML Ⅱ	-203.926(0.064)	49.823(0.046)	-1131.055(0.066)
SVM Ⅲ	-215.386(0.049)	38.438(0.055)	-1141.540(0.041)
SVML Ⅲ	-205.118(0.052)	39.304(0.046)	-1139.085(0.045)
SV	-224.784(0.090)	21.126(0.122)	-1155.443(0.049)
SVL	-214.247(0.195)	22.549(0.060)	-1154.006(0.153)
GARCH-M	-215.283(0.003)	10.194(0.005)	-1125.723(0.007)
GARCH-M Ⅱ	-238.244(0.006)	1.940(0.005)	-1129.148(0.010)
GARCH-M Ⅲ	-233.935(0.133)	14.178(0.231)	**-1119.157(0.161)**
GARCH	-259.238(0.011)	-21.886(0.009)	-1165.468(0.010)
EGARCH-M	-229.199(0.043)	5.297(0.003)	-1144.959(0.006)
EGARCH-M Ⅱ	-244.028(0.009)	4.743(0.007)	-1151.550(0.006)
EGARCH-M Ⅲ	-253.988(0.010)	-2.3118(0.005)	-1143.756(0.005)
EGARCH	-254.035(0.005)	-14.376(0.004)	-1168.813(0.008)

Table 6: Log marginal likelihood estimation and standard error (in parentheses). TB, DGS, and AAA data. SVML, Ⅱ and Ⅲ, and SVL model include the leverage effect ρ. The bold font indicates the largest marginal likelihood.

7 Conclusion

In this paper, we have successfully extended the mixture sampler for the SV model to the SVM model of which the mean equation is described by the standard deviation of the error term as an independent variable. Our main point is the approximation of the distribution of $\log \chi_1^2(\beta^2)$ by mixture of normal distributions which is dependent on the parameter β. This approximation facilitates efficient sampling, leveraging well-established methods for the linear Gaussian state-space model. It is shown in simulation studies that our proposed method is implemented easily and works efficiently. In the empirical studies of the excess holding yield data, we conducted the extensive model comparison including the SVM, GARCH-M and EGARCH-M models, and found that our proposed SVM outperforms other models in terms of the marginal likelihood. It also shows that there exists the positive risk premiums and

time-varying volatility in the excess holding yield data.

Acknowledgement

This work is partially supported by JSPS KAKENHI [Grant number: 24H00142]. The computational results are obtained using Rcpp and Ox (see Doornik (2007)).

References

Abanto-Valle, C. A., H. Migon, and V. Lachos (2011). Stochastic volatility in mean models with scale mixtures of normal distributions and correlated errors: A bayesian approach. *Journal of Statistical Planning and Inference 141*(5), 1875–1887.

Abanto-Valle, C. A., H. S. Migon, and V. H. Lachos (2012). Stochastic volatility in mean models with heavy-tailed distributions.

Abanto-Valle, C. A., G. Rodríguez, L. M. Castro Cepero, and H. B. Garrafa-Aragón (2023). Approximate bayesian estimation of stochastic volatility in mean models using hidden markov models: Empirical evidence from emerging and developed markets. *Computational Economics*, 1–27.

Abanto-Valle, C. A., G. Rodríguez, and H. B. Garrafa-Aragón (2021). Stochastic volatility in mean: Empirical evidence from latin-american stock markets using hamiltonian monte carlo and riemann manifold hmc methods. *The Quarterly Review of Economics and Finance 80*, 272–286.

Berument, H., Y. Yalcin, and J. Yildirim (2009). The effect of inflation uncertainty on inflation: Stochastic volatility in mean model within a dynamic framework. *Economic Modelling 26*(6), 1201–1207.

Chan, J. C. (2017). The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics 35*(1), 17–28.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the american statistical association 90*(432), 1313–1321.

Chib, S. and E. Greenberg (1995). Understanding the metropolis-hastings algorithm. *The american statistician 49*(4), 327–335.

Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the metropolis–hastings output. *Journal of the American statistical association 96*(453), 270–281.

Chib, S., F. Nardari, and N. Shephard (2002). Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics 108*(2), 281–316.

Cross, J. L., C. Hou, G. Koop, and A. Poon (2023). Large stochastic volatility in mean vars. *Journal of Econometrics 236*(1), Article 105469.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*(2), 339–350.

Del Negro, M. and G. E. Primiceri (2015). Time varying structural vector autoregressions and monetary policy: a corrigendum. *The review of economic studies 82*(4), 1342–1345.

Doornik, J. A. (2007). *Object-Oriented Matrix Programming Using Ox, 3rd ed.* London: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–616.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987). Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica: journal of the Econometric Society*, 391–407.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous univariate distributions, volume 2*, Volume 289. John wiley & sons.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies 65*(3), 361–393.

Koopman, S. J. and E. Hol Uspensky (2002). The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of applied Econometrics 17*(6), 667–689.

Leão, W. L., C. A. Abanto-Valle, and M.-H. Chen (2017). Bayesian analysis of stochastic volatility-in-mean model with leverage and asymmetrically heavy-tailed error using generalized hyperbolic skew student's t-distribution. *Statistics and its Interface 10*, 529.

