REALIZED NON-LINEAR STOCHASTIC VOLATILITY MODELS WITH ASYMMETRIC EFFECTS AND GENERALIZED STUDENT’S T-DISTRIBUTIONS

Didit Budi Nugroho* and Takayuki Morimoto**

This study proposes a class of realized non-linear stochastic volatility models with asymmetric effects and generalized Student’s t-error distributions by applying three families of power transformation—exponential, modulus, and Yeo-Johnson—to lagged log volatility. The proposed class encompasses a raw version of the realized stochastic volatility model. In the Markov chain Monte Carlo algorithm, an efficient Hamiltonian Monte Carlo (HMC) method is developed to update the latent log volatility and transformation parameter, whereas the other parameters that could not be sampled directly are updated by an efficient Riemann manifold HMC method. Empirical studies on daily returns and four realized volatility estimators of the Tokyo Stock Price Index (TOPIX) over 4-year and 8-year periods demonstrate statistical evidence supporting the incorporation of skew distribution into the error density in the returns and the use of power transformations of lagged log volatility.

Keywords and phrases: Hamiltonian Monte Carlo, non-linear stochastic volatility, realized variance, skew distribution, TOPIX.

1. Introduction

Volatility modelling of asset returns is one of the most prolific topics in the financial econometrics time series literature, including the autoregressive conditional heteroskedasticity (ARCH) model proposed by Engle (1982), the generalized ARCH model (GARCH) by Bollerslev (1986), the stochastic volatility (SV) model proposed by Taylor (1982) and Kim et al. (1998), and the realized variance (RV) approach introduced by Andersen et al. (2001). Theoretically, the SV model is much more flexible, realistic, and better performing than the ARCH-type models (Ghysels et al. (1996)) and generalized ARCH (GARCH)-type models (Kim et al. (1998), Carnero et al. (2001), Yu (2002)) since an innovation term is embedded in the volatility dynamics.

Recently, models joining returns and realized measures have been developed via a measurement equation that relates the conditional variance of the returns to the realized measure, which also includes ‘asymmetry’ to shocks for making a very flexible and rich representation. Koopman and Scharth (2013) provide a short overview of the joint models outside the SV methodology. In particular, Hansen et al. (2011) introduced the Realized GARCH model that substantially...

Received February 12, 2014. Revised May 21, 2014. Accepted June 25, 2014.

*Department of Mathematics, Satya Wacana Christian University, Jl. Diponegoro 52-60, Salatiga, Central Java 50711, Indonesia. Email: didit.budinugroho@staff.uksw.edu

**Department of Mathematical Sciences, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan. Email: morimot@kwansei.ac.jp
improves the empirical fit compared to the standard GARCH models that only use return series. In the context of the SV model, very closely related studies of joint models have been proposed by Takahashi et al. (2009), Dobrev and Szerszen (2010), and Koopman and Scharth (2013). Their model is known as the realized stochastic volatility (RSV) model.

This study focuses on Takahashi et al.’s (2009) version. Recently, Takahashi et al. (2014) extended their model by applying a general non-linear bias correction in the RV measure and the generalized hyperbolic skew Student’s t-distribution (SKT distribution), which includes normal and Student’s t-distributions as special cases. The models studied by Takahashi et al. (2009) and Takahashi et al. (2014) have already incorporated the asymmetric volatility phenomenon, which captures the negative correlation between current return and future volatility. We label Takahashi et al.’s (2014) asymmetric version of the so called RSVC model as the R-ASV model and then formulate this model as

$$
\begin{align*}
R_t &= e^{(1/2)h_t} \epsilon_t, \\
Y_t &= \xi_0 + \xi_1 h_t + \sigma_y u_t, \\
h_{t+1} &= \alpha + \phi (h_t - \alpha) + \sigma_h v_{t+1}, \\
h_1 &\sim \mathcal{N}(\alpha, \sigma_h^2/(1 - \phi^2)), \\
\begin{bmatrix}
\epsilon_t \\
u_t \\
v_{t+1}
\end{bmatrix} &\sim \mathcal{N}
\begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}, \\
\begin{bmatrix}
1 & 0 & \rho \\
0 & 1 & 0 \\
0 & \rho & 0
\end{bmatrix}
\end{align*}
$$

