On a Buffered Threshold Autoregressive Stochastic Volatility Model

Qingzheng Wang[1], Ka-Fai Cedric Yiu*[1], and Heung Wong[1]

[1]Department of Applied Mathematics,The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Abstract

This paper introduces a new autoregressive stochastic volatility model with a new piecewise linear structure such that the regime-switching mechanism has a buffer zone where regime-switching is delayed. The proposed model allows us to model the hysteretic phenomenon of the regime-switching existing on both the mean equation and the volatility equation. A full description of the proposed MCMC method is given. In the empirical study, we consider the daily closing prices of NIKKEI stock average, the exchange rate for US Dollar to Japanese Yen and Hang Seng Index. Deviance Information Criterion (DIC) measure shows that our proposed model outperforms the classical threshold stochastic volatility models.

Keywords: Bayesian inference, Kalman filter, Stochastic volatility, Threshold estimation, Buffer zone

1 Introduction

Volatility modelling is a prevailing and ubiquitous topic in both academic research and financial industry. In the literature, there are numerous models for forecasting the volatility. Shephard [1] categorized volatility models into two groups. One is the observation-driven model including the ARCH model, GARCH model and so forth. The other is the parameter-driven model including the stochastic volatility (SV) model. With theoretical and practical development, many generalization of ARCH type and SV type models are introduced. Tsay [2] gave a comprehensive explanation of the stochastic volatility (SV) model which was proposed by Taylor [3]. In this paper, we consider a new SV model which can capture the non-linear and asymmetric properties of volatility. The mean and variance equation of the fundamental SV model are defined by

$$\begin{aligned} \text{mean equation:} \quad y_t &= \exp(h_t/2)u_t, \\ \text{variance equation:} \quad h_t &= \phi_0 + \phi_1 h_{t-1} + \eta_t, \end{aligned}$$

where y_t is the observation. The random noises u_t and η_t are assumed to be independent of one another. Usually, u_t follows a normal distribution with mean zero and variance one. η_t follows a normal distribution with mean μ_h and variance σ_h^2. Instead of directly formulating the variance process, in the SV model, the log-volatility is specified by a first-order autoregressive process. This setting makes sure the positiveness of the variance. If $|\phi_1| < 1$, the log-volatility, h_t, is a stationary process. Under this mild condition, the unconditional mean $\mathbb{E}[h_t]$, which equals $\phi_0/(1-\phi_1)$, and the unconditional variance $\mathrm{Var}(h_t)$, which equals $\sigma_h^2/(1-\phi_1^2)$, are directly obtained by using the definition of the autoregressive model.

The non-linear behavior in financial time series has attracted a lot of attention. The major characteristics of non-linear time series are the jump phenomena and asymmetric behaviour which cannot be described by a linear model. Tong [4] asserted that a piecewise linear structure partitioned by a threshold variable was able to mimic these non-linear characteristics. The change between different linear structures is called regime switching. Following Tong's seminal work, Nelson [5] proposed the exponential GARCH model in which a regime switching framework was first brought into the time-varying volatility model. So *et. al.*[6] introduced the threshold autoregressive stochastic volatility model allowing the existence of regime switching

in the SV model. So and Choi [7] generalized the threshold autoregressive stochastic volatility model to a multi-dimensional case. Chen *et. al.* [8] gave a review on the development of the threshold time-varying volatility model. While threshold heteroscedastic time series models have achieved a great success, it is noticed that threshold models usually have weak performance when the value of the threshold variable is close to the boundaries between different regimes. In order to improve the performance of the classical threshold model, Li *et. al.* [9] and Zhu *et. al.* [10] introduced hysteresis to the regime-switching structure which enjoys the piecewise linear structure of a threshold model but has a more flexible regime switching mechanism. Note that the standard threshold model is a special case of the buffered model when the buffered zone degenerates to a single value. Lo *et. al.*[11] and Zhu *et. al.* [12] suggested the buffered threshold GARCH model to investigate the financial time series based on the idea in [9]. Chen and Truong [13] proposed a Bayesian method to estimate the buffered threshold GARCH model. Truong and Chen [14] and Chen *et. al.* [15] introduced the hysteretic integer-valued GARCH model which applied the hysteretic regime-switching structure to the integer-valued time series. Chen *et. al.* [16] extended the buffered threshold GARCH model to the multi-dimensional time series with the consideration of the dynamic conditional correlation structure. Chen *et. al.*[17] improved the multivariate buffered threshold GARCH model via the investigation of the leverage effect in financial assets. To our best knowledge, this paper is the first to investigate the hysteretic regime-switching property in the autoregressive stochastic volatility model.

For both classical and buffered threshold, the regime switching can be characterized by the step functions which are shown as follows:

$$\underbrace{s_t = \begin{cases} 1, & \text{if } z_t \geq r_{thres}, \\ 0, & \text{if } z_t < r_{thres}, \end{cases}}_{\text{Classical threshold}} \qquad \underbrace{s_t = \begin{cases} 1, & \text{if } z_t \geq r_U, \\ s_{t-1}, & \text{if } r_L \leq z_t < r_U, \\ 0, & \text{if } z_t < r_U, \end{cases}}_{\text{Buffered threshold}} \tag{1}$$

where s_t is the regime indicator, z_t is the threshold variable, r_{thres} is the threshold value in the classical threshold model, while r_L and r_U are the lower bound and upper bound of a buffer zone for the threshold. If $r_L = r_U$, the buffered threshold degenerates into the classical threshold due to the disappearance of the buffer zone. Therefore, the buffered threshold model

is a generalization of the classical threshold model. If $r_L < r_U$, the buffer zone exists. The regime-switching is triggered when the the time series value impales both the lower bound and upper bound of the buffered zone. To be specific, the regime indicator s_t equals 1 when the threshold variable $z_t \geq r_U$; the regime indicator is equal to 0 if $z_t < r_U$ and s_t is unchanged as long as z_t falls in $[r_L, r_U)$. Since the regime-switching might not be immediately triggered when one threshold value is impaled, the buffer zone is able to describe the hysteretic phenomenon for the regime-switching. In this case, regime-switching can be considered as an analogy of the transformation from the quantitative change to qualitative change wherein the width of the buffer zone can be recognized as the minimum requirement of the accumulated quantitative change. The larger the width of the buffer zone, the more accumulated quantitative change is required. Besides that, hysteretic regime-switching can mitigate the side-effect of the intrinsic error of the estimate of the threshold values.

In this paper, we propose a new model called buffered threshold stochastic volatility (BARSV) model which does not only mitigate the side-effect of the estimation error of the threshold variable but also captures the hysteretic regime switching on both mean and volatility equations. The latent structure of the stochastic volatility model and the discontinuity of the likelihood caused by the buffer zone lead to great difficulty in the frequentist approach, thus we employ the Bayesian method for parameter estimation. More specifically, Kalman filter is adopted to estimate the log-volatility h_t which is an unobserved variable. Random Walk Metropolis-Hastings (RM-MH) algorithm is used to estimate the lower and upper bound of the buffer zone simultaneously. Since the choice of a proposal distribution is crucial for the convergence of the Markov chain associated with the RM-MH algorithm, an adaptive proposal distribution is utilized in the burn-in stage. Due to the unknown dependent relationship between r_L and r_U, a stochastic sampling scheme is deployed on the RW-MH algorithm. Gibbs sampling is applied to estimate the rest of the parameters. The empirical study of the BARSV model employed the daily closing prices of NIKKEI stock average, the exchange rate for US Dollar to Japanese Yen and Hang Seng Index from 1$^{\text{st}}$ January 2006 to 31$^{\text{st}}$ December 2014. The deviance information criterion is adopted to compare the BARSV model and other threshold time series models. The results show that the BARSV model outperforms others threshold models with a more flexible threshold structure.

The rest of this paper is organized as follows. The model specification

and estimation process are given in Section 2. The simulation studies are demonstrated in Section 3. The empirical study is implemented in Section 4. Conclusion is presented in Section 5.

2 Model and Estimation

Table 1 shows the variables in the definition of the BARSV model. The BARSV model is defined as

$$r_t = \begin{cases} \theta_1 + \theta_2 r_{t-1} + y_t, & \text{if } s_t = 1, \\ \theta_3 + \theta_4 r_{t-1} + y_t, & \text{if } s_t = 0, \end{cases} \tag{2a}$$

$$y_t = \exp(h_t/2)u_t, \quad u_t \sim N(0,1), \tag{2b}$$

$$h_t = \begin{cases} \beta_1 + \beta_2 h_{t-1} + \eta_t, & \text{if } s_t = 1, \\ \beta_3 + \beta_4 h_{t-1} + \eta_t, & \text{if } s_t = 0, \end{cases} \tag{2c}$$

$$\eta_t \sim \begin{cases} N(0, \sigma_1^2), & \text{if } s_t = 1, \\ N(0, \sigma_2^2), & \text{if } s_t = 0, \end{cases} \tag{2d}$$

$$s_t = \begin{cases} 1, & \text{if } z_t \geq r_U, \\ s_{t-1}, & \text{if } r_L \leq z_t < r_U, \\ 0, & \text{if } z_t < r_L, \end{cases} \tag{2e}$$

where r_t is observable; the log-volatility, h_t, can be recognized as the hidden state at time t; the threshold variable, z_t, can be derived by a previous observation, i.e., $z_t = r_{t-1}$, or a weighted summation of previous observations. The white-noise processes, $\{u_t\}$ and $\{\eta_t\}$, are independent of one another. $\Theta_1 = (\theta_1, \theta_2)'$ and $\Theta_2 = (\theta_3, \theta_4)'$ are the vectors of parameters in the mean equation (2a) for Regime 1 and 2, respectively. $B_1 = (\beta_1, \beta_2)'$ and $B_2 = (\beta_3, \beta_4)'$ are the vectors of parameters in the variance equation (2c) for Regime 1 and 2, respectively. σ_1^2 and σ_2^2 are the variances of the innovation process for Regime 1 and 2. The buffer zone is denoted by $[r_L, r_U)$. It is noted that the proposed model can be extended to a general s-regime model, i.e., $s_t = j$ where $j = 1, 2, \ldots, s$, and $-\infty = r_{U_0} < r_{L_1} < r_{U_1} <$

$r_{L_2} < r_{U_2} \cdots < r_{L_s} < r_{U_s} < r_{L_{s+1}} = \infty$. The total $s-1$ buffer zones are $[r_{L_k}, r_{U_k})$ with $k = 1, 2, \ldots, s$. Assume the regime indicator s_t is equal to j and $r_{U_{j-1}} \leq z_{t-1} < r_{L_j}$, the times series will stay in the same regime at time $t+1$ if the threshold variable z_t increases to the buffer zone $[r_{L_j}, r_{U_j})$ or decreases to the buffer zone $[r_{L_{j-1}}, r_{U_{j-1}})$.

Following a suggestion in Kim *et. al.* [18], Equation (2b) can be transformed to a linear equation as $\log y_t^2 = h_t + \log u_t^2$, where $\log u_t^2$ follows the $\log \chi_1^2$ distribution. To put this model into a Gaussian state space form, Harvey *et. al.* [19] proposed a normal distribution with mean -1.27 and variance $\pi^2/2$ to approximate the $\log \chi^2$ distribution. To improve the fit of the $\log \chi^2$ distribution, Kim *et. al.* [18] proposed the 7-component mixture normal distribution to replace the one in Harvey *et. al.* [19]. In this paper, we adopt a 10-component mixture normal distribution, which was introduced by Omori *et. al.* [20], to tightly approximate the density of the $\log \chi^2$ distribution. Equation (2b) is formulated as

$$
\begin{aligned}
&\log y_t^2 = h_t + \epsilon_t, \\
&\epsilon_t | \rho_t \sim N(\mu_{\rho_t}, \tau_{\rho_t}^2), \\
&\Pr(\rho_t = i | \log y_t^2, h_t) \propto q_i f(\log y_t^2 | \mu_i, \tau_i^2), \\
&i \in \mathrm{I}_{10} \\
&\mathrm{I}_{10} = \{1, 2, \ldots, 10\}
\end{aligned}
$$

where q_i, μ_i and τ_i^2 are given by Omori *et. al.* [20]. ρ_t is the indicator variable denoting which Gaussian distribution ϵ_t belongs to. The prior distribution of ρ_t is a multinomial distribution specified by q_i. Given the observation value, y_t, and h_t which is drawn from the last MCMC iteration, ρ_t can be drawn directly from its posterior distribution.