Mumtaz, H. and F. Zanetti (2013). The impact of the volatility of monetary policy shocks. *Journal of Money, Credit and Banking 45*(4), 535–558.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics 140*(2), 425–449.

Omori, Y. and T. Watanabe (2008). Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics & Data Analysis 52*(6), 2892–2910.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association 94*(446), 590–599.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika 84*(3), 653–667.

Takahashi, M., Y. Omori, and T. Watanabe (2023). *Stochastic volatility and realized stochastic volatility models*. Springer. SpringerBriefs in Statistics (JSS Research Series in Statistics).

Taylor, S. J. (2008). *Modelling financial time series*. world scientific.

Appendix

A MH step to correct the approximation error

The posterior density is given by

$$\begin{aligned}
\pi(h,\theta|y) &\propto f(y,h|\theta)\pi(\theta) \\
&\propto (1+\phi)^{a-\frac{1}{2}}(1-\phi)^{b-\frac{1}{2}}(\sigma^2)^{-(\frac{n_1}{2}+1)}\exp\left\{-\frac{1}{2\sigma^2}\{s_0+(1-\phi^2)(h_1-\mu)^2\}\right\} \\
&\quad\times\exp\left\{-\frac{1}{2}\sum_{t=1}^{n}[h_t+\{y_t-\beta\exp(h_t/2)\}^2\exp(-h_t)]\right\} \\
&\quad\times\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n-1}[h_{t+1}-\mu(1-\phi)-\phi h_t]^2\right\} \\
&\quad\times\exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}\exp\left\{-\frac{(\beta-b_0)^2}{2B_0}\right\},
\end{aligned}$$

where $n_1 = n_0+n$. To correct the approximation error in Step 3 of Algorithm 1, we implement the additional MH step (Step 4) as in the following Algorithm 2.

Algorithm 2. Let us denote $\theta = (\alpha, \beta)$ where $\alpha = (\mu, \phi, \sigma^2)$. The Markov chain Monte Carlo simulation is implemented in four blocks:

1. Initialize h and $\theta = (\alpha, \beta)$.
2. Generate $\beta|\alpha, h, y \sim \pi(\beta|\alpha, h, y)$ as in Algorithm 1.
3. Generate $(\alpha, h)|\beta, y \sim \pi(\alpha, h|\beta, y)$ as in Algorithm 1.
4. Conduct MH algorithm to correct the approximation error.
5. Go to step 2.

Step 4. Generation of $(\alpha, h)|\beta, y$

Since the mixture sampler is based on the approximation, we can correct the approximation error after the MCMC simulation as in Kim et al. (1998) and Omori et al. (2007)) Instead, we use the data augmentation method to correct it within the MCMC simulation by MH algorithm with the pseudo target density. We note that the similar approach has been considered for the SV model without leverage (Del Negro and Primiceri (2015)) and with leverage (Takahashi et al. (2023)). Define the pseudo target density

$$\tilde{\pi}(\alpha, h, s|\beta, y) = \pi(\alpha, h|\beta, y) \times q(s|h, \alpha, \beta, y^*),$$
$$q(s|h, \alpha, \beta, y^*) = \prod_{t=1}^{n} \frac{\tilde{p}_{s_t} g(y_t^*|h_t, \alpha, \beta, s_t)}{\sum_{i=1}^{10} \sum_{j=0}^{2} \tilde{p}_{i,j} g(y_t^*|h_t, \alpha, \beta, s_t = (i, j))},$$

Note that the marginal density $\pi(\alpha, h|\beta, y)$ is our target density, $\pi(\alpha, h|\beta, y) = \sum_s \tilde{\pi}(\alpha, h, s|\beta, y)$. We generate sample (α, h, s) from the pseudo target density as follows. Using Step 3 of Algorithm 1, we have a sample from $\pi^*(h|\alpha, s, \beta, y^*)\pi^*(\alpha|s, \beta, y)$ and let us denote it as $(\alpha^\dagger, h^\dagger)$, and let

$$f(y_t|h_t, \alpha, \beta) = f_N(y_t|\beta \exp(h_t/2), \exp(h_t)), \quad t = 1, \ldots, n.$$

Given the current value (α, h), accept the candidate $(\alpha^\dagger, h^\dagger)$ with probability