for \( t = 1, 2, \ldots, T \), where \( R_t \) are the log returns over a unit time period from which the autocorrelations are removed, \( \exp(h_t) \) characterizes the unobservable variance of returns, \( Y_t \) denotes the log RV, \( \xi_0 \) is a linear bias correction parameter representing the effect of microstructure noise (if \( \xi_0 > 0 \)) or that of non-trading hours (if \( \xi_0 < 0 \)), \( \xi_1 \) is a non-linear bias correction parameter representing the persistence (correlation) of RV, \( \alpha \) and \( \phi \) represent the drift and persistence of log volatilities \( h_t \), respectively, and \( \rho \) captures the correlation between current returns and future volatility. Jacquier and Miller (2010) and Takahashi et al. (2014) found that, in the presence of microstructure noise, the posterior mean of \( \xi_1 \) deviates from the assumption of Takahashi et al. (2009) that \( \xi_1 = 1 \) for the different realized estimators.

This study presents three main contributions. First, we propose some extensions of the above R-ASV model. In contrast to the Takahashi et al. (2014) study, we apply the SKT distribution without standardization as in Nakajima and Omori (2012). Moreover, we implement another generalized Student’s t-distribution that also accommodate flexible skewness and heavy-tailedness, i.e., non-central Student’s t-distribution (NCT distribution) as in Tsiotas (2012). In the context of the SV model, Tsiotas (2012) showed that the SV model with asymmetric effect and NCT distribution can outperform the SV model with asymmetric effect and SKT distribution on model selection. Furthermore, this study proposes a class of non-linear stochastic volatility (NSV) process from Tsiot-
tas (2009) instead of the linear version as in Takahashi et al. (2014). We refer to this extension of the R-ASV framework as the realized non-linear asymmetric stochastic volatility (R-NASV) model. The empirical results from Tsiotas (2009) provide evidence that the nonlinear version is a better fit than the linear version in the SV context. This study does not only investigate the use of exponential and Yeo-Johnson transformations (as in Tsiotas (2009)) but also the use of modulus transformations for transformation of lagged volatility in Model (1.1). These transformations are indexed by the parameter $\lambda$ and were selected on the basis of the fact that these families permit transformed data to be non-positive and contain a $\lambda$ value that does not correspond to transformation (linear version) because the main idea is to transform log volatility $h_t$. Notice that in the case of the Yeo-Johnson transformation (Yeo and Johnson (2000)), interpretation of the transformation parameter is difficult as it has a different function for negative and non-negative values of transformed data. However, this family can be useful in procedures for selecting a transformation for linearity or normality (Weisberg (2014)).

Second, we apply two Hamiltonian Monte Carlo (HMC)-based methods within the Markov chain Monte Carlo (MCMC) algorithm to update the parameters that cannot be sampled directly. Our MCMC simulation employs a Hamiltonian Monte Carlo (HMC) algorithm for updating the log volatility and transformation parameter. To update the other parameters that cannot be sampled directly, the implementation is to sample these parameters using the Riemann manifold HMC (RMHMC) algorithm proposed by Girolami and Calderhead (2011). In contrast to previous studies that have proposed the single-move or multi-move Metropolis-Hastings algorithm to estimate the SV model, including Tsiotas (2009), Takahashi et al. (2009, 2014), we propose the HMC-based methods since these methods update the entire log volatility at once. The HMC method has been shown to be more appropriate than the single move Metropolis-Hastings algorithm when sampling from high-dimensional, strongly correlated target densities (e.g., Takaishi (2009), Neal (2011), Girolami and Calderhead (2011)). Moreover, Girolami and Calderhead (2011) demonstrated that, among various sampling methods (including HMC), RMHMC sampling yields the best performance for parameters and latent volatilities, in terms of time-normalized effective sample size.