2.1 Bayesian analysis

Let $R_T = (r_1, r_2, ..., r_T)^{'}$, $H_T = (h_1, h_2, ..., h_T)^{'}$ and $S_T = (s_1, s_2, ..., s_T)^{'}$ be the vector of the time series $\{r_t\}$, $\{h_t\}$ and $\{s_t\}$ for $t \geq 1$, respectively. Let $\gamma = (\theta_1, \theta_2, \theta_3, \theta_4, \beta_1, \beta_2, \beta_3, \beta_4, \sigma_1^2, \sigma_2^2)'$ be the vector of the regular parameters which can be segmented into $\Theta_1^{'}, \Theta_2^{'}, B_1^{'}, B_2^{'}, \sigma_1^2$ and σ_2^2. $\omega = (H_T^{'}, \gamma^{'}, r_0, r_L, r_U)^{'}$ is the variable of our interest. ω_i denotes one of the element in ω. ω_{-i} denotes the full conditional set of ω without ω_i. By the Bayes theorem, the full conditional distribution of ω_i is (See So *et. al.* [6] as well

as So and Choi [7])

$$
\begin{aligned}
f(\omega_i|\omega_{-i}, R_T) &= \tfrac{f(\omega_i,\omega_{-i},R_T)}{f(\omega_{-i},R_T)} = \tfrac{f(\omega,R_T)}{f(\omega_{-i},R_T)} \\
&\propto f(\omega, R_T) = f(r_0, R_T, H_T, \gamma, r_L, r_U) \\
&= f(r_T, h_T, r_0, \ldots, r_{T-1}, h_1, \ldots, h_{T-1}, \gamma, r_L, r_U) \\
&= \tfrac{f(r_T,h_T,r_0,\ldots,r_{T-1},h_1,\ldots,h_{T-1},\gamma,r_L,r_U)}{f(r_0,\ldots,r_{T-1},h_1,\ldots,h_{T-1},\gamma,r_L,r_U)} \times \tfrac{f(r_{T-1},h_{T-1},r_0,\ldots,r_{T-2},h_1,\ldots,h_{T-2},\gamma,r_L,r_U)}{f(r_0,\ldots,r_{T-2},h_1,\ldots,h_{T-2},\gamma,r_L,r_U)} \\
&\quad \times \cdots \times \tfrac{f(r_3,h_3,r_0,r_1,r_2,h_1,h_2,\gamma,r_L,r_U)}{f(r_2,h_2,r_0,r_1,h_1,\gamma,r_L,r_U)} \times \tfrac{f(r_2,h_2,r_0,r_1,h_1,\gamma,r_L,r_U)}{f(r_0,r_1,h_1,\gamma,r_L,r_U)} \times f(r_0, r_1, h_1, \gamma, r_L, r_U) \\
&= \textstyle\prod_{t=2}^{T} f(r_t, h_t|r_0, R_{t-1}, H_{T-1}, \gamma, r_L, r_U) \\
&\quad \times f(r_1|r_0, h_1, \gamma, r_L, r_U) f(h_1|r_0, \gamma, r_L, r_U) f(r_0, \gamma, r_L, r_U) \\
&= \textstyle\prod_{t=2}^{T} f(r_t, |h_t, r_0, R_{t-1}, H_{T-1}, \gamma, r_L, r_U) f(h_t|r_0, R_{t-1}, H_{T-1}, \gamma, r_L, r_U) \\
&\quad \times f(r_1|r_0, h_1, \gamma, r_L, r_U) f(h_1|r_0, \gamma, r_L, r_U) f(r_0) f(\gamma) f(r_L, r_U) \\
&= \textstyle\prod_{t=2}^{T} f(r_t, |h_t, r_{t-1}, \gamma, r_L, r_U) f(h_t|h_{t-1}, \gamma, r_L, r_U) \\
&\quad \times f(r_1|r_0, h_1, \gamma, r_L, r_U) f(h_1|r_0, \gamma, r_L, r_U) f(r_0) f(\gamma) f(r_L, r_U) \\
&= \textstyle\prod_{t=1}^{T} f(r_t, |h_t, r_{t-1}, \gamma, r_L, r_U) \prod_{t=2}^{T} f(h_t|h_{t-1}, \gamma, r_L, r_U) \\
&\quad \times f(h_1|r_0, \gamma, r_L, r_U) f(r_0) f(\gamma) f(r_L, r_U),
\end{aligned}
\tag{3}
$$

where $f(r_L, r_U)$ and $f(\gamma)$ denote the p.d.f. of the prior distribution of the buffer zone and regular parameters, respectively. Rather than assuming the threshold as an unknown scalar in So *et. al.* [6] as well as So and Choi [7], we consider the threshold as a hysteretic structure which is determine by variables r_L and r_U. The prior distribution of the buffer zone is assumed to be a bivariate uniform distribution, which has been investigated by Chen and Truong [13], is

$$\{r_L, r_U : (r_L, r_U) \in [Q_a, Q_c] \times [Q_b, Q_d] \quad and \quad r_L \leq r_U\}.$$

Q_a, Q_b, Q_c and Q_d denote respectively the $100a$-th, $100b$-th, $100c$-th and $100d$-th percentile of R_T, where $a \leq b < c \leq d$.

Given the buffer zone $[r_L, r_U)$, the regime indicator s_t of the buffered

threshold in Equations (1) can be defined by

$$I(s_t = 1) = I(z_t \geq r_U) + I(r_L \leq z_t < r_U)I(s_{t-1} = 1),$$
$$I(s_t = 0) = I(z_t < r_L) + I(r_L \leq z_t < r_U)I(s_{t-1} = 0),$$

where $I(\cdot)$ denotes the indicator function. Given h_{t-1}, r_{t-1}, γ and s_t, the one-step-ahead prediction distribution of h_t is

$$N\big(\mu_{h_t}, \sigma^2_{h_t}\big),$$

where

$$\mu_{h_t} = (\beta_1 + \beta_3 h_{t-1})\mathrm{I}(s_t = 1) + (\beta_2 + \beta_4 h_{t-1})\mathrm{I}(s_t = 0),$$
$$\sigma^2_{h_t} = \sigma^2_1 \mathrm{I}(s_t = 1) + \sigma^2_2 \mathrm{I}(s_t = 0).$$

The conditional p.d.f. of h_t is

$$f(h_t|r_{t-1}, h_{t-1}, \gamma, r_L, r_U) = \frac{1}{\sqrt{2\pi\sigma^2_{h_t}}} \exp\big\{ - \frac{(h_t - \mu_{h_t})^2}{2\sigma^2_{h_t}}\big\}. \tag{4}$$

Given h_t, r_{t-1}, γ and s_t, the one-step-ahead prediction distribution of r_t is

$$N\big(\mu_{r_t}, \exp(h_t)\big),$$

where

$$\mu_{r_t} = (\theta_1 + \theta_3 r_{t-1})\mathrm{I}(s_t = 1) + (\theta_2 + \theta_4 r_{t-1})\mathrm{I}(s_t = 1).$$

The conditional p.d.f. of r_t is

$$f(r_t|h_t, r_{t-1}, \gamma, r_L, r_U) = \frac{1}{\sqrt{2\pi \exp(h_t)}} \exp\big\{ - \frac{(r_t - \mu_{r_t})^2}{2\exp(h_t)}\big\}. \tag{5}$$

Thus, the log-likelihood function corresponding to the BARSV model can be expressed as

$$\begin{aligned} l(R_T, H_T, \gamma, r_0, r_{thres}) \quad &= \textstyle\sum_{t=2}^{T} \mathrm{I}(s_t = 1)[\log f(r_t, |h_t, r_{t-1}, \gamma) + \log f(h_t|\rho_t, h_{t-1}, \gamma)] \\ &\quad + \mathrm{I}(s_1 = 1)[\log f(r_1, |h_1, r_0, \gamma) + \log f(h_1|r_0, \gamma)] \\ &\quad + \textstyle\sum_{t=2}^{T} \mathrm{I}(s_t = 0)[\log f(r_t, |h_t, r_{t-1}, \gamma) + \log f(h_t|\rho_t, h_{t-1}, \gamma)] \\ &\quad + \mathrm{I}(s_t = 0)[\log f(r_1, |h_1, r_0, \gamma) + \log f(h_1|r_0, \gamma)] \\ &\quad + \log f(r_0) + \log f(\gamma) + \log f(r_L, r_U). \end{aligned} \tag{6}$$

Define

$$G_{1,1} = \log f(r_1|h_1, r_0, \gamma) + \log f(h_1|r_0, \gamma),$$
$$G_{1,t} = \log f(r_t|h_t, r_{t-1}, \gamma) + \log f(h_t|\rho_t, h_{t-1}, \gamma), \text{ for } t = 2, 3, \ldots, T.$$

An alternative expression of Equation (6) is

$$l(R_T, H_T, \gamma, r_0, r_L, r_U) = \sum_{t=1}^{T} \mathrm{I}(s_t = 1) G_{1,t} + \sum_{t=1}^{T} \mathrm{I}(s_t = 0) G_{1,t} + \log f(r_0) + \log f(\gamma) + \log f(r_L, r_U).$$

The partial derivatives of $l(R_T, H_T, \gamma, r_0, r_L, r_U)$ in the direction r_L and r_U are

$$\frac{\partial l(R_T, H_T, \gamma, r_0, r_L, r_U)}{\partial r_L} = \begin{cases} \text{not defined}, & \text{if } r_L = r_0, r_1, \ldots, r_{T-1}, \\ 0, & \text{otherwise;} \end{cases}$$

$$\frac{\partial l(R_T, H_T, \gamma, r_0, r_L, r_U)}{\partial r_U} = \begin{cases} \text{not defined}, & \text{if } r_U = r_0, r_1, \ldots, r_{T-1}, \\ 0, & \text{otherwise.} \end{cases}$$

It is noted that the underlying model has T step discontinuities in the log-likelihood function along the direction of either r_L or r_U. From a realistic point of view, changes in either r_L or r_U could cause at least one pair of r_t and h_t to switch from one regime to another. For the same observation set, regime-switching leads to a discontinuous change in the probability density such that the log-likelihood function of the BARSV model has a discrete shift. In this case, it is difficult to implement the gradient-based estimation method, such as MLE, to search for the optimal value for the buffer zone. Therefore, the Bayesian method is adopted to estimate the threshold value in the BARSV model.

Since $\{h_t\}$ is unobserved and the size of $\{h_t\}$ is increasing over time, it is natural to compute the likelihood function of the BARSV model by integrating out H_T, i.e.

$$L(r_1, \ldots, r_T, \gamma) = \int_{h_1} \cdots \int_{h_T} \prod_{t=1}^{T} f(r_t|h_t, r_{t-1}, \gamma) \prod_{t=2}^{T} f(h_t|\rho_t, h_{t-1}, \gamma) \times f(h_1|r_0, \gamma) f(r_0) f(\gamma)\, dh_1 \ldots dh_T.$$

However, it is impossible to directly evaluate the T dimensional integral because of its complicated and intractable from. Therefore, it is necessary to do a careful, tractable and discernable study in handling the unobserved variable, h_t.

2.2 Implementation of the MCMC technique

Due to the latent structure of the SV family models (See So [21]) and the change-point problem in the proposed model (See Chen and Lee [22]), we prefer the MCMC method which is relatively easy to implement compared to others (See Broto and Ruiz [23]). The main attraction of the MCMC method is that it splits a high dimensional problem into low dimensional tasks. In the framework of the MCMC method, Markov chains are created on the block of the unknown parameters and samples are drawn from the associated posterior distributions. The MCMC algorithm for the BARSV model is sketched out in Table 2, followed by a detailed description of each step.

Step 1-4: Implement Gibbs sampler sampling Θ_1, Θ_2, B_1, B_2, σ_1^2, σ_2^2 and r_0. In Gibbs sampler, samples are drawn from the full conditional distributions which is directly derived from Equation (3) after specifying the prior distributions. All related posterior distributions are shown in the Appendix.

Step 5: Draw ρ_t from a multinomial distribution. The detailed sampling scheme is shown below.