$$\begin{aligned}
&\min\left\{1, \frac{\tilde{\pi}(\alpha^\dagger, h^\dagger|s, \beta, y)\pi^*(h|\alpha, s, \beta, y^*)\pi^*(\alpha|s, \beta, y)}{\tilde{\pi}(\alpha, h|s, \beta, y)\pi^*(h^\dagger|\alpha^\dagger, s, \beta, y^*)\pi^*(\alpha^\dagger|s, \beta, y)}\right\} \\
&= \min\left\{1, \frac{\pi(\alpha^\dagger, h^\dagger|\beta, y)q(s|h^\dagger, \alpha^\dagger, \beta, y^*)\pi^*(h|\alpha, s, \beta, y^*)\pi^*(\alpha|s, \beta, y)}{\pi(\alpha, h|\beta, y)q(s|h, \alpha, \beta, y^*)\pi^*(h^\dagger|\alpha^\dagger, s, \beta, y^*)\pi^*(\alpha^\dagger|s, \beta, y)}\right\} \\
&= \min\left\{1, \frac{q(s|h^\dagger, \alpha^\dagger, \beta, y^*)\prod_{t=1}^{n} f(y_t|h_t^\dagger, \alpha^\dagger, \beta)g(y_t^*|h_t, \alpha, \beta, s_t)}{q(s|h, \alpha, \beta, y^*)\prod_{t=1}^{n} f(y_t|h_t, \alpha, \beta)g(y_t^*|h_t^\dagger, \alpha^\dagger, \beta, s_t)}\right\} \\
&= \min\left\{1, \frac{\prod_{t=1}^{n} f(y_t|h_t^\dagger, \alpha^\dagger, \beta)\sum_{i=1}^{10}\sum_{j=0}^{2} \tilde{p}_{i,j} g(y_t^*|h_t, \alpha, \beta, s_t = (i, j))}{\prod_{t=1}^{n} f(y_t|h_t, \alpha, \beta)\sum_{i=1}^{10}\sum_{j=0}^{2} \tilde{p}_{i,j} g(y_t^*|h_t^\dagger, \alpha^\dagger, \beta, s_t = (i, j))}\right\}.
\end{aligned}$$

Below we compare estimates using Algorithms 1 and 2 in illustrative examples. The results are quite close to each other, implying that the approximation is highly accurate.

		Algorithm 1				Algorithm 2			
Par	True	Mean	Std Dev	95% interval	IF	Mean	Std Dev	95% interval	IF
μ	0	0.090	0.319	(-0.575, 0.725)	17	0.101	0.311	(-0.537, 0.731)	24
ϕ	0.97	0.971	0.011	(0.948, 0.989)	15	0.972	0.010	(0.949, 0.988)	29
σ	0.3	0.259	0.037	(0.195, 0.338)	13	0.257	0.036	(0.194, 0.335)	41
β	0.3	0.315	0.033	(0.251, 0.380)	1	0.327	0.033	(0.263, 0.392)	3
h_{250}	2.310	1.735	0.460	(0.843, 2.647)	5	1.722	0.454	(0.858, 2.632)	16
h_{750}	2.077	1.674	0.412	(0.900, 2.520)	5	1.715	0.415	(0.932, 2.549)	14

Table 7: True values, posterior means, posterior standard deviations, 95% credible intervals, and inefficiency factors. $\beta = 0.3$.

		Algorithm 1				Algorithm 2			
Par	True	Mean	Std Dev	95% interval	IF	Mean	Std Dev	95% interval	IF
μ	0	0.102	0.319	(-0.562, 0.754)	18	0.104	0.303	(-0.516, 0.740)	48
ϕ	0.97	0.971	0.011	(0.947, 0.988)	13	0.971	0.010	(0.950, 0.988)	53
σ	0.3	0.263	0.037	(0.198, 0.342)	13	0.262	0.035	(0.199, 0.333)	89
β	0.5	0.511	0.035	(0.443, 0.579)	3	0.530	0.034	(0.462, 0.597)	12
h_{250}	2.310	1.811	0.459	(0.928, 2.719)	4	1.758	0.450	(0.891, 2.631)	38
h_{750}	2.077	1.641	0.410	(0.866, 2.476)	4	1.710	0.417	(0.930, 2.574)	41

Table 8: True values, posterior means, posterior standard deviations, 95% credible intervals, and inefficiency factors. $\beta = 0.5$.

		Algorithm 1				Algorithm 2			
Par	True	Mean	Std Dev	95% interval	IF	Mean	Std Dev	95% interval	IF
μ	0	0.109	0.326	(-0.616, 0.756)	31	0.103	0.305	(-0.479, 0.802)	84
ϕ	0.97	0.971	0.010	(0.949, 0.989)	23	0.971	0.011	(0.945, 0.990)	110
σ	0.3	0.265	0.035	(0.202, 0.339)	15	0.263	0.035	(0.199, 0.343)	139
β	0.7	0.704	0.037	(0.631, 0.776)	5	0.731	0.037	(0.659, 0.804)	37
h_{250}	2.310	1.881	0.454	(1.007, 2.792)	4	1.813	0.457	(0.966, 2.722)	74
h_{750}	2.077	1.662	0.404	(0.901, 2.493)	5	1.660	0.399	(0.924, 2.441)	69

Table 9: True values, posterior means, posterior standard deviations, 95% credible intervals, and inefficiency factors. $\beta = 0.7$.