Third, we perform a detailed empirical study of the R-NASV models using daily Tokyo Stock Price Index (TOPIX) data over 4-year and 8-year periods. In contrast to the Takahashi et al. (2014) study that analyses the R-ASV models on the realized kernel estimator, the RV equation in our models is fitted to the data sampled at four different frequencies. The 95% and 92.5% highest posterior density (HPD) interval proposed by Chen and Shao (1999) is used to provide evidence for the existence of some parameters that build the extended model. Similar to the empirical evidence found by Takahashi et al. (2014), we find that the 95% HPD intervals of $\xi_1$ exclude 1, which indicates the statistically significant general bias of the RV, but only for the RV data sampled at very high frequency.
Regarding the parameters of the generalized Student’s \( t \)-distribution and power transformations, the data provide some evidence supporting the NCT distribution and the non-linear volatility specifications. Thereafter, the performance of the new model is compared with those of competing new and old RSV models using a modified harmonic mean method proposed by Gelfand and Dey (1994). We find that our empirical test presents significant evidence against the linear RSV model introduced by Takahashi et al. (2009) and Takahashi et al. (2014) and particularly favors an SKT distribution. In addition, to address concerns about the sensitivity of MCMC output to prior choices, we show considerable robustness for priors of power parameter \( \lambda \) with very diffused distributional behavior.

This study is organized as follows. Section 2 introduces the four RV approaches, existing leveraged RSV models, and new specifications of RSV models. In Section 3, we develop the HMC and RMHMC algorithms to estimate the parameters and latent variables of the proposed model. Section 4 discusses the computation of the marginal likelihood. The empirical results are presented in Section 5, and in Section 6, we present our conclusions and suggestions for extending this study.

2. Volatility models

2.1. RV measures

At present, RV has become the benchmark volatility measure of intra-day high-frequency data (IHFD). RV is a non-parametric measure proposed by Andersen et al. (2001) as a proxy for non-observable integrated volatility. Various versions of the RV measure, incorporating many improvements and modifications, have been proposed previously. These include the bipower variation (BV, Barndorff-Nielsen and Shephard (2004)), the two-scales realized volatility (TSRV, Zhang et al. (2005)), and the realized kernel (Barndorff-Nielsen et al. (2008)). In this study, we compute the IHFD-based volatility measures using data sampled at four different frequencies: RV 1-min, RV 5-min, skip-one BV, and TSRV 5-min.

Suppose that a process occurring in day \( t \) is observed on a full grid \( \{0 \leq \tau_{t,0} \leq \tau_{t,1} \leq \cdots \leq \tau_{t,m}\} \) and \( p_{t,j} \) denotes the log price on the \( j \)th observation grid in day \( t \). The \( j \)th percentage intra-day return is then defined as follows:

\[
R_{t,j} = 100 \times (p_{t,j} - p_{t,j-1}).
\]

The basic daily RV is defined as the summation of the corresponding high-frequency intra-daily squared returns:

\[
RV_t = \sum_{j=1}^{m} R_{t,j}^2.
\]

Assuming the absence of jumps and microstructure noise and on the basis of quadratic variation theory, Andersen et al. (2001) showed that \( RV_t \) converges to the integrated variance of the price process \( IV_t \) as observation frequency increases.
To extract the jump component from RV, Barndorff-Nielsen and Shephard (2004) proposed an empirical estimator called bipower variation (BV) that is defined as the sum of the product of adjacent absolute high-frequency intra-daily returns:

\[ \text{BV}_t = \frac{\pi}{2} \sum_{j=2}^{m} |R_{t,j}| |R_{t,j-1}|. \]