At iteration n:

- Calculate y_t, $t = 1, 2, \ldots, T$, by

  $$y_t = r_t - \theta_1^{(n)} * s_t^{(n-1)} - \theta_3^{(n)} * (1 - s_t^{(n-1)}) - (\theta_2^{(n)} * s_t^{(n-1)} + \theta_4^{(n)} * (1 - s_t^{(n-1)})) * r_{t-1}$$

  where $\theta_1^{(n)}, \theta_2^{(n)}, \theta_3^{(n)}$ and $\theta_4^{(n)}$ are drawn from the current iteration; $s_t^{(n-1)}$ is the regime indicator for r_t drawn from the previous iteration.

- Calculate $\log u_t^2 = \log y_t^2 - h_t^{(n-1)}$, where $h_t^{(n-1)}$ denotes the sample of h_t drawn from the previous iteration.

- Calculate $P_{\rho_t=i} = q_i * \frac{1}{\sqrt{2*\pi}\tau_i} \exp\{\frac{(\log u_t^2 - \mu_i)^2}{2\tau_i^2}\}$, $i \in \{1, 2, \ldots, 10\}$.

- Normalize $P_{\rho_t=i}$ by $P^*_{\rho_t=i} = \frac{P_{\rho_t=i}}{\sum_{i=1}^{10} P_{\rho_t=i}}$.

- Draw a random sample, u, from the uniform distribution, $U(0,1)$. If $u \in (\sum_{i=0}^{i-1} P^*_{\rho_t=i}, \sum_{i=0}^{i} P^*_{\rho_t=i}]$ where $P^*_{\rho_t=0} = 0$, let $\rho_t = i$. As a result, $\log u_t^2$ is approximated by $N(\mu_i, \tau_i^2)$.

Step 6-7: Implement both the Random-walk Metropolis-Hastings (RW-MH) algorithm and Kalman filter sampling r_L, r_U and h_t. In the RW-MH algorithm, it is not necessary to have a clear picture on the marginal distribution or the full conditional distribution of the parameter of interest because it solely works with the joint distribution (Equation (3)). If it is not easy to generate samples from the joint distribution, an alternative distribution, which is easy to sample from, is deployed to generate random samples. In order to compensate for the difference between the target distribution and the sampling distribution, an acceptance ratio is introduced to fulfill the detailed balance equation of the Metropolis-Hastings algorithm. Under the acceptance-rejection scheme, large moves would be rejected and they slow down the chain, but small accepted moves do not help much for exploring the parameter space. Therefore, it is important to get a reasonable acceptance rate such that more proper moves can be found. The optimal acceptance rate should be between 0.234 and 0.44 suggested by Roberts *et. al.* [24] as well as Roberts and Rosenthal [25]. With the increase of the dimension of proposal distribution, the optimal acceptance rate is decreased. A simple way to get the optimal acceptance rate is to tune the variance of the proposal distribution. In the burn-in stage, we implement an adaptive proposal distribution whose variance is changed based on the historical acceptance rate. Let N_0 denote the iteration number in the burn-in stage. Let σ_{MH1} and σ_{MH2} denote the step sizes associated with r_L and r_U, respectively, in the RW-MH algorithm. The burn-in stage is splitted into k batches such that each batch contains N_0/b iterations. The details of implementing the adaptive proposal distribution are shown as follows.

- Initialize $\sigma^{(0)}_{\text{MH1}}$ and $\sigma^{(0)}_{\text{MH2}}$.
- Count c, the number of accepted candidate samples, in the n_b^{th} batch where $n_b = 1, 2, \ldots, k$.
- Calculate the acceptance rate, $\frac{c}{N_0/b}$, in terms of the n_b^{th} batch.
- If $\frac{c}{N_0/b} > 0.44$, update $\sigma^{(n_b+1)}_{\text{MH1}} = \sigma^{(n_b)}_{\text{MH1}} * 1.1$ and $\sigma^{(n_b+1)}_{\text{MH2}} = \sigma^{(n_b)}_{\text{MH2}} * 1.1$ on the $(n_b+1)^{\text{th}}$ batch.

- If $\frac{c}{N_0/b} < 0.234$, update $\sigma_{\text{MH1}}^{(n_b+1)} = \sigma_{\text{MH1}}^{(n_b)} * 0.9$ and $\sigma_{\text{MH2}}^{(n_b+1)} = \sigma_{\text{MH2}}^{(n_b)} * 0.9$ on the $(n_b+1)^{\text{th}}$ batch.

Besides that, we use a stochastic sampling scheme for r_L and r_U due to the intrinsic dependence between those two parameters. In the context of MCMC method, estimation and inference methods of unobserved variable, h_t, can be classified into two algorithms. The first one is the single-move sampler proposed by Jacquier *et. al.* [26]. Another algorithm is the multi-move sampler proposed by Shephard and Pitt [27], followed by So *et. al.* [6] and Tsang [28], and further developed by Watanabe and Omori [29]. Since the single-move sampler generates high level correlation from the drawn samples of h_t, we apply the multi-move sampler to draw h_t based on the standard Kalman filter proposed by Kalman [30]. The detailed sampling scheme is illustrated below and equations of Kalman filtering are summarized in Equations (8).

- Using a diffuse prior for the initial state h_0, set $a_1 = \log y_1^2 - \mu_{\rho_1}$ and $p_1 = \tau_{\rho_1}^2$.
- Using the prediction function of Kalman filter (Equation (8a) and (8b)), calculate $a_{t|t-1}$ and $p_{t|t-1}$.
- Using the filtering function of Kalman filter (Equation (8c) and (8d)), calculate $a_{t|t}$ and $p_{t|t}$.
- Calculate the quasi-likelihood of BARSV model based on the results of prediction function of Kalman filter. i.e.

$$f(r_L, r_U|R_T, \gamma) \propto \prod_{t=1}^{T} \left[F_t^{-1/2} \exp\left\{ -\frac{V_t^2}{2F_t} \right\} \right] f(r_L, r_U), \tag{7}$$

  where both $V_t = \log y_t^2 - \mu_{\rho_t} - a_{t|t-1}$ and $F_t = p_{t|t-1} + \sigma_{\rho_t}^2$ are derived by Kalman filter directly and $f(r_L, r_U)$ is the p.d.f. of the joint prior distribution of r_L and r_U.

- Draw a random sample, b, from the binomial distribution $B(1, 0.5)$.
  - If $b = 1$, generate r_L^* using $r_L^* = r_L^{(t-1)} + \sigma_{\text{MH1}} u_t$ in which u_t follows a truncated normal distribution $N(0,1)\text{I}(\frac{Q_a - r_L^{(t-1)}}{\sigma_{\text{MH1}}} < u_t <$

$\frac{Q_c - r_L^{(t-1)}}{\sigma_{\text{MH1}}}$). After that, we generate r_U^* using $r_U^* = r_U^{(t-1)} + \sigma_{\text{MH2}} v_t$ in which v_t follows a truncated normal distribution $N(0,1)\text{I}\big(\frac{\max(r_L^*, Q_b) - r_U^{(t-1)}}{\sigma_{\text{MH2}}} < v_t < \frac{Q_d - r_U^{(t-1)}}{\sigma_{\text{MH2}}}\big)$.

– If $b = 0$, generate r_U^* using $r_U^* = r_U^{(t-1)} + \sigma_{\text{MH2}} v_t$ in which v_t follows a truncated normal distribution $N(0,1)\text{I}(\frac{Q_b - r_U^{(t-1)}}{\sigma_{\text{MH2}}} < v_t < \frac{Q_d - r_U^{(t-1)}}{\sigma_{\text{MH2}}})$. After that, we generate r_L^* using $r_L^* = r_L^{(t-1)} + \sigma_{\text{MH1}} u_t$ in which u_t follows a truncated normal distribution $N(0,1)\text{I}\big(\frac{Q_a - r_L^{(t-1)}}{\sigma_{\text{MH1}}} < u_t < \frac{\min(r_U^*, Q_c) - r_L^{(t-1)}}{\sigma_{\text{MH1}}}\big)$.

Both $r_L^{(t-1)}$ and $r_U^{(t-1)}$ are the samples derived from previous iterations.

- Generate s_t^*, $t = 1, 2, \ldots, T$, by substituting r_L^* and r_U^* into Equation (2e).
- Calculate $a_{t|t-1}^*$, $p_{t|t-1}^*$, $a_{t|t}^*$ and $p_{t|t}^*$ by substituting s_t^* into the prediction and filtering function of Kalman filter.
- Using Equation (7), Calculate a new quasi-likelihood of BARSV model in terms of r_L^* and r_U^*.
- Calculate the acceptance ratio a_r by

$$a_r = \min\left[1, \frac{f(r_L^{(*)}, r_U^{(*)})|R_T, \gamma)}{f(r_L^{(t-1)}, r_U^{(t-1)})|R_T, \gamma)}\right].$$

.

- Draw a random sample, u, from the uniform distribution, $U(0,1)$.
  - If $u \leq a_r$, let $r_L^{(t)} = r_L^*$, $r_U^{(t)} = r_U^*$, $s_t = s_t^*$, $a_{t|t-1} = a_{t|t-1}^*$, $p_{t|t-1} = p_{t|t-1}^*$, $a_{t|t} = a_{t|t}^*$ and $p_{t|t} = p_{t|t}^*$.
  - Otherwise, let $r_L^{(t)} = r_L^{(t-1)}$ and $r_U^{(t)} = r_U^{(t-1)}$.
- Using the smoothing function of Kalman filter (Equation (8f) and (8g)), calculate $a_{t|t+1}$ and $p_{t|t+1}$.
- Sample h_T from $N(a_{T|T}, p_{T|T})$ and sample $h_t, t = T-1, \ldots, 1$, from $N(a_{t|t+1}, p_{t|t+1})$.

$$a_{t|t-1} = \phi_{1,t} + \phi_{2,t} a_{t-1|t-1}, \tag{8a}$$
$$p_{t|t-1} = \phi_{2,t}^2 p_{t-1|t-1} + \sigma_{h_t}^2, \tag{8b}$$
$$a_{t|t} = a_{t|t-1} + KG_t(\log y_t^2 - \mu_{\rho_t} - a_{t|t-1}), \tag{8c}$$
$$p_{t|t} = (1 - KG_t)^2 p_{t|t-1} + KG_t^2 \tau_{\rho_t}^2, \tag{8d}$$
$$KG_t = \frac{p_{t|t-1}}{p_{t|t-1} + \tau_{\rho_t}^2} \tag{8e}$$
$$a_{t|t+1} = a_{t|t} + \frac{\phi_{2,t+1} p_{t|t}}{p_{t+1|t}}(h_{t+1} - a_{t+1|t}), \tag{8f}$$
$$p_{t|t+1} = p_{t|t} - \frac{\phi_{2,t+1}^2 p_{t|t}^2}{p_{t+1|t}}, \tag{8g}$$
$$\text{for } t = 1, 2, \ldots, T-1.$$

where $\phi_{1,t} = \beta_1 s_t + \beta_3(1-s_t)$, $\phi_{2,t} = \beta_2 s_t + \beta_4(1-s_t)$ and $\sigma_{h_t}^2 = \sigma_1^2 s_t + \sigma_2^2(1-s_t)$. KG_t stands for the Kalman gain at time t which makes a trade-off between the mean equation and the variance equation.

2.3 Forecast of future returns

Given the estimated parameters, BARSV model can predict the return directly. The one-step-ahead prediction at time t is demonstrated below.

- Based on the filtering function of Kalman filter (Equation (8c) and (8d)), draw volatility samples h_t^i, $i = M+1, M+2, \ldots, M+N$, from $N(a_{t|t}, p_{t|t})$ where M is the iteration number of the burning stage and N is the iteration number of the sampling stage.
- For each h_t^i, draw the prediction h_{t+1}^i from $f(h_{t+1}|r_t, h_t, \gamma, r_L, r_U)$ in Equation (4).
- Calculate the smoothing log-volatilities (or hidden state) by $\frac{1}{N}\sum_{i=M+1}^{M+N} h_{t+1}^i$
- Draw the one-step-ahead predicted return, r_{t-1}, from $f(r_{t+1}|h_{t+1}, r_t, \gamma, r_L, r_U)$ in Equation (5).

3 Simulation studies

In this section, we demonstrate the performance of our proposed Bayesian method with a simulation study. Data are generated by 6 different parameter

sets which are shown in Table 3. Without loss of generality, the threshold variable z_t is assumed to be a lag variable of the observed time series, r_{t-1}. Note that, in our simulation experiment, we try with many different parameters in the mean equation, the variance equation, and the threshold. We also try with different sample sizes and number of replications.