B MCMC algorithm and particle filter for SVM with leverage

B.1 MCMC algorithm

For the SVM with leverage, we set $\theta = (\mu, \phi, \sigma^2, \rho)$. For the prior distribution of ρ, we assume $\rho \sim U(-1, 1)$ where $U(a, b)$ denotes uniform distribution over (a, b), and the posterior density function of (h, θ) is given by

$$
\begin{aligned}
&\pi(h, \theta|y) \\
&\quad \propto f(y, h|\theta)\pi(\theta) \\
&\quad \propto (1+\phi)^{a-\frac{1}{2}}(1-\phi)^{b-\frac{1}{2}}(1-\rho^2)^{-\frac{n-1}{2}}(\sigma^2)^{-\left(\frac{n_1}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}\{s_0 + (1-\phi^2)(h_1-\mu)^2\}\right\} \\
&\qquad \times \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}[h_t + \{y_t - \beta\exp(h_t/2)\}^2\exp(-h_t)]\right\} \\
&\qquad \times \exp\left\{-\frac{1}{2\sigma^2(1-\rho^2)}\sum_{t=1}^{n-1}[h_{t+1} - \mu(1-\phi) - \phi h_t - \rho\sigma\{y_t - \beta\exp(h_t/2)\}\exp(-h_t/2)]^2\right\} \\
&\quad \times \exp\left\{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right\}\exp\left\{-\frac{(\beta-b_0)^2}{2B_0}\right\},
\end{aligned}
$$

where $n_1 = n_0 + n$.

Algorithm 3. Let us denote $\theta = (\alpha, \beta)$ where $\alpha = (\mu, \phi, \sigma^2, \rho)$. The Markov chain Monte Carlo simulation is implemented in four blocks:

1. Initialize h and $\theta = (\alpha, \beta)$.
2. Generate $\beta|\alpha, h, y \sim \pi(\beta|\alpha, h, y)$.
3. Generate $(\alpha, h)|\beta, y \sim \pi(\alpha, h|\beta, y)$.
4. Go to step 2.

Step 2. Generation of $\beta|\alpha, h, y$

The conditional posterior distribution of β is normal with mean b_1 and variance B_1 where

$$
b_1 = B_1\left(X'\Omega^{-1}\tilde{y} + B_0^{-1}b_0\right), \quad B_1^{-1} = X'\Omega^{-1}X + B_0^{-1},
$$

and

$$\tilde{y} = \begin{pmatrix} y_1 - \rho\exp(h_1/2)\sigma^{-1}\{h_2 - \mu - \phi(h_1-\mu)\} \\ y_2 - \rho\exp(h_2/2)\sigma^{-1}\{h_3 - \mu - \phi(h_2-\mu)\} \\ \vdots \\ y_{n-1} - \rho\exp(h_{n-1}/2)\sigma^{-1}\{h_n - \mu - \phi(h_{n-1}-\mu)\} \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} \exp(h_1/2) \\ \exp(h_2/2) \\ \vdots \\ \exp(h_n/2) \end{pmatrix},$$

$$\Omega = \mathrm{diag}\left((1-\rho^2)\exp(h_1), (1-\rho^2)\exp(h_2), \cdots, (1-\rho^2)\exp(h_{n-1}), \exp(h_n)\right),$$

Thus we generate $\beta \sim N(b_1, B_1)$.

Step 3. Generation of $(\alpha, h)|\beta, y$

We define the pseudo target density

$$\tilde{\pi}(\alpha, h, s|\beta, y) = \pi(\alpha, h|\beta, y) \times q(s|\alpha, h, \beta, y^*, d),$$

$$q(s|\alpha, h, \beta, y^*, d) = \prod_{t=1}^{n} \frac{\tilde{p}_{s_t} g(y_t^*, h_{t+1}|h_t, \alpha, \beta, s_t, d)}{\sum_{i=1}^{10}\sum_{j=0}^{2} \tilde{p}_{i,j} g(y_t^*, h_{t+1}|h_t, \alpha, \beta, s_t = (i,j), d)},$$

where $\theta = (\alpha, \beta)$ and

$$g(y_t^*, h_{t+1}|h_t, \alpha, \beta, s_t, d) = \begin{cases} f_N(y_t^*|\tilde{m}_{s_t} + h_t, v_{s_t}^2) f_N(h_{t+1}|\overline{h}_{s_t,t}, \sigma^2(1-\rho^2)), & t < n, \\ f_N(y_t^*|\tilde{m}_{s_t} + h_t, v_{s_t}^2) & t = n, \end{cases}$$

$$\overline{h}_{s_t,t} = \mu(1-\phi) + \phi h_t + \rho\sigma[d_t \exp(\tilde{m}_{s_t}/2)\{a_{s_t} + b_{s_t}(y_t^* - h_t - \tilde{m}_{s_t})\} - \beta],$$

where $y_t^* = \log(y_t^2)$, $d_t = I(y_t \geq 0) - I(y_t < 0)$. $\tilde{m}_{s_t} = \tilde{m}_{s_{1t}, s_{2t}}$ and $\tilde{p}_{s_t} = \tilde{p}_{s_{1t}, s_{2t}}$ are defined in (10) and $(p_{s_{1t}}, m_{s_{1t}}, v_{s_{1t}}^2, a_{s_{1t}}, b_{s_{1t}})$ are given in Table 1. Only $\tilde{p}_{s_t}$, which depends on β, needs to be updated according to the formula in (10) before sampling. Note that the marginal density $\pi(\alpha, h|\beta, y)$ is our target density, $\pi(\alpha, h|\beta, y) = \sum_s \tilde{\pi}(\alpha, h, s|\beta, y)$. We generate sample (α, h, s) from the pseudo target density in two steps.