This approach is known to be robust in the presence of jumps in prices. Andersen et al. (2005) proposed a modification of the BV estimator and called it “staggered BV”. Assuming that microstructure noise is absent, these staggered BV measures converge to the integrated conditional variance as observation frequency increases. In this study, we use the staggered (skip-one) BV defined by

\[ \text{BV}_t = \frac{\pi}{2} \cdot \left(1 - \frac{2}{m}\right)^{-1} \sum_{j=3}^{m} |R_{t,j}| |R_{t,j-2}|, \]

where the normalization factor in front of the sum reflects the loss of two observations due to the staggering.

To control independent microstructure noise, Zhang et al. (2005) suggested a TSRV estimator based on sub-sampling, averaging, and bias-correction. This estimator is the bias-adjusted average of lower frequency realized volatilities computed on K non-overlapping observation sub-grids

\[ \text{TSRV}_t = \frac{1}{K} \sum_{j=1}^{K} \text{RV}_t^{(j),(avg)} - \frac{m - K + 1}{K} \text{RV}_t^{(all)}, \]

which combines the two time scales (all) and (avg). The natural way to select the jth sub-grid, \( j = 1, \ldots, K \), is to start at \( \tau_{t,j-1} \) and subsequently observe every Kth sample point.

### 2.2. RSV models with asymmetric effect and generalized t-distribution

In the SV models, Harvey and Shephard (1996), Jacquier et al. (2004), and Yu (2005) found strong evidence supporting the asymmetrical hypothesis of stock returns, which is a negative correlation between the errors of conditional returns and conditional log-squared volatility. Takahashi et al. (2009) confirmed that the asymmetry is also crucial in the RSV model. As explained by Yu (2005), if we assume \( \text{corr}(\epsilon_t, v_{t+1}) = \rho \neq 0 \) as in Harvey and Shephard (1996) instead of assuming \( \text{corr}(\epsilon_t, v_t) = \rho \) in Jacquier et al. (2004), the SV model becomes a martingale difference sequence, by which asymmetry can be interpreted.

According to previous studies, it is convenient to write

\[ v_{t+1} = \rho \epsilon_t + \sqrt{1 - \rho^2} \eta_{t+1}, \quad \text{for} \quad t = 1, 2, \ldots, T - 1, \]

where \( \eta_{t+1} \sim \mathcal{N}(0, 1) \) and \( \text{corr}(\epsilon_t, \eta_{t+1}) = 0 \). By this transformation, \( \text{var}(v_{t+1}) = 1 \) and \( \text{corr}(v_{t+1}, \epsilon_t) = \rho \), which satisfies the specifications of the R-ASV model.
A reparameterization of the model yields

\[
\begin{aligned}
    h_{t+1} &= \alpha + \phi(h_t - \alpha) + \varphi R_t e^{-(1/2)h_t} + \psi \eta_{t+1}, \\
    h_1 &\sim \mathcal{N}(\alpha, \psi^2/(1 - \phi^2)),
\end{aligned}
\]

where \( \varphi = \rho \sigma_h \) and \( \psi^2 = (1 - \rho^2) \sigma_h^2 \).

Furthermore, many empirical studies have shown that asset returns are characterized by heavy tails (excess kurtosis) that cannot be captured by the normal distribution. Geweke (1993), Watanabe and Asai (2001), and Jacquier et al. (2004) found that the Student’s \( t \)-distribution can capture the heavy-tailedness of the conditional distribution of daily returns. However, as the return distribution may exhibit skewness, a skewed version of the Student’s \( t \)-distribution might be more adequate in some cases. Several skew Student’s \( t \)-distributions have been previously proposed; for a short overview see Aas and Haff (2006).