To glean some idea of the sampling behaviour of the MCMC estimator for this model, the replication number is chosen as 100, 300 and 500 in Trial 1. Unless otherwise stated, each trail contains 500 replications and the sample size in each replication is set to 4000. For the experiment of each replication, the total number of iterations is 1.5×10^5 and the number of the burn-in iteration is 5×10^4. The thinning parameter is 10. Finally, we got 1×10^4 drawn samples to estimate the posterior distribution for each parameter in terms of one replication and got 500 posterior estimates for each parameter in total.

The flat priors for parameters are specified as follows:

$$\Theta_1 \sim N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}10 & 0\\0 & 10\end{pmatrix}\right] I(|\theta_2| < 1), \qquad \Theta_2 \sim N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}10 & 0\\0 & 10\end{pmatrix}\right], I(|\theta_4| < 1)$$

$$B_1 \sim N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}10 & 0\\0 & 10\end{pmatrix}\right] I(|\beta_2| < 1), \qquad B_2 \sim N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}10 & 0\\0 & 10\end{pmatrix}\right] I(|\beta_4| < 1),$$

$$\sigma_1^2 \sim IG\Big(\frac{5}{2}, \frac{c \times (5-2)}{2}\Big), \qquad \sigma_2^2 \sim IG\Big(\frac{5}{2}, \frac{c \times (5-2)}{2}\Big),$$

$$r_0 \sim N(\mu_0, \lambda_0),$$

where μ_0 and λ_0 are the sample mean and variance of one replication, respectively. c is a hyperparameter in the inverse gamma distribution. The prior distribution of r_L and r_U is a bivariate uniform distribution where the joint p.d.f. is defined by

$$f_{r_L, r_U}(r_L, r_U) = \frac{1}{(Q_c - Q_a)(Q_d - Q_b)}$$

on the support $\{(r_L, r_U) : Q_a \leq r_L \leq Q_c, Q_b \leq r_U \leq Q_d, r_L \leq r_U\}$. Usually, Q_a, Q_b, Q_c and Q_d are the 20^{th}, 30^{th}, 70^{th} and 80^{th} percentile of one replication, respectively. This setting of the prior distribution of r_L and r_U makes sure that each regime contains at least twenty percent observations.

To study the sampling behaviour of the BARSV model, the diagnostic work based on an arbitrary replication is shown. Figure 1(a) demonstrates

the time series plots for the unknown parameters. As we can see from the time series plots, the drawn samples for each parameters fluctuate in the fixed interval and behave like uncorrelated sequences which provides indirect evidence that all the Markov chains are mixed well and attain the stationary (equilibrium) state. Figure 1(b) displays the autocorrelation function (ACF) plots for the unknown parameters. It is observed that the ACF values of $\theta_1, \theta_2, \theta_3$ and θ_4 are close to zero and the ACF values of the rest of parameters drop quickly to zero. ACF plots indicate that the sampling algorithm generates approximate random samples for the unknown parameters and makes sure that the fast convergence of the Markov chains is attained (See [6]). The same diagnostic techniques are also conducted in the empirical study.

Table 4 shows the summary statistics, in terms of Trial 1, including the initial values, the posterior means and variances and the 90% credible intervals. As we can see, all the posterior means are close to the corresponding true values and all the posterior variances are small. Although a minor departure can be observed from the estimate of σ_1^2, the true value of σ_1^2 fall into the corresponding 90% credible intervals. When the replication number is varied among 100, 300 and 500, all the estimates are close to the corresponding true values. It could be recognized that each replication bears a certain information of the BARSV model and 100 replications are sufficient to estimate the underlying model quite accurately. Figure 1(c) displays kernel density plots of all the unknown parameters, based on 500 posterior estimates. Most of the kernel density plots are represented by bell-sharped curve lines.

Figure 2 displays the time series plot of the true h_t and corresponding estimates $\hat{h}_t$ associated with one replication of Trial 1. It is noted that the red line, denoting the estimates $\hat{h}_t$, overlaps the black line, which represents the true value of h_t, meaning that our estimation method is able to reproduce the unobserved variable h_t. Figure 3 represents the time series of the observation r_t in contrast to the negative estimated volatility $-\exp(\hat{h}_t/2)$. As we can see, the negative estimated volatility could be considered as an approximate lower bound of the observation series.

In order to study the sensitivity and robustness of our proposed MCMC method further, another four trials are deployed. Table 5 presents the estimation results by employing four different sets of parameters. In each trial, it can be observed that the estimated parameters are close to the corresponding true values. It is evident that our proposed MCMC method is robust to the change of parameter values. Particularly, it is the only difference between

Trial 4 and 5 that the buffer zone is changed from $[-0.3, 0.1)$ to $[0, 0.1)$. It can be recognized that our estimated method performs well when the true buffer zone is varied.

4 Empirical study

In the empirical study, we consider the daily closing prices of NIKKEI stock average (NIKKEI225), the exchange rate for US Dollar to Japanese Yen (USD/JPY) and Hang Seng Index (HSI). These three time series are from 1^{st} January 2006 to 31^{st} December 2014. We focus on the series of log returns in percentage which are calculated by $r_t = 100 \times \log \frac{P_t}{P_{t-1}}$ and P_t represents the daily adjusted closing prices at time t. For easy illustration, time series $\{r_{\text{NIKKEI225}}\}$, $\{r_{\text{USD/JPY}}\}$ and $\{r_{\text{HSI}}\}$ are indicated by r_t on the following context. Figure 4 shows the time series plots of r_t wherein the volatility persistence phenomenon is observed. Figure 5 shows the ACF plots of $\{r_t\}$, $\{r_t^2\}$ and $\{\log(r_t^2)\}$. It is noted that a significant positive autocorrelation exists in the series $\{r_t^2\}$ and $\{\log(r_t^2)\}$. Thus, the specification of the series $\{r_t\}$ merely using the autoregressive model is inadequate. An autoregressive modelling for the series $\{r_t^2\}$ or $\{\log(r_t^2)\}$ should be taken into consideration. Table 6 presents the summary statistics for all of the empirical times series. Since time series of log returns have kurtosis greater than three, the distribution with respect to each series has thick tails. These series skewnesses are far from zero. It indicates the existence of the asymmetric feature in these time series. Therefore, it is reasonable to consider implementing the threshold volatility model into the series $\{r_t\}$ and $\{r_t^2\}$.

In the estimation work of the empirical study, 150000 samples are generated from our MCMC method for each parameter. The first 50000 samples are discarded, i.e. the first 50000 iterations are treated as the burn-in stage. To reduce the sample autocorrelation, the thinning parameter is also put to 10. The prior distributions, here, are initialized as the same as the counterparts in the simulation study.

We first consider the BARSV model defined by Equations (2). To be compared, two threshold autoregressive stochastic volatility (TARSV) models are studied in this paper. In the TARSV model, Equation (2e) is redefined

as the classical threshold which is denoted by

$$s_t = \begin{cases} 1, & \text{if } z_t \geq r_{thres}, \\ 0, & \text{if } z_t < r_{thres}. \end{cases}$$

Apart form that, we also study the TARSV model in So *et. al.* [6]. It is denoted by the TARSV-SO model in which r_{thres} is fixed at zero and σ_1^2 is the same as σ_2^2.

Table 7 reports the summary statistics of BARSV, TARSV and TARSV-SO models when the series $\{r_{\text{NIKKEI225}}\}$ is considered. We can observe that the posterior estimates for θ_2 and θ_4 are close to zero , which coincide with the conclusion drawn from the ACF plots for the series $\{r_{\text{NIKKEI225}}\}$ in Figure 5. In all the three models, it is noted that all the posterior estimates of β_2 and β_4 are close to one and the posterior estimates for σ_1^2 and σ_2^2 are small. It can be concluded that the series $\{r_{\text{NIKKEI225}}\}$ exhibits the phenomenon of volatility persistence. The buffer zone of the BARSV model is $[0.7712, 0.9790)$ and the threshold of the TARSV model is 0.8159. Due to the page limitation, we omit the summary statistics table and only report the estimates of threshold variables when the series $\{r_{\text{USD/JPY}}\}$ and $\{r_{\text{HSI}}\}$ are employed. For the series $\{r_{\text{USD/JPY}}\}$, the estimate of buffer zone in BARSV model is [-1.1282,1.1880) and the estimate of the classical threshold in TARSV model is 0.0935. For the series $\{r_{\text{HSI}}\}$, the estimate of buffer zone in BARSV model is [0.3594, 0.7252) and the estimate of the classical threshold in TARSV model is 0.5971. For these three empirical time series, the buffer zone in BARSV model includes the classical threshold in the TARSV model.Figure 7 shows an example of regime switching in the TARSV and BARSV model where the blue dash line denotes the TARSV model and the solid red line denote the BARSV model. The lower sub-figure shows the log-return, classical threshold and buffer zone. The upper sub-figure shows the regime indicators. First, we can observe the delay switching in the BARSV model. Moreover, we argue that the TARSV model is too sensitive to frequently switch the regimes when the observation slightly fluctuate around the threshold.

Table 8 reports the DIC measure including the mean deviance ($\bar{D}$), the effective dimension(pD) and the deviance information criterion (DIC) value. The BARSV model attains the smallest DIC value among three models. This leads us to the conclusion that the BARSV model performs better than the TARSV models. It indicates that the BARSV model is more reasonable to

fit the empirical series $\{r_t\}$, comparing with the TARSV and TARSV-SO model. To be more precise, the BARSV model is the best model after taking care of the model complexity. Figure 8 shows the time series of the observation $r_{\text{NIKKEI225}}$ in contrast to the 99^{th} percentile of the negative estimated volatility $-\exp(\hat{h}_t/2)$. It is visible that only a few violation can be detected. The BARSV model provides us a way to do the risk management.

Visual MCMC diagnostics are applied to check whether our proposed MCMC method gives estimates convergent to the target distribution. Figure 6(a) is the MCMC plot in which drawn samples in terms of each parameter do not exhibit upward or downward tendency and fluctuate in intervals. Figure 6(b) is the autocorrelation function (ACF) plots in which the ACF values for θ_1, θ_2, θ_3, θ_4 and β_2 are close to zero and the ACF values for other parameters drop to zero rapidly. Figure 6(b) gives evidence that the autocorrelations in the drawn samples fade out quickly. Figure 6(c) is the ergodic average plot in which the value of cumulative average for each parameter remains steady as a horizontal line approximately. According to the above analysis, our proposed MCMC method generates approximately random samples for each parameter and all the Markov chains have converged to the stationary distribution when the BARSV model is fitted to the series $\{r_{\text{NIKKEI225}}\}$. Figure 6(d) displays the empirical kernel density plots for each parameter which provides more information about the posterior estimates. As we can see, empirical kernel densities exhibit a bell-sharped curve fashion. An obviously asymmetric pattern is observed from the empirical kernel density plots of β_2 and β_4 since both estimates are drawn from a truncated normal distribution whose support is $(-1, 1)$. Similar conclusions of convergence diagnostics can be derived when both the TARSV and TARSV-SO model are fitted to the series $\{r_{\text{NIKKEI225}}\}$.

5 Conclusion

This paper proposes a new type of threshold stochastic volatility model in which hysteretic switching is allowed between different regimes. This (BARSV) model is a generalization of the threshold autoregressive stochastic volatility model proposed by So *et. al.* [6] and allows us to model the asymmetric volatility structure in a more flexible way. A full description of the MCMC estimation method is given and several simulation studies are conducted to verify the validation of our method. Three empirical time series are

implemented to illustrate the advantages of the BARSV model. Based on the results of DIC measure, the BARSV model outperforms the TARSV models in the empirical studies of the daily closing prices of NIKKEI stock average (NIKKEI225), the exchange rate for US Dollar to Japanese Yen (USD/JPY) and Hang Seng Index (HSI).

Appendix

A) Define notations:

r_{j,π_i} the ith observation belonging to the regime j in chronological order,

h_{j,π_i} the ith latent variable belonging to the regime j in chronological order,

π_i the associated time index in the time series $\{r_t\}$,

Z_j the design matrix in the measurement equation for regime j,

X_j the design matrix in the state equation for regime j,

j the regime indicator where $j = 1$ or $j = 2$,

T_j the size of observations belonging to regime j where $T_1 + T_2 = T$.