(a) Generate $s \sim q(s|h, \theta, y^*, d)$.

(b) Generate $(\alpha, h)|\theta, s, y \sim \tilde{\pi}(\alpha, h|\beta, s, y)$.

(i) Generate $\alpha \sim \pi^*(\alpha|s, \beta, y^*, d)$. The target density here is given by

$$\pi^*(\alpha|s, \beta, y^*, d) \propto m(y^*|\alpha, s, \beta, d)\pi(\alpha),$$

where

$$m(y^*|\alpha, s, \beta, d) = \int \prod_{t=1}^{n} g(y_t^*, h_{t+1}|h_t, \alpha, s_t, \beta, d) \times f_N\left(h_1 \middle| \mu, \frac{\sigma^2}{1-\phi^2}\right) dh,$$

which we evaluate using Kalman filter algorithm. We first transform α to $\vartheta = (\mu, \log\{(1+\phi)/(1-\phi)\}, \log\sigma^2, \log\{(1+\rho)/(1-\rho)\})$ to remove parameter constraints, and conduct MH algorithm to sample from the conditional posterior distribution with density $\pi^*(\vartheta|s, \beta, y) = \pi^*(\alpha|s, \beta, y) \times |d\alpha/d\vartheta|$ where $|d\alpha/d\vartheta|$ is the Jacobian of the transformation. Compute the posterior mode $\hat{\vartheta}$ and define ϑ_* and Σ_* as

$$\vartheta_* = \hat{\vartheta}, \quad \Sigma_*^{-1} = -\frac{\partial^2 \log \pi^*(\vartheta|s, \beta, y)}{\partial\vartheta\partial\vartheta'}\bigg|_{\vartheta=\hat{\vartheta}}.$$

Given the current value ϑ, generate a candidate $\vartheta^\dagger$ from the distribution $N(\vartheta_*, \Sigma_*)$ and accept it with probability

$$\alpha(\vartheta, \vartheta^\dagger|s, \beta, y) = \min\left\{1, \frac{\pi^*(\vartheta^\dagger|s, \beta, y) f_N(\vartheta|\vartheta_*, \Sigma_*)}{\pi^*(\vartheta|s, \beta, y) f_N(\vartheta^\dagger|\vartheta_*, \Sigma_*)}\right\},$$

where $f_N(\cdot|\vartheta_*, \Sigma_*)$ is the probability density of $N(\vartheta_*, \Sigma_*)$. If candidate $\vartheta^\dagger$ is rejected, we take the current value ϑ as the next draw. When the Hessian matrix is not negative definite, we may take a flat normal proposal $N(\vartheta_*, c_0 I)$ using some large constant c_0. The obtained draw is denoted as $\alpha^\dagger$.

(ii) Generate $h|\alpha, s, \beta, y \sim \pi^*(h|\alpha, s, \beta, y)$. Given $\alpha = \alpha^\dagger$, we propose a candidate $h^\dagger = (h_1^\dagger, ..., h_n^\dagger)$ using a simulation smoother introduced by de Jong and Shephard (1995) and Durbin and Koopman (2002) for the linear space Gaussian state space model as in (14)-(16). The $h^\dagger$ is a sample from

$$\pi^*(h|\alpha^\dagger, s, \beta, y^*, d) = \frac{\prod_{t=1}^{n} g(y_t^*, h_{t+1}|h_t, \alpha^\dagger, \beta, s_t, d)}{m(y^*|\alpha^\dagger, s, \beta, d)} \times f_N\left(h_1 \middle| \mu^\dagger, \frac{\sigma^{2\dagger}}{1-\phi^{\dagger 2}}\right),$$

(iii) Generate $(\alpha, h) \sim \tilde{\pi}(\alpha, h|s, \beta, y^*, d)$. From (i) and (ii), we have a sample $(\alpha^\dagger, h^\dagger)$ from $\pi^*(h|\alpha, s, \beta, y^*, d)\pi^*(\alpha|s, \beta, y^*, d)$. Let

$$f(y_t, h_{t+1}|h_t, \alpha, \beta)$$
$$= \begin{cases} f_N(y_t|\beta\exp(h_t/2), \exp(h_t)) f_N(h_{t+1}|\overline{h}_t, \sigma^2(1-\phi^2)), & t < n \\ f_N(y_t|\beta\exp(h_t/2), \exp(h_t)), & t = n, \end{cases}$$
$$\overline{h}_t = \mu(1-\phi) + \phi h_t + \rho\sigma\{y_t - \beta\exp(h_t/2)\}\exp(-h_t/2).$$