To accommodate flexible skewness and heavy-tailedness in the conditional distribution of returns, Barndorff-Nielsen (1977) proposed the generalized hyperbolic (GH) distribution that may be represented as a normal variance-mean mixture, which is given by

\[
T_{\mu, \beta, a, b, c} = \mu + \beta Z + \sqrt{Z} V,
\]

where \( V \sim \mathcal{N}(0, 1) \) and \( Z \sim \mathcal{IG}(a, b, c) \) are independent and \( \mathcal{IG} \) represents the generalised inverse Gaussian distribution.

A special case of the GH distribution is the SKT distribution that is simple, flexible, and easily incorporated into the SV model based on a Bayesian estimation scheme using the MCMC algorithm (Nakajima and Omori (2012)). The SKT distribution is obtained by assuming \( a = -\frac{1}{2} \nu \) and \( c = 0 \), where \( \nu \) denotes degrees of freedom (Aas and Haff (2006)). Following Nakajima and Omori (2012), we then assume \( b = \sqrt{\nu} \), which yields \( Z_\nu \sim \mathcal{IG}(\nu/2, \nu/2) \), where \( \mathcal{IG} \) denotes the inverse gamma distribution, and set \( \mu = -\beta \mu_\nu \), where \( \mu_Z = E(Z_\nu) = \frac{\nu}{\nu-2} \) and \( \nu > 4 \). These yield the SKT distribution formulated as

\[
T_{\nu, \beta} = \beta(Z_\nu - \mu_z) + \sqrt{Z_\nu} V.
\]

Its density is given as (Aas and Haff (2006))

\[
f(x) = \frac{2^{(1-\nu)/2}\nu^{\nu/2}|\beta|^{(\nu+1)/2}K_{(\nu+1)/2}(\sqrt{\beta^2(\nu + (x + \beta \mu_z))^2})\exp(\beta(x + \beta \mu_z))}{\sqrt{\pi}\Gamma \left( \frac{\nu}{2} \right) (\nu + (x + \beta \mu_z)^2)^{(\nu+1)/2}},
\]

where \( K_j \) is the modified Bessel function of the third kind of order \( j \).

Using the SKT distribution, we propose a R-ASV model with SKT-errors distribution, called the R-ASV-SKT model, that takes the following form:

\[
\begin{aligned}
    R_t &= e^{(1/2)h_t}(\beta(z_t - \mu_z) + z_t^{1/2} \zeta_t), \\
    Y_t &= \xi_0 + \xi_1 h_t + \sigma_y u_t, \\
    h_{t+1} &= \alpha + \phi(h_t - \alpha) + \varphi(R_t e^{-(1/2)h_t} - \beta(z_t - \mu_z))z_t^{-1/2} + \psi \eta_{t+1}, \\
    h_1 &\sim \mathcal{N}(\alpha_0, \psi^2/(1 - \phi^2)),
\end{aligned}
\]
where $z_t \sim IG(\nu/2, \nu/2)$ and $(\zeta_t, u_t, \eta_{t+1}) \sim \mathcal{N}(0, I_3)$. This model is an extension of the so called SVSKt model introduced by Nakajima and Omori (2012) for asymmetric SV models with SKT distribution that do not incorporate an RV equation. Incorporating an RV equation, the model is an extension of the so called ASV-RVC model introduced by Takahashi et al. (2009) for the asymmetric RSV model with normal distribution and $\xi_1 = 1$. Takahashi et al. (2014) considered a similar approach to the above model with a standardized SKT distribution the so called RSVCskt model. When $\beta = 0$, the above model reduces to the model with the central Student’s $t$-distribution (R-ASV-T model).