B) Denote the BARSV model in matrix form:
For Regime 1 ($j = 1$),

$$
\begin{aligned}
r^*_{1|\cdot} &= \begin{pmatrix} r_{1,\pi_1}\exp(-\frac{1}{2}h_{1,\pi_1}) \\ \vdots \\ r_{1,\pi_t}\exp(-\frac{1}{2}h_{1,\pi_t}) \\ \vdots \\ r_{1,\pi_{T_1}}\exp(-\frac{1}{2}h_{1,\pi_{T_1}}) \end{pmatrix} \\
&= \begin{pmatrix} \exp(-\frac{1}{2}h_{1,\pi_1}) & r_{1,\pi_1-1}\exp(-\frac{1}{2}h_{1,\pi_1}) \\ \vdots & \vdots \\ \exp(-\frac{1}{2}h_{1,\pi_t}) & r_{1,\pi_t-1}\exp(-\frac{1}{2}h_{1,\pi_t}) \\ \vdots & \vdots \\ \exp(-\frac{1}{2}h_{1,\pi_{T_1}}) & r_{1,\pi_{T_1}-1}\exp(-\frac{1}{2}h_{1,\pi_{T_1}}) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \\ \theta_2 \end{pmatrix} + \begin{pmatrix} y_{1,\pi_1}\exp(-\frac{1}{2}h_{1,\pi_1}) \\ \vdots \\ y_{1,\pi_t}\exp(-\frac{1}{2}h_{1,\pi_t}) \\ \vdots \\ y_{1,\pi_{T_1}}\exp(-\frac{1}{2}h_{1,\pi_{T_1}}) \end{pmatrix} \\
&= Z_1\Theta_1 + y^*_{1|\cdot},
\end{aligned}
$$

where

$$
y^*_{1|\cdot} \sim N(\mathbf{0}_{T_1}, \Sigma_{y_{1|\cdot}}) \text{ with } \Sigma_{y_{1|\cdot}} = \underbrace{\begin{pmatrix} \exp(-h_{1,\pi_1}) & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & & 0 \\ \vdots & & \exp(-h_{1,\pi_t}) & & \vdots \\ 0 & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \exp(-h_{1,\pi_{T_1}}) \end{pmatrix}}_{T_1 \times T_1 \text{ matrix}}.
$$

$$
h_{1|\cdot} = \begin{pmatrix} h_{1,\pi_1} \\ \vdots \\ h_{1,\pi_t} \\ \vdots \\ h_{1,\pi_{T_1}} \end{pmatrix} = \begin{pmatrix} 1 & h_{1,\pi_1-1} \\ \vdots & \vdots \\ 1 & h_{1,\pi_t-1} \\ \vdots & \vdots \\ 1 & h_{1,\pi_{T_1}-1} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \eta_{1,\pi_1} \\ \vdots \\ \eta_{1,\pi_t} \\ \vdots \\ \eta_{1,\pi_{T_1}} \end{pmatrix} = X_1 B_1 + \eta^{(1)},
$$

where

$$\eta^{(1)} \sim N(\mathbf{0}_{T_1}, \Sigma_1) \text{ with } \mathbf{0}_{T_1} = \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{T_1 \times 1 \text{ vector}} \quad \text{and } \Sigma_1 = \underbrace{\begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & & & 0 \\ \vdots & & \sigma_1^2 & & \vdots \\ 0 & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \sigma_1^2 \end{pmatrix}}_{T_1 \times T_1 \text{ matrix}} \succ 0.$$

Similarly, it is easily obtained the matrix form of the BARSV model for Regime 2.

C1) Posterior distributions of B_1 and B_2:

According to the Bayes theorem, the p.d.f. of the posterior distribution of B_1 is proportional to the product of the likelihood function and the p.d.f. of the prior distribution of B_1, which is denoted as

$$\begin{aligned} f_{pos}(B_1|\cdot) &\propto L(B_1|\cdot) \times f(B_1|\mu_0, \Lambda_0) \\ &\propto \exp\big\{ -\frac{1}{2}(h_{1|\cdot} - X_1 B_1)' \Sigma_1^{-1} (h_{1|\cdot} - X_1 B_1) \big\} \exp\big\{ -\frac{1}{2}(B_1 - \mu_0)' \Lambda_0^{-1} (B_1 - \mu_0) \big\}. \end{aligned}$$

Then,

$$\log(f_{pos}(B_1|\cdot)) \quad = constant - \tfrac{1}{2}(h_{1|\cdot} - X_1 B_1)' \Sigma_1^{-1} (h_{1|\cdot} - X_1 B_1) - \tfrac{1}{2}(B_1 - \mu_0)' \Lambda_0^{-1} (B_1 - \mu_0),$$

and the mean of B_1 is

$$\mu_{B_1} \equiv \mathbb{E}[B_1] = \arg\max_{B_1} \log(f_{pos}(B_1|\cdot)).$$

By the standard matrix operations, it can be derived that

$$\begin{aligned}
\frac{\partial \log(f_{pos}(B_1|\cdot))}{\partial B_1} &= -(h_{1|\cdot} - X_1 B_1)^{'} \Sigma_1^{-1}(-X_1) - (B_1 - \mu_0)^{'} \Lambda_0^{-1} \\
&= h_{1|\cdot}^{'} \Sigma_1^{-1} X_1 - B_1^{'} X_1^{'} \Sigma_1^{-1} X_1 - B_1^{'} \Lambda_0^{-1} + \mu_0^{'} \Lambda_0^{-1} \\
&= (h_{1|\cdot}^{'} \Sigma_1^{-1} X_1 + \mu_0^{'} \Lambda_0^{-1}) - B_1^{'}(X_1^{'} \Sigma_1^{-1} X_1 + \Lambda_0^{-1}) \\
&= (h_{1|\cdot}^{'} \Sigma_1^{-1} X_1 + \mu_0^{'} \Lambda_0^{-1}) - B_1^{'}(X_1^{'} \Sigma_1^{-1} X_1 + \Lambda_0^{-1}) \\
&= 0. \\
\Rightarrow \mu_{B_1} = \mathbb{E}[B_1] &= (X_1^{'} \Sigma_1^{-1} X_1 + \Lambda_0^{-1})(X_1^{'} \Sigma_1^{-1} h_{1|\cdot} + \Lambda_0^{-1} \mu_0^{'}) \\
&= (X_1^{'} \sigma_1^{-2} I X_1 + \Lambda_0^{-1})(X_1^{'} \sigma_1^{-2} I h_{1|\cdot} + \Lambda_0^{-1} \mu_0^{'}) \\
&= (\sigma_1^{-2} X_1^{'} X_1 + \Lambda_0^{-1})(\sigma_1^{-2} X_1^{'} h_{1|\cdot} + \Lambda_0^{-1} \mu_0^{'}).
\end{aligned}$$

Since the support of B_1 does not depend on itself, the Fisher information of B_1 is

$$\begin{aligned}
\mathcal{I}(B_1) = \sigma_{B_1}^{-2} &= \mathbb{E}\Big[\Big(\frac{\partial \log(f_{pos}(B_1|\cdot))}{\partial B_1}\Big)^2\Big] = \mathbb{E}\Big[-\frac{\partial^2 \log(f_{pos}(B_1|\cdot))}{\partial B_1^2}\Big] \\
&= (X_1^{'} \Sigma_1^{-1} X_1 + \Lambda_0^{-1})^{'} \\
&= X_1^{'} \Sigma_1^{-1} X_1 + \Lambda_0^{-1}. \\
\Rightarrow \sigma_{B_1}^2 &= \mathcal{I}(B_1)^{-1} = (X_1^{'} \Sigma_1^{-1} X_1 + \Lambda_0^{-1})^{-1} \\
&= (X_1^{'} \sigma_1^{-2} I X_1 + \Lambda_0^{-1})^{-1} \\
&= (\sigma_1^{-2} X_1^{'} X_1 + \Lambda_0^{-1})^{-1}.
\end{aligned}$$

Therefore, the posterior distribution of B_1 is given by

$$N_2\big((\sigma_1^{-2} X_1^{'} X_1 + \Lambda_0^{-1})(\sigma_1^{-2} X_1^{'} h_{1|\cdot} + \Lambda_0^{-1} \mu_0^{'}), (\sigma_1^{-2} X_1^{'} X_1 + \Lambda_0^{-1})^{-1}\big) I(|\beta_2| < 1). \tag{A.1}$$

Similarly, the posterior distribution of B_2 is

$$N_2\big((\sigma_2^{-2} X_2^{'} X_2 + \Lambda_0^{-1})(\sigma_2^{-2} X_2^{'} h_{2|\cdot} + \Lambda_0^{-1} \mu_0^{'}), (\sigma_2^{-2} X_2^{'} X_2 + \Lambda_0^{-1})^{-1}\big) I(|\beta_4| < 1). \tag{A.2}$$

C2) Posterior distributions of σ_1^2 and σ_2^2:

The posterior distribution of σ_1^2 is

$$\begin{aligned}
f_{pos}(\sigma_1^2|\cdot) &\propto L(\sigma_1^2|\cdot) \times f(\sigma_1^2|\upsilon/2, \upsilon\lambda/2) \\
&= \frac{1}{(2\pi)^{T_1/2}\det|\Sigma|^{1/2}} \exp\Big\{-\frac{1}{2}(h_{1|\cdot} - X_1B_1)'\Sigma^{-1}(h_{1|\cdot} - X_1B_1)\Big\} \\
&\qquad \times \frac{(\upsilon\lambda/2)^{\upsilon/2}}{\Gamma(\upsilon/2)}(\sigma_1^2)^{-\upsilon/2-1}\exp(-\frac{\upsilon\lambda}{2\sigma^2}) \\
&\propto (\sigma_1^{-2})^{T_1/2}\exp\Big\{-\frac{1}{2}\sigma_1^{-2}(h_{1|\cdot} - X_1B_1)'(h_{1|\cdot} - X_1B_1)\Big\} \times (\sigma_1^2)^{-\upsilon/2-1}\exp(-\frac{\upsilon\lambda}{2\sigma^2}) \\
&= (\sigma_1^2)^{-(T_1+\upsilon)/2-1}\exp\Big\{-\frac{1}{\sigma_1^2}[(h_{1|\cdot} - X_1B_1)'(h_{1|\cdot} - X_1B_1)/2 + \upsilon\lambda/2]\Big\}.
\end{aligned}$$

Since the posterior p.d.f. of σ_1^2 is proportional to the kernel function of the inverse gamma distribution, it is straightforward to obtain that

$$\sigma_1^2 \sim IG\bigg(\frac{(T_1+\upsilon)}{2}, \frac{T_1 s_1^2 + \upsilon\lambda}{2}\bigg), \tag{A.3}$$

where

$s_1^2 = \dfrac{1}{T_1}(h_{1|\cdot} - X_1B_1)'(h_{1|\cdot} - X_1B_1)$ is the mean of squared error of h_t in Regime 1.

Similarly, the posterior distribution of σ_2^2 is

$$\sigma_2^2 \sim IG\bigg(\frac{(T_2+\upsilon)}{2}, \frac{T_2 s_2^2 + \upsilon\lambda}{2}\bigg), \tag{A.4}$$

where

$s_2^2 = \dfrac{1}{T_2}(h_{2|\cdot} - X_2B_2)'(h_{2|\cdot} - X_2B_2)$ is the mean of squared error of h_t in Regime 2.