Given the current value (α, h), accept the candidate $(\alpha^\dagger, h^\dagger)$ with probablity

$$\begin{aligned}
&\min\left\{1, \frac{\tilde{\pi}(\alpha^\dagger, h^\dagger|s,\beta,y)\pi^*(h|\alpha,s,\beta,y^*,d)\pi^*(\alpha|\beta,s,y^*,d)}{\tilde{\pi}(\alpha,h|s,\beta,y)\pi^*(h^\dagger|\alpha^\dagger,s,\beta,y^*,d)\pi^*(\alpha^\dagger|\beta,s,y^*,d)}\right\} \\
&= \min\left\{1, \frac{\pi(\alpha^\dagger, h^\dagger|\beta,y)q(s|h^\dagger,\alpha^\dagger,\beta,y^*)\pi^*(h|\alpha,s,\beta,y^*)\pi^*(\alpha|\beta,s,y^*,d)}{\pi(\alpha,h|\beta,y)q(s|h,\alpha,\beta,y^*)\pi^*(h^\dagger|\alpha^\dagger,s,\beta,y^*)\pi^*(\alpha^\dagger|\beta,s,y^*,d)}\right\} \\
&= \min\left\{1, \frac{q(s|h^\dagger,\alpha^\dagger,\beta,y^*)\prod_{t=1}^n f(y_t,h_{t+1}^\dagger|h_t^\dagger,\alpha^\dagger,\beta)g(y_t^*,h_{t+1}|h_t,\alpha,\beta,s_t,d)}{q(s|h,\alpha,\beta,y^*)\prod_{t=1}^n f(y_t,h_{t+1}|h_t,\alpha,\beta)g(y_t^*,h_{t+1}^\dagger|h_t^\dagger,\alpha^\dagger,\beta,s_t,d)}\right\} \\
&= \min\left\{1, \frac{\prod_{t=1}^n f(y_t|h_t^\dagger,\alpha^\dagger,\beta)\sum_{i=1}^{10}\sum_{j=0}^{2}\tilde{p}_{i,j}g(y_t^*|h_t,\alpha,\beta,s_t=(i,j))}{\prod_{t=1}^n f(y_t|h_t,\alpha,\beta)\sum_{i=1}^{10}\sum_{j=0}^{2}\tilde{p}_{i,j}g(y_t^*|h_t^\dagger,\alpha^\dagger,\beta,s_t=(i,j))}\right\} \\
&= \min\left\{1, \frac{\prod_{t=1}^n f(y_t,h_{t+1}^\dagger|h_t^\dagger,\alpha^\dagger,\beta)\sum_{i=1}^{10}\sum_{j=0}^{2}\tilde{p}_{i,j}g(y_t^*,h_{t+1}|h_t,\alpha,\beta,s_t=(i,j),d)}{\prod_{t=1}^n f(y_t,h_{t+1}|h_t,\alpha,\beta)\sum_{i=1}^{10}\sum_{j=0}^{2}\tilde{p}_{i,j}g(y_t^*,h_{t+1}^\dagger|h_t^\dagger,\alpha^\dagger,\beta,s_t=(i,j),d))}\right\}
\end{aligned}$$

Remark. As in Algorithm 1, we may skip (iii) of Step 3b since the approximation error is usually small.

B.2 Associated particle filter

We describe how to compute the likelihood $f(y|\theta)$ when there is a leverage effect. Let

$$\begin{aligned}
f(y_t|h_t,\theta) &= \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}h_t - \frac{1}{2}\{y_t - \beta\exp(h_t/2)\}^2\exp(-h_t)\right] \\
f(h_{t+1}|h_t,y_t,\theta) &= \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma}\exp\left\{-\frac{(h_{t+1}-\mu_{t+1})^2}{2(1-\rho^2)\sigma^2}\right\}, \\
\mu_{t+1} &= \mu + \phi(h_t-\mu) + \rho\sigma\exp(-h_t/2)\{y_t - \beta\exp(h_t/2)\},
\end{aligned}$$

and consider the importance function for the auxiliary particle filter

$$\begin{aligned}
q(h_{t+1},h_t^i|Y_{t+1},\theta) &\propto f(y_{t+1}|\mu_{t+1}^i,\theta)f(h_{t+1}|h_t^i,y_t,\theta)\hat{f}(h_t^i|Y_t,\theta) \\
&\propto f(h_{t+1}|h_t^i,y_t,\theta)q(h_t^i|Y_{t+1},\theta)
\end{aligned}$$

where

$$\begin{aligned}
q(h_t^i|Y_{t+1},\theta) &= \frac{f(y_{t+1}|\mu_{t+1}^i,\theta)\hat{f}(h_t^i|Y_t,\theta)}{\sum_{j=1}^I f(y_{t+1}|\mu_{t+1}^j,\theta)\hat{f}(h_t^j|Y_t,\theta)}, \\
f(y_{t+1}|\mu_{t+1}^i,\theta) &= \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\mu_{t+1}^i - \frac{1}{2}\{y_t - \beta\exp(h_t^i/2)\}^2\exp(-\mu_{t+1}^i)\right], \\
\mu_{t+1}^i &= \mu + \phi(h_t^i-\mu) + \rho\sigma\exp(-h_t^i/2)\{y_t - \beta\exp(h_t^i/2)\}.
\end{aligned}$$

This leads to the following particle filtering.