An alternative distribution to accommodate flexible skewness and heavy-tailedness in the conditional distribution of returns is the NCT distribution proposed by Johnson et al. (1995). If $V$ and $Q_\nu$ are (statistically) independent standard normal and chi-square random variables, respectively, the latter with $\nu$ degrees of freedom, then

$$T_{\nu, \mu} = \frac{\mu + V}{\sqrt{Q_\nu/\nu}} = (\mu + V)\sqrt{Z_\nu},$$

where $Z_\nu \sim IG(\nu/2, \nu/2)$, is said to have an NCT distribution with $\nu$ degrees of freedom and non-centrality parameter $\mu$. The probability density function of $T_{\nu, \mu}$ is

$$f(x) = e^{-\mu^2/2} \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\mu^2}{\nu + \mu^2}\right)^j \frac{1}{2} \left(\nu + j, \frac{\nu}{2}\right),$$

where $I(x | \nu, \mu)$ is the incomplete beta function with parameters $\nu$ and $\mu$.

Replacing a normal random variable $\epsilon_t$ in Model (1.1) with an NCT distribution yields a new R-ASV model called the R-ASV-NCT model, which takes the following form:

$$\begin{align*}
R_t &= e^{(1/2)h_t} z_t^{1/2} (\mu + \zeta_t), \quad z_t \sim IG(\nu/2, \nu/2), \\
Y_t &= \xi_0 + \xi_1 h_t + \sigma_y u_t, \\
h_{t+1} &= \alpha + \phi(h_t - \alpha) + \varphi(R_t e^{-(1/2)h_t} z_t^{-1/2} - \mu) + \psi \eta_{t+1}, \\
h_1 &= \mathcal{N}(\alpha, \psi^2/(1-\phi^2)), \\
(\zeta_t, u_t, \eta_{t+1}) &\sim \mathcal{N}(0, I_3).
\end{align*}$$

This model is an extension of the so called ASV-nct model introduced by Tsiotas (2012) for the asymmetric SV model with NCT distribution that do not incorporate an RV equation. When $\mu = 0$, the R-ASV-NCT model reduces to the R-ASV-T model with the same degrees of freedom.

Figure 1 shows the densities of NCT and SKT distributions proposed, respectively, in equations (2.4) and (2.2) for a few different choices of parameter sets $(\mu, \nu)$ and $(\beta, \nu)$. It is shown that the skewness and heavy-tailedness of both the NCT and SKT distributions are determined jointly by the combination of those parameters values. As mentioned, $\mu = 0$ and $\beta = 0$ correspond to a
Figure 1. The NCT and SKT densities proposed, respectively, in equations (2.4) and (2.2) generated by combinations of parameters.

central Student’s $t$-distribution. The higher value of $\mu$ and the lower value of $\beta$ imply a more positive skewness or right-skewness and a more negative skewness or left-skewness as well as heavier tails, respectively. As $\nu$ becomes larger, both distributions become less skewed and have lighter tails.
2.3. Power transformations

In statistics, data transformation is the usual method applied so that the data more closely satisfy the theoretical assumptions made in an analysis. Since Box and Cox (1964) published their seminal paper on power transformations, a number of modifications of the Box-Cox (BC) transformations have been proposed, both for theoretical work and practical applications. Because the purpose of this study is to transform log volatility $h_t$, we select transformation families from Manly (1976), John and Draper (1980), and Yeo and Johnson (2000) that consider any value in the real line and include a linear case.

Manly (1976) introduced a family of exponential transformations (ET) which is claimed to be quite effective at turning skewed unimodal distributions into nearly symmetric normal-like distributions taking the form:

$$ P_{ET}(x, \lambda) = \begin{cases} \frac{e^{\lambda x} - 1}{\lambda}, & \text{if } \lambda \neq 0; \\ x, & \text{if } \lambda = 0. \end{cases} $$