C3) Posterior distributions of Θ_1 and Θ_2:

The posterior distribution of Θ_1 is

$$\begin{aligned}
f_{pos}(\Theta_1|\cdot) &\propto L(\Theta_1|\cdot) \times f(\Theta_1|\mu_0, \Lambda_0) \\
&\propto \exp\Big\{-\frac{1}{2}(r^*_{1|\cdot} - Z_1\Theta_1)'I^{-1}(r^*_{1|\cdot} - Z_1\Theta_1)\Big\}\exp\Big\{-\frac{1}{2}(\Theta_1 - \mu_1)'\Lambda_1^{-1}(\Theta_1 - \mu_1)\Big\}.
\end{aligned}$$

Then,

$$\log(f_{pos}(\Theta_1|\cdot)) = constant - \frac{1}{2}(r^*_{1|\cdot} - Z_1\Theta_1)'(r^*_{1|\cdot} - Z_1\Theta_1) - \frac{1}{2}(\Theta_1 - \mu_1)'\Lambda_1^{-1}(\Theta_1 - \mu_1),$$

and the mean of Θ_1 is

$$\begin{aligned}
\mu_{\Theta_1} &\equiv \mathbb{E}[\Theta_1] = \arg\max_{\Theta_1} \log(f_{pos}(\Theta_1|\cdot)). \\
\frac{\partial \log(f_{pos}(\Theta_1|\cdot))}{\partial \Theta_1} &= -(r^*_{1|\cdot} - Z_1\Theta_1)'(-Z_1) - (\Theta_1 - \mu_1)'\Lambda_1^{-1} \\
&= (r^*_{1|\cdot})'Z_1 - \Theta_1'Z_1'Z_1 - \Theta_1'\Lambda^{-1} + \mu_1'\Lambda_1^{-1} \\
&= ((r^*_{1|\cdot})'Z_1 + \mu_1'\Lambda_1^{-1}) - \Theta_1'(Z_1'Z_1 + \Lambda_1^{-1}) \\
&= 0 \\
\Rightarrow \mu_{\Theta_1} = \mathbb{E}[\Theta_1] &= (Z_1'Z_1 + \Lambda_1^{-1})^{-1}(Z_1'r^*_{1|\cdot} + \Lambda_1^{-1}\mu_1')
\end{aligned}$$

Since the support of Θ_1 does not depend on itself, the Fisher information of Θ_1 is

$$\begin{aligned}
\mathcal{I}(\Theta_1) = \sigma_{\Theta_1}^{-2} &= \mathbb{E}\Big[\Big(\frac{\partial \log(f_{pos}(\Theta_1|\cdot))}{\partial \Theta_1}\Big)^2\Big] = \mathbb{E}\Big[-\frac{\partial^2 \log(f_{pos}(\Theta_1|\cdot))}{\partial \Theta_1^2}\Big] \\
&= (Z_1'Z_1 + \Lambda_1^{-1})' \\
&= Z_1'Z_1 + \Lambda_1^{-1} \\
\Rightarrow \sigma^2_{\Theta_1} &= \mathcal{I}(\Theta_1)^{-1} = (Z_1'Z_1 + \Lambda_1^{-1})^{-1}
\end{aligned}$$

Therefore, the posterior distribution of Θ_1 is given by

$$N_2\big((Z_1'Z_1 + \Lambda_1^{-1})^{-1}(Z_1'r^*_{1|\cdot} + \Lambda_1^{-1}\mu_1'), (Z_1'Z_1 + \Lambda_1^{-1})^{-1}\big) I(|\theta_2| < 1). \quad \text{(A.5)}$$

Similarly, the posterior distribution of Θ_2 is

$$N_2\big((Z_2'Z_2 + \Lambda_1^{-1})^{-1}(Z_2'r^*_{2|\cdot} + \Lambda_1^{-1}\mu_1'), (Z_2'Z_2 + \Lambda_1^{-1})^{-1}\big) I(|\theta_4| < 1). \quad \text{(A.6)}$$

C4) Posterior distribution of r_0:

Since $r_1 = \psi_{1,1} + \psi_{2,1}r_0 + y_1$, $r_0 = (r_1 - \psi_{1,1} - y_1)/\psi_{2,1}$. The distribution of r_0 derived from the measurement equation is denoted as

$$N\left(\frac{r_1 - \psi_{1,1}}{\psi_{2,1}}, \frac{\exp(h_1)}{\psi_{2,1}^2}\right).$$

Combined with the prior distribution of r_0, the posterior mean of r_0 is

$$\begin{aligned}\mu_{r_0} &= \frac{\frac{r_1-\psi_{1,1}}{\psi_{2,1}}\lambda_0 + \mu_0\frac{\exp(h_1)}{\psi_{2,1}^2}}{\frac{\exp(h_1)}{\psi_{2,1}^2}+\lambda_0} \\ &= \frac{\lambda_0\exp(h_1^{-1})\psi_{2,1}(r_1-\psi_{1,1})+\mu_0}{1+\lambda_0\exp(h_1^{-1})\psi_{2,1}^2}.\end{aligned}$$

The posterior variance of r_0 is

$$\sigma_{r_0}^2 = \frac{\lambda_0\exp(h_1)\psi_{2,1}^{-2}}{\lambda_0+\exp(h_1)\psi_{2,1}^{-2}} = \frac{\lambda_0}{1+\lambda_0\exp(h_1^{-1})\psi_{2,1}^2}.$$

Therefore, the posterior distribution of r_0 is given by

$$N\left(\frac{\lambda_0\exp(h_1^{-1})\psi_{2,1}(r_1-\psi_{1,1})+\mu_0}{1+\lambda_0\exp(h_1^{-1})\psi_{2,1}^2}, \frac{\lambda_0}{1+\lambda_0\exp(h_1^{-1})\psi_{2,1}^2}\right). \tag{A.7}$$

6 Acknowledgements

This work is supported by RGC Grant PolyU 15223419 and PolyU grant UAHF.

References

[1] Shephard N. Statistical aspects of ARCH and stochastic volatility. *Monographs on Statistics and Applied Probability* 1996; 65: 1–68.

[2] Tsay RS. *Analysis of financial time series*. 543. John Wiley & Sons . 2005.

[3] Taylor S. *Modelling financial time series*. Chichester: Wiley . 1986.

[4] Tong H. On a threshold model. In: Chen CH. , ed. *Pattern Recognition and Signal Processing*Amsterdam: Sijthoff & Noordhoff; 1978: 575–586.

[5] Nelson D. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 1991; 59(2): 347–370.

[6] So MK, Li W, Lam K. A threshold stochastic volatility model. *Journal of Forecasting* 2002; 21(7): 473–500.

[7] So MK, Choi C. A threshold factor multivariate stochastic volatility model. *Journal of Forecasting* 2009; 28(8): 712–735.

[8] Chen CW, So MK, Liu FC. A review of threshold time series models in finance. *Statistics and its Interface* 2011; 4(2): 167–181.

[9] Li G, Guan B, Li WK, Yu PL. Hysteretic autoregressive time series models. *Biometrika* 2015; 102(3): 717–723.

[10] Zhu K, Yu PL, Li WK. Testing for the buffered autoregressive processes. *Statistica Sinica* 2014: 971–984.

[11] Lo PH, Li WK, Yu PL, Li G. On buffered threshold GARCH models. *Statistica Sinica* 2016; 26(4): 1555–1567.

[12] Zhu K, Li WK, Yu PL. Buffered autoregressive models with conditional heteroscedasticity: An application to exchange rates. *Journal of Business & Economic Statistics* 2017; 35(4): 528–542.

[13] Chen CW, Truong BC. On double hysteretic heteroskedastic model. *Journal of Statistical Computation and Simulation* 2016; 86(13): 2684–2705.

[14] Truong BC, Chen CW, Sriboonchitta S. Hysteretic Poisson INGARCH model for integer-valued time series. *Statistical Modelling* 2017; 17(6): 401–422.

[15] Chen CW, Lee S, Khamthong K. Bayesian inference of nonlinear hysteretic integer-valued GARCH models for disease counts. *Computational Statistics* 2021; 36(1): 261–281.

[16] Chen CW, Than-Thi H, So MK, Sriboonchitta S. Quantile forecasting based on a bivariate hysteretic autoregressive model with GARCH errors and time-varying correlations. *Applied Stochastic Models in Business and Industry* 2019; 35(6): 1301–1321.

[17] Chen CW, Than-Thi H, Asai M. On a bivariate hysteretic AR-GARCH model with conditional asymmetry in correlations. *Computational Economics* 2021; 58(2): 413–433.

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

[19] Harvey A, Ruiz E, Shephard N. Multivariate stochastic variance models. *The Review of Economic Studies* 1994; 61(2): 247–264.

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

[21] So Kp. On the statistical modelling of stochastic volatility and its applications to financial markets. *HKU Theses Online (HKUTO)* 1996.

[22] Chen CW, Lee JC. Bayesian inference of threshold autoregressive models. *Journal of time series analysis* 1995; 16(5): 483–492.

[23] Broto C, Ruiz E. Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys* 2004; 18(5): 613–649.

[24] Roberts GO, Gelman A, Gilks WR. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability* 1997; 7(1): 110–120.

[25] Roberts GO, Rosenthal JS. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical science* 2001; 16(4): 351–367.

[26] Jacquier E, Polson NG, Rossi PE. Bayesian analysis of stochastic volatility models (with Discussion). *Journal of Business & Economic Statistics* 1994; 12(4): 69-86.

[27] Shephard N, Pitt MK. Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 1997; 84(3): 653–667.

[28] Tsang Wy. Aspects of modelling stochastic volatility. *HKU Theses Online (HKUTO)* 2000.

[29] Watanabe T, Omori Y. A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997). *Biometrika* 2004; 91(1): 246–248.

[30] Kalman RE. A new approach to linear filtering and prediction problems. *Journal of basic Engineering* 1960; 82(1): 35–45.

Variable	Definition
r_t	time series of the observations
y_t	time series of the innovations
h_t	time series of the log-volatilities (or hidden states)
s_t	time series of regime indicators
u_t, η_t	white-noise processes
$\theta_1, \theta_2, \theta_3, \theta_4$	parameters in observation equation
$\beta_1, \beta_2, \beta_3, \beta_4$	parameters in state equation
σ_1, σ_2	variances of the white-noise processes
z_t	threshold variable
r_L	the lower bound of the buffer zone
r_U	the upper bound of the buffer zone

Table 1: The list of variables in the BARSV model

Step 0	Initialize the Gibbs sampler by specifying $\Theta_1^{(0)}, \Theta_2^{(0)}, B_1^{(0)}, B_2^{(0)}, \sigma_1^{2^{(0)}}, \sigma_2^{2^{(0)}}$, $r_0^{(0)}$ and $r_{thres}^{(0)}$.
	At iteration n:
Step 1	Draw $\Theta_1^{(n)}$ and $\Theta_2^{(n)}$.
Step 2	Draw $B_1^{(n)}$ and $B_2^{(n)}$.
Step 3	Draw $\sigma_1^{2^{(n)}}$ and $\sigma_2^{2^{(n)}}$.
Step 4	Draw $r_0^{(n)}$.
Step 5	Draw $\rho_t^{(n)}, t = 1, \ldots, T$.
Step 6	Draw $r_L^{(n)}$ and $r_U^{(n)}$.
Step 7	Draw $h_t^{(n)}, t = T, \ldots, 1$.
Step 8	Repeat step 1 to step 7 and keep samples after convergence is achieved.

Table 2: MCMC algorithm for the BARSV model

	θ_1	θ_2	θ_3	θ_4	β_1	β_2	β_3	β_4	σ_1^2	σ_2^2	r_L	r_U
Trial 1	0.05	-0.05	-0.05	0.05	-0.35	0.7	0.35	0.9	0.3	0.1	-0.1	0.1
Trial 2	0.15	-0.15	0	0.05	-0.15	0.75	0.3	0.95	0.5	0.1	-0.15	0.15
Trial 3	0.2	0.2	-0.1	0.1	-0.3	0.7	0.3	0.9	0.2	0.05	-0.1	0.2
Trial 4	0	0.25	-0.2	0.35	-0.25	0.75	0.25	0.95	0.2	0.05	-0.3	0.1
Trial 5	0	0.25	-0.2	0.35	-0.25	0.75	0.25	0.95	0.2	0.05	0	0.1

Table 3: BARSV (Simulation): parameter values of each trial.