1. Compute $\hat{f}(y_1|\theta)$ and $\hat{f}(h_1^i|Y_1,\theta)=\pi_1^i$ for $i=1,\ldots,I$.

   (a) Generate $h_1^i \sim f(h_1|\theta)$ $(= N(\mu,\sigma^2/(1-\phi^2)))$ for $i=1,\ldots,I$.

   (b) Compute

$$\pi_1^i = \frac{w_i}{\sum_{i=1}^I w_i}, \quad w_i = f(y_1|h_1,\theta), \quad W_i = F(y_1|h_1,\theta),$$
$$\hat{f}(y_1|\theta) = \overline{w}_1 = \frac{1}{I}\sum_{i=1}^I w_i, \quad \hat{F}(y_1|\theta) = \overline{W}_1 = \frac{1}{I}\sum_{i=1}^I W_i,$$

   where $f(y_1|\theta)$ and $F(y_1|\theta)$ are the marginal density function and the marginal distribution function of y_1 given θ. Let $t=1$.

2. Compute $\hat{f}(y_{t+1}|\theta)$ and $\hat{f}(h_{t+1}^i|Y_{t+1},\theta)=\pi_{t+1}^i$ for $i=1,\ldots,I$.

   (a) Sample $h_t^i \sim q(h_t|Y_t,\theta)$, $i=1,\ldots,I$.

   (b) Generate $h_{t+1}^i|h_t^i,y_t,\theta \sim f(h_{t+1}|h_t^i,y_t,\theta)$ $(= N(\mu_{t+1}^i,\sigma^2(1-\rho^2)))$ for $i=1,\ldots,I$.

   (c) Compute

$$\pi_{t+1}^i = \frac{w_i}{\sum_{i=1}^I w_i}, \quad w_i = \frac{f(y_{t+1}|h_{t+1}^i,\theta)f(h_{t+1}^i|h_t^i,y_t,\theta)\hat{f}(h_t^i|Y_t,\theta)}{f(h_{t+1}^i|h_t^i,y_t,\theta)q(h_t^i|Y_{t+1},\theta)} = \frac{f(y_{t+1}|h_{t+1}^i,\theta)\hat{f}(h_t^i|Y_t,\theta)}{q(h_t^i|Y_{t+1},\theta)},$$
$$W_i = \frac{F(y_{t+1}|h_{t+1}^i,\theta)\hat{f}(h_t^i|Y_t,\theta)}{q(h_t^i|Y_{t+1},\theta)},$$
$$\hat{f}(y_{t+1}|Y_t,\theta) = \overline{w}_{t+1} = \frac{1}{I}\sum_{i=1}^I w_i, \quad \hat{F}(y_{t+1}|\theta) = \overline{W}_{t+1} = \frac{1}{I}\sum_{i=1}^I W_i.$$

3. Increment t and go to 2.

C MCMC algorithm to compute the posterior ordinate

This section describes MCMC algorithm which may be used when computing the marginal likelihood. It is a little less efficient than Algorithm 3, but still efficient enough to compute the posterior ordinate.

Algorithm 4. The Markov chain Monte Carlo simulation is implemented in four blocks:

1. Initialize h and θ.

2. Generate $\theta|h, y \sim \pi(\theta|h, y)$.

3. Generate $h|\theta, y \sim \pi(h|\theta, y)$.

4. Go to step 2.

Step 2. Generation of $\theta|h, y$

We first transform θ to $\vartheta = (\mu, \log\{(1+\phi)/(1-\phi)\}, \log\sigma^2, \beta, \log\{(1+\rho)/(1-\rho)\})$, to remove parameter constraints, and conduct Metropolis-Hastings (MH) algorithm to sample from the conditional posterior distribution with density $\pi(\vartheta|h, y) = \pi(\theta|h, y) \times |d\theta/d\vartheta|$ where $|d\theta/d\vartheta|$ is the Jacobian of the transformation. Compute the posterior mode $\hat{\vartheta}$ and define ϑ_* and Σ_* as

$$\vartheta_* = \hat{\vartheta}, \quad \Sigma_*^{-1} = -\frac{\partial^2 \log \pi(\vartheta|h, y)}{\partial\vartheta\partial\vartheta'}\bigg|_{\vartheta=\hat{\vartheta}}.$$

Given the current value ϑ, generate a candidate $\vartheta^\dagger$ from the distribution $N(\vartheta_*, \Sigma_*)$ and accept it with probability

$$\alpha(\vartheta, \vartheta^\dagger|h, y) = \min\left\{1, \frac{\pi(\vartheta^\dagger|h, y) f_N(\vartheta|\vartheta_*, \Sigma_*)}{\pi(\vartheta|h, y) f_N(\vartheta^\dagger|\vartheta_*, \Sigma_*)}\right\},$$

where $f_N(\cdot|\vartheta_*, \Sigma_*)$ is the probability density of $N(\vartheta_*, \Sigma_*)$. If candidate $\vartheta^\dagger$ is rejected, we take the current value ϑ as the next draw. When the Hessian matrix is not negative definite, we may take a flat normal proposal $N(\vartheta_*, c_0 I)$ using some large constant c_0.