Parameter	True Value	Initial Value	Replication Number $N = 100$	$N = 300$	$N = 500$
θ_1	0.05	0	0.0494 [0.0005417] (0.0106621, 0.0896005)	0.0478 [0.0005690] (0.0098041, 0.0875631)	0.0474 [0.000 (0.0077009, 0.08
θ_2	-0.05	0	-0.0496 [0.0006226] (-0.0976819,-0.0070380)	-0.0491 [0.0006286] (-0.0880208,-0.0047143)	-0.0485 [0.000 (-0.0914900,-0.0
θ_3	-0.05	0	-0.0480 [0.0011101] (-0.0994707, 0.0108876)	-0.0465 [0.0013067] (-0.1047969, 0.0151787)	-0.0486 [0.001 (-0.1083979, 0.0
θ_4	0.05	0	0.0538 [0.0012714] (-0.0040833, 0.1076159)	0.0518 [0.0016289] (-0.0194214, 0.1204969)	0.0509 [0.001 (-0.0194214, 0.12
β_1	-0.35	0	-0.3596 [0.0006487] (-0.4105130,-0.3248975)	-0.3558 [0.0007373] (-0.4067506,-0.3150478)	-0.3557 [0.000 (-0.4002444,-0.31
β_2	0.7	0.5	0.6911 [0.0007030] (0.6435550, 0.7329862)	0.6936 [0.0009110] (0.6414956, 0.7404433)	0.6940 [0.000 (0.6433786, 0.74
β_3	0.35	0	0.3542 [0.0004055] (0.3243129, 0.3893340)	0.3553 [0.0005008] (0.3199368, 0.3938148)	0.3555 [0.000 (0.3204410, 0.39
β_4	0.9	0.5	0.8978 [0.0005342] (0.8593499, 0.9375143)	0.8961 [0.0005534] (0.8570659, 0.9343741)	0.8959 [0.000 (0.8564474, 0.93
σ_1^2	0.3	0.5	0.3126 [0.0021296] (0.2470537, 0.3834312)	0.3110 [0.0021619] (0.2305248, 0.3828782)	0.3101 [0.002 (0.2331755, 0.38
σ_2^2	0.1	0.5	0.1050 [0.0006892] (0.0658894, 0.1547968)	0.1023 [0.0007595] (0.0628706, 0.1547058)	0.1030 [0.000 (0.0621859, 0.15
r_L	-0.1	0.5	-0.0962 [0.0003811] (-0.1210256,-0.0652547)	-0.0978 [0.0004803] (-0.1286806,-0.0610273)	-0.0981 [0.000 (-0.1296079,-0.06
r_U	0.1	-0.5	0.1013 [0.0002103] (0.0780835, 0.1222115)	0.0992 [0.0002749] (0.0762078, 0.1224401)	0.0985 [0.000 (0.0706633, 0.1

Table 4: BARSV (Simulation): posterior means , posterior variance (in square brackets) and 90% credible interval (in parentheses) associated with Trial 1.

Parameter	θ_1	θ_2	θ_3	θ_4	β_1	β_2	β_3	β_4	σ_1^2	σ
Trial 2										
True	0.15	-0.15	0	0.05	-0.15	0.75	0.3	0.95	0.5	0
Est.	0.1424	-0.1456	0.0092	0.0556	-0.1537	0.7465	0.3068	0.9453	0.5059	0.1
Variance	0.0008	0.0006	0.0019	0.0016	0.0008	0.0006	0.0007	0.0003	0.0039	0.0
Trial 3										
True	0.2	0.2	-0.1	0.1	-0.3	0.7	0.3	0.9	0.2	0.
Est.	0.1961	0.2020	-0.0890	0.1090	-0.3088	0.6916	0.3043	0.8903	0.2081	0.0
Variance	0.0006	0.0006	0.0016	0.0021	0.0005	0.0008	0.0003	0.0006	0.0011	0.0
Trial 4										
True	0	0.25	-0.2	0.35	-0.25	0.75	0.25	0.95	0.2	0.
Est.	-0.0041	0.2521	-0.1934	0.3541	-0.2573	0.7430	0.2553	0.9441	0.2087	0.0
Variance	0.0006	0.0007	0.0017	0.0013	0.0004	0.0005	0.0002	0.0003	0.0010	0.0
Trial 5										
True	0	0.25	-0.2	0.35	-0.25	0.75	0.25	0.95	0.2	0.
Est.	-0.0078	0.2545	-0.1971	0.3520	-0.2554	0.7436	0.2564	0.9435	0.2085	0.0
Variance	0.0011	0.0009	0.0013	0.0010	0.0005	0.0007	0.0003	0.0003	0.0011	0.0

Table 5: BARSV (Simulation): Estimation results in terms of Trial 2, 3, 4 and 5.

	# of Obs.	Mean	Variance	Skewness	Kurtosis
$r_{\text{NIKKEI225}}$	2208	0.0036	2.6855	-0.5347	10.6958
$r^2_{\text{NIKKEI225}}$	2208	2.6843	69.8590	11.4401	178.3183
$\log(r^2_{\text{NIKKEI225}})$	2208	-0.7796	5.6487	-0.9014	3.8855
r_{USDJPY}	2348	0.00073	0.4494	0.1209	7.6500
r^2_{USDJPY}	2348	0.4492	1.3427	10.8614	213.8575
$\log(r^2_{\text{USD/JPY}})$	2467	-2.5187	5.3522	-0.7137	3.2246
r_{HSI}	2467	0.0202	2.7311	0.0456	11.8906
r^2_{HSI}	2467	2.7303	81.2175	11.7472	191.5400
$\log(r^2_{\text{HSI}})$	2467	-1.0695	7.1169	-0.9631	4.0118

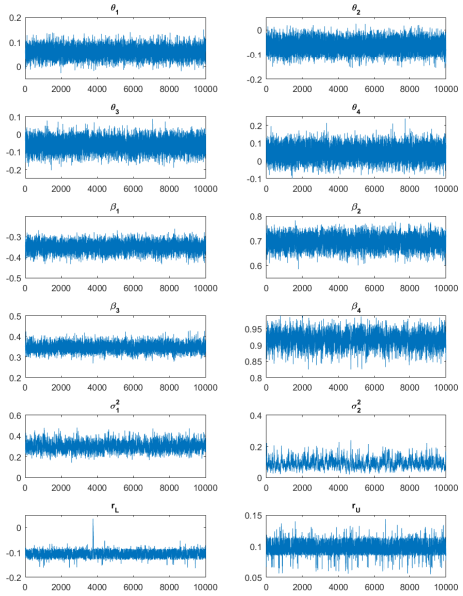
Table 6: Summary statistics for the log daily return (%) of NIKKEI225, USD/JPY and HSI

Parameter	BARSV	TARSV	TARSV-SO
θ_1	-0.0357 [0.0209544] (-0.2764898, 0.2007253)	-0.0297 [0.0171934] (-0.2459309, 0.1832046)	-0.0381 [0.0026210] (-0.1234659, 0.0466065)
θ_2	0.0317 [0.0067379] (-0.1035775, 0.1664631)	0.0306 [0.0059503] (-0.0961136, 0.1580550)	0.0227 [0.0020591] (-0.0515929, 0.0966541)
θ_3	0.0169 [0.0008984] (-0.0322285, 0.0663064)	0.0131 [0.0009921] (-0.0393157, 0.0652673)	0.0335 [0.0031238] (-0.0575817, 0.1252002)
θ_4	-0.0817 [0.0010522] (-0.1352179,-0.0284628)	-0.0839 [0.0011796] (-0.1396947,-0.0264909)	-0.0641 [0.0022819] (-0.1429155, 0.0150084)
β_1	-0.2323 [0.0015606] (-0.2978694,-0.1675909)	-0.2237 [0.0015561] (-0.2885019,-0.1590855)	-0.0787 [0.0005360] (-0.1171084,-0.0406688)
β_2	0.9782 [0.0002972] (0.9447314, 0.9983791)	0.9821 [0.0002286] (0.9520063, 0.9988807)	0.9538 [0.0004326] (0.9193658, 0.9883975)
β_3	0.0819 [0.0001340] (0.0635793, 0.1015549)	0.0922 [0.0001956] (0.0703340, 0.1156657)	0.1318 [0.0008226] (0.0858596, 0.1798716)
β_4	0.9904 [0.0000551] (0.9760163, 0.9991605)	0.9844 [0.0001009] (0.9662259, 0.9980935)	0.9622 [0.0005102] (0.9209933, 0.9945673)
σ_1^2	0.0977 [0.0006281] (0.0621638, 0.1436012)	0.0786 [0.0005766] (0.0459669, 0.1230970)	0.0442 [0.0000663] (0.0321936, 0.0586047)
σ_2^2	0.0292 [0.0000600] (0.0184105, 0.0430850)	0.0376 [0.0000822] (0.0250591, 0.0541950)	0.0442 [0.0000663] (0.0321936, 0.0586047)
r_L	0.7712 [0.0016408] (0.7130471, 0.8451786)	–	–
r_U	0.9790 [0.0022706] (0.8950070, 1.0511998)	–	–
r_{thres}	–	0.8159 [0.0019609] (0.7417181, 0.8800605)	0 [ ] –

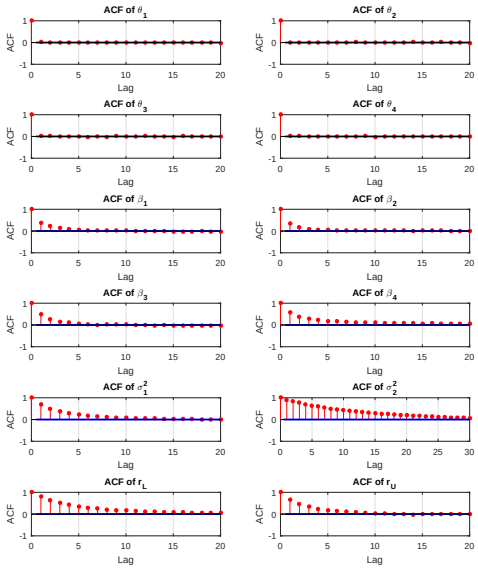
Table 7: Summary statistics (NIKKEI225): posterior means , posterior variance (in square brackets) and 90% credible interval (in parentheses)

Model	$\bar{D}$	pD	DIC
NIKKEI			
BARSV	7388.729	158.0590	7546.788
TARSV	7392.399	162.9333	7555.332
TARSV-SO	7464.614	153.2396	7617.853
USD/JPY			
BARSV	3868.661	228.2101	4096.871
TARSV	3901.433	208.8672	4110.301
TARSV-SO	3926.655	193.7104	4120.365
HSI			
BARSV	7967.365	162.8512	8130.216
TARSV	7981.356	166.7117	8148.068
TARSV-SO	8040.516	150.4509	8190.967

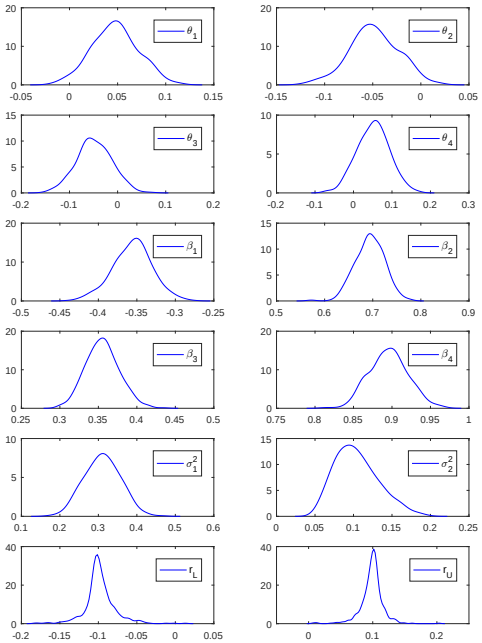
Table 8: Model Comparison (NIKKEI225, USD/JPY and HSI)



(a) MCMC plots of parameters



(b) ACF plots of parameters



(c) Kernel density plots of parameters

Figure 1: BARSV (Simulation): One replication of Trial 1

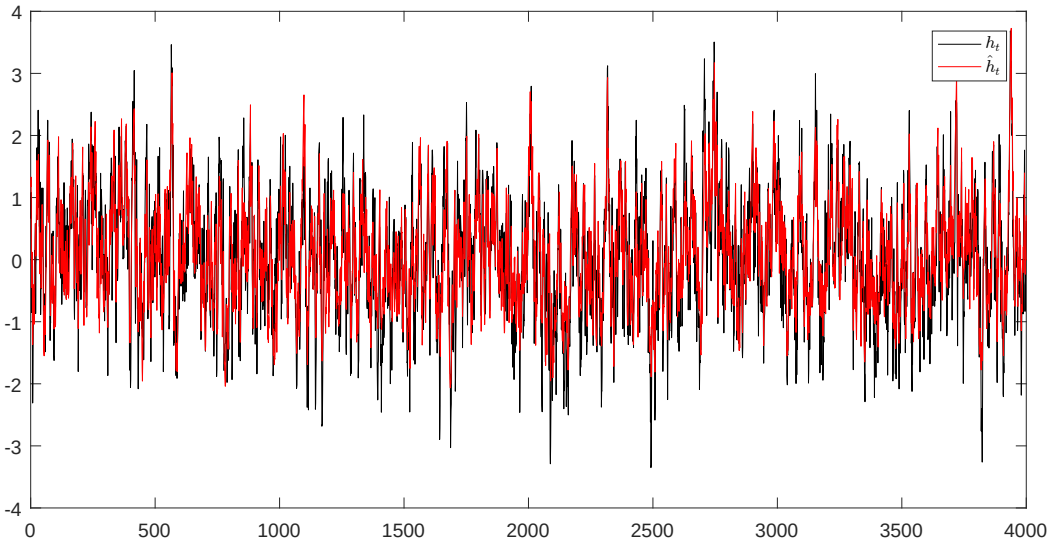


Figure 2: BARSV (Simulation): time series plot of true h_t and its estimated value $\hat{h}_t$ in one replication of Trial 1