Step 3. Generation of $h|\theta, y$

We sample h using the mixture sampler using the mixture of normal distributions as discussed in the previous section. Define the pseudo target density

$$\tilde{\pi}(h, s|\theta, y) = \pi(h|\theta, y) \times q(s|h, \theta, y^*, d),$$
$$q(s|h, \theta, y^*, d) = \prod_{t=1}^{n} \frac{\tilde{p}_{s_t} g(y_t^*, h_{t+1}|h_t, \theta, s_t, d)}{\sum_{i=1}^{10}\sum_{j=0}^{2} \tilde{p}_{i,j} g(y_t^*, h_{t+1}|h_t, \theta, s_t = (i, j), d)},$$

where

$$g(y_t^*, h_{t+1}|h_t, \theta, s_t, d) = \begin{cases} f_N(y_t^*|\tilde{m}_{s_t} + h_t, v_{s_t}^2) f_N(h_{t+1}|\overline{h}_{s_t,t}, \sigma^2(1-\rho^2)), & t < n, \\ f_N(y_t^*|\tilde{m}_{s_t} + h_t, v_{s_t}^2) & t = n, \end{cases}$$

$$\overline{h}_{s_t,t} = \mu(1+\phi) + \phi h_t + \rho\sigma[d_t \exp(\tilde{m}_{s_t}/2)\{a_{s_t} + b_{s_t}(y_t^* - h_t - \tilde{m}_{s_t})\} - \beta].$$

We generate sample (h, s) from the pseudo target density in two steps.

(a) Generate $s \sim q(s|h, \theta, y^*, d)$.

(b) Generate $h|\theta, s, y \sim \tilde{\pi}(h|\theta, s, y)$.

i. Propose a candidate $h^\dagger = (h_1^\dagger, ..., h_n^\dagger)$ using a simulation smoother introduced by de Jong and Shephard (1995) and Durbin and Koopman (2002) for the linear space Gaussian state space model as in (14)-(16). The $h^\dagger$ is a sample from

$$\pi^*(h|\theta, s, y^*, d) = \frac{\prod_{t=1}^n g(y_t^*, h_{t+1}|h_t, \theta, s_t, d)}{m(y^*|\theta, s)} \times f_N\left(h_1 \middle| \mu, \frac{\sigma^2}{1-\phi^2}\right),$$

where $m(y^*|\theta, s)$ is a normalizing constant given by

$$m(y^*|\theta, s) = \int \prod_{t=1}^n g(y_t^*, h_{t+1}|h_t, \theta, s_t, d) \times f_N\left(h_1 \middle| \mu, \frac{\sigma^2}{1-\phi^2}\right) dh.$$

ii. Let

$$f(y_t, h_{t+1}|h_t, \theta) = \begin{cases} f_N(y_t|\beta \exp(h_t/2), \exp(h_t)) f_N(h_{t+1}|\overline{h}_t, \sigma^2(1-\rho^2)), & t < n \\ f_N(y_t|\beta \exp(h_t/2), \exp(h_t)), & t = n, \end{cases}$$

$$\overline{h}_t = \mu(1+\phi) + \phi h_t + \rho\sigma\{y_t - \beta \exp(h_t/2)\} \exp(-h_t/2).$$

Given the current value h, accept the candidate $h^\dagger$ with probability

$$\begin{aligned} &\min\left\{1, \frac{\tilde{\pi}(h^\dagger|\theta, s, y)\pi^*(h|\theta, s, y^*, d)}{\tilde{\pi}(h|\theta, s, y)\pi^*(h^\dagger|\theta, s, y^*, d)}\right\} \\ &= \min\left\{1, \frac{\pi(h^\dagger|\theta, y) q(s|h^\dagger, \theta, y^*, d)\pi^*(h|\theta, s, y^*, d)}{\pi(h|\theta, y) q(s|h, \theta, y^*, d)\pi^*(h^\dagger|\theta, s, y^*, d)}\right\} \\ &= \min\left\{1, \frac{q(s|h^\dagger, \theta, y^*, d) \prod_{t=1}^n f(y_t, h_{t+1}^\dagger|h_t^\dagger, \theta) g(y_t^*, h_{t+1}|h_t, \theta, s_t, d)}{q(s|h, \theta, y^*, d) \prod_{t=1}^n f(y_t, h_{t+1}|h_t, \theta) g(y_t^*, h_{t+1}^\dagger|h_t^\dagger, \theta, s_t, d)}\right\} \\ &= \min\left\{1, \frac{\prod_{t=1}^n f(y_t, h_{t+1}^\dagger|h_t^\dagger, \theta) \sum_{i=1}^{10} \sum_{j=0}^2 \tilde{p}_{i,j} g(y_t^*, h_{t+1}|h_t, \theta, s_t = (i,j), d))}{\prod_{t=1}^n f(y_t, h_{t+1}|h_t, \theta) \sum_{i=1}^{10} \sum_{j=0}^2 \tilde{p}_{i,j} g(y_t^*, h_{t+1}^\dagger|h_t^\dagger, \theta, s_t = (i,j), d))}\right\}. \end{aligned}$$