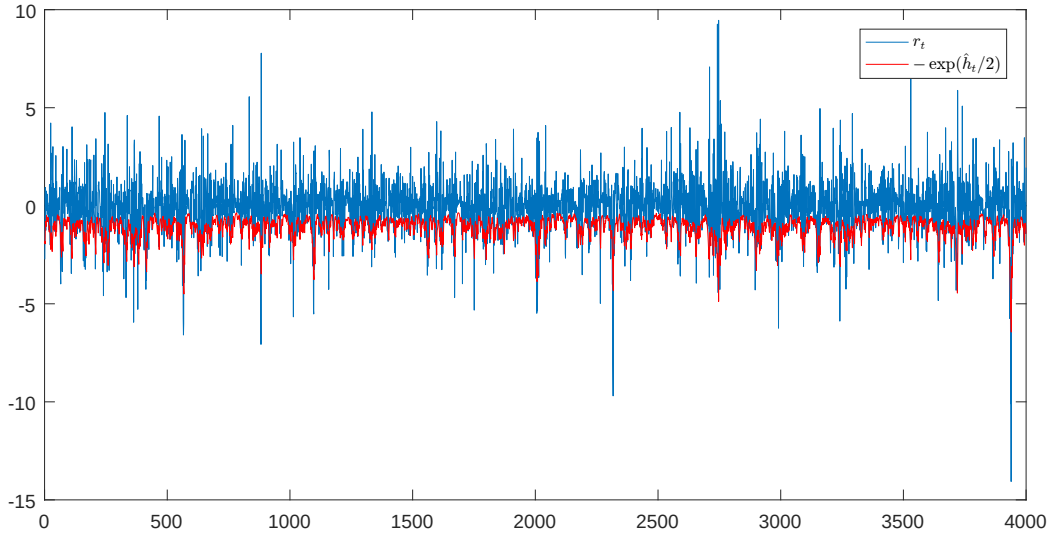


Figure 3: BARSV (Simulation): time series plot of the observation r_t and the negative estimated volatility $-\exp(\hat{h}_t/2)$ in one replication of Trial 1

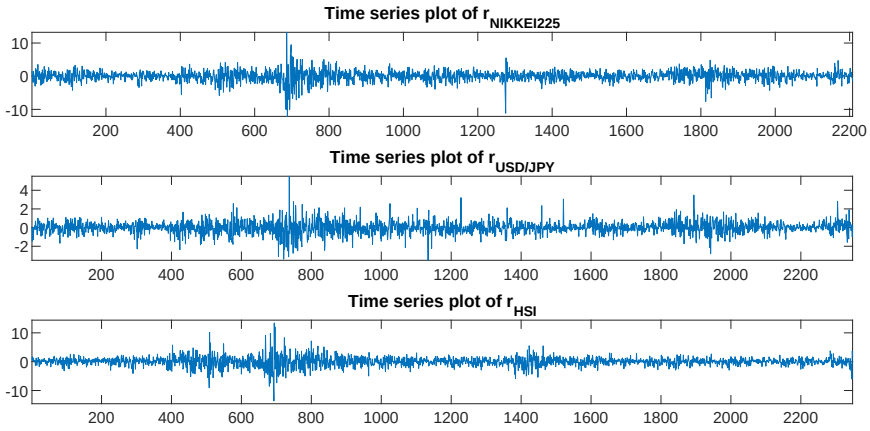


Figure 4: Time series plots for the log-daily-return (%) of NIKKEI225, USD/JPY and HSI

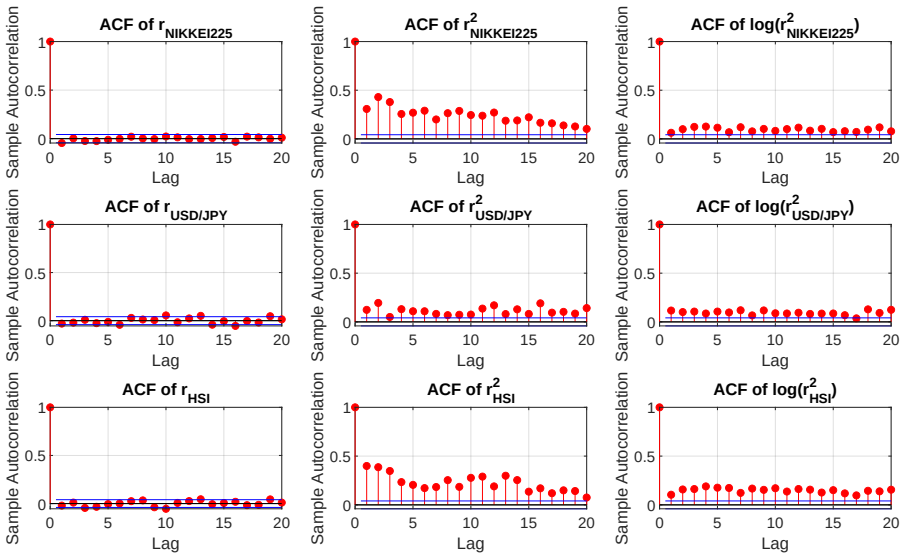
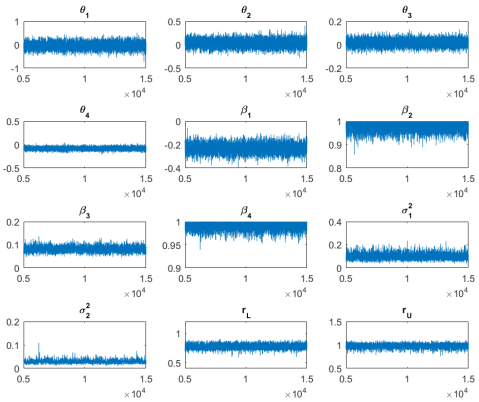
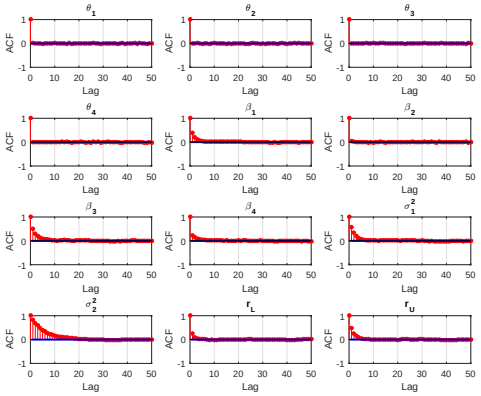


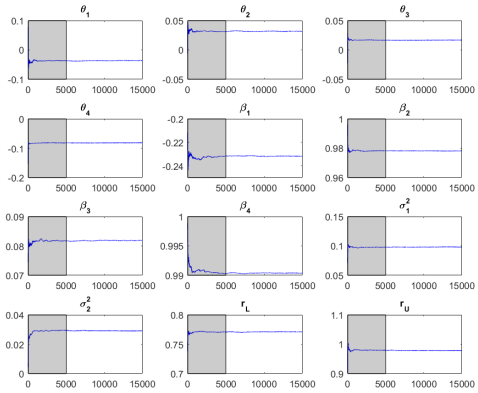
Figure 5: ACF plots for time series $\{r_{\text{NIKKEI225}}\}$, $\{r^2_{\text{NIKKEI225}}\}$, $\{\log r^2_{\text{NIKKEI225}}\}$, $\{r_{\text{USD/JPY}}\}$, $\{r^2_{\text{USD/JPY}}\}$, $\{\log r^2_{\text{USD/JPY}}\}$, $\{r_{\text{HSI}}\}$, $\{r^2_{\text{HSI}}\}$, and $\{\log r^2_{\text{HSI}}\}$



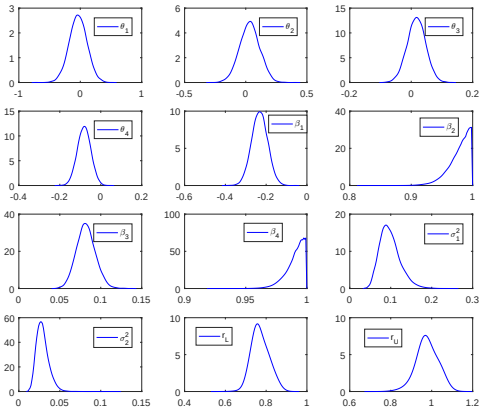
(a) MCMC plot



(b) ACF plot



(c) Ergodic average plots



(d) Kernel density plots

Figure 6: BARSV (NIKKEI225): Convergence diagnostic

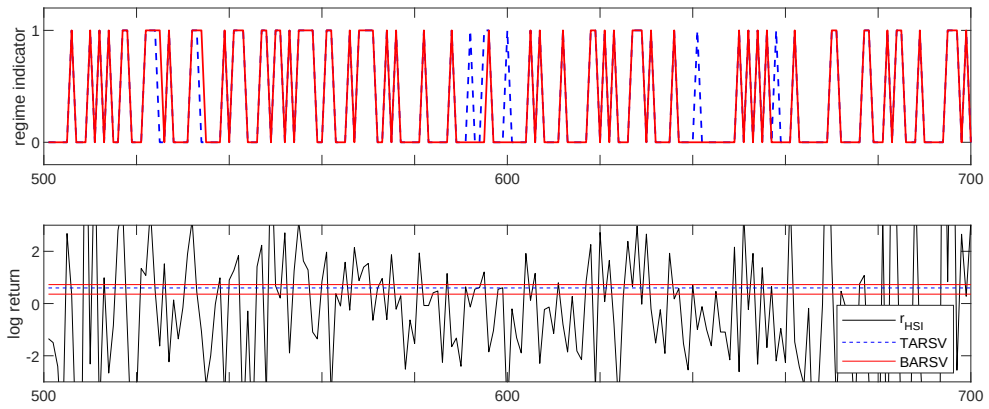


Figure 7: Regime switching of TARSV and BARSV models among the 501^{st} - 700^{th} observations in series r_{HSI}

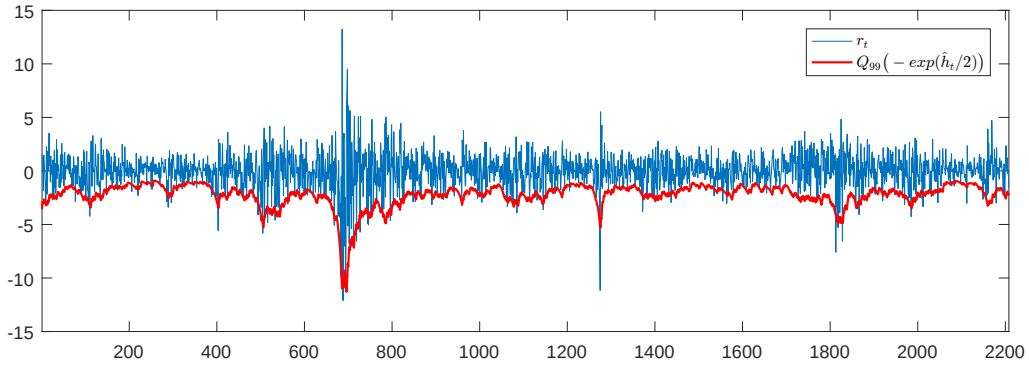


Figure 8: BARSV (NIKKEI225): the 99^{th} percentile of the negative estimated volatility $-\exp(\hat{h}_t/2)$