John and Draper (1980) proposed the so-called modulus transformation (MT), which is effective for normalizing a distribution already possessing approximate symmetry about some central point, and it takes the form:

$$ P_{MT}(x, \lambda) = \begin{cases} \text{sign}(x) \frac{(|x| + 1)^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0; \\ \text{sign}(x) \log(|x| + 1), & \text{if } \lambda = 0, \end{cases} $$

where $\text{sign}(x) = 1$ when $x \geq 0$ or $\text{sign}(x) = -1$ when $x < 0$. Recently, Yeo and Johnson (2000) proposed a family of power transformations:

$$ P_{YJ}(x, \lambda) = \begin{cases} \frac{(x + 1)^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0, \ x \geq 0; \\ \log(x + 1), & \text{if } \lambda = 0, \ x \geq 0; \\ \frac{(1 - x)^{2 - \lambda} - 1}{\lambda - 2}, & \text{if } \lambda \neq 2, \ x < 0; \\ - \log(-x + 1), & \text{if } \lambda = 2, \ x < 0, \end{cases} $$

which is appropriate for reducing skewness toward normality and has many good properties of the BC power transformation family for positive variables. Thus, if the Box and Cox, Manly, and Yeo-Johnson transformations are used to make skew distributions more symmetrical and normal, the transformation family proposed by John-Draper can be used to eliminate any residual (positive) kurtosis. Figure 2 shows these transformations for the values of $\lambda = -1, 0, 0.5, 1,$ and $2$.

Applying the above transformations families in the lagged state of the log volatility process in the R-ASV model, we propose an R-NASV model formulated as in equation (1.1) but the log volatility process now takes the form:

$$ h_{t+1} = \alpha_0 + \phi(P_t - \alpha_0) + \sigma_h v_{t+1}, $$
where $P_t = P(h_t, \lambda)$ represents a family of power transformations. This model is a non-linear extension of the ASV-RVC and RSVCsikt models proposed by Takahashi et al. (2009, 2014), respectively, for the asymmetric RSV model with linear volatility process. Removing RV equation and asymmetric effects, the R-NASV model reduces to the so called nl-SV model proposed by Tsiotas (2009) for basic SV model with non-linear volatility process. We then call R-ETASV, R-MTASV, and R-YJASV for an R-NASV model corresponding to the exponential, modulus, and Yeo-Johnson transformations, respectively. In addition, the values of $\lambda^{ET} = 0$, $\lambda^{MT} = 1$, and $\lambda^{YJ} = 1$ correspond to no transformation.

3. MCMC algorithm for R-NASV models

In this section, we detail the MCMC scheme using Hamiltonian dynamics for the R-NASV models.
3.1. HMC and RMHMC methods

The HMC method alternately combines Gibbs updates with Metropolis updates and avoids the random walk behavior. This method proposes a new state by computing a trajectory obeying Hamiltonian dynamics (Neal (2011)). The trajectory is guided by a first-order gradient of the log of the posterior by applying time discretization in the Hamiltonian dynamics. This gradient information encourages the HMC trajectories in the direction of high probabilities, resulting in a high-acceptance rate and ensuring that the accepted draws are not highly correlated (Marwala (2012)).

Let us consider position variables $\theta \in \mathbb{R}^D$ with density $L(\theta | y) \pi(\theta)$, where $L(\theta | y)$ is the likelihood function of given data $y$ and $\pi(\theta)$ is our prior density, and introduce an independent auxiliary variable $\omega \in \mathbb{R}^D$ with density $p(\omega) = N(\omega | 0, M)$. In a physical analogy, the negative logarithm of the joint probability density for the variables of interest, $-L(\theta) \equiv -\ln(L(\theta | y)\pi(\theta))$, denotes a potential energy function, the auxiliary variable $\omega$ is analogous to a momentum variable, and the covariance matrix $M$ denotes a mass matrix. For standard HMC, the mass matrix is set to the identity.

The Hamiltonian dynamics system is described by a function of two variables known as the Hamiltonian function, $H(\theta, \omega)$, which is a sum of the potential energy $U(\theta)$ and kinetic energy $K(\omega)$ (Neal (2011)),

$$H(\theta, \omega) = U(\theta) + K(\omega),$$

where $U(\theta) = -L(\theta) + \frac{1}{2} \log \{|(2\pi)^D|M|\}$ and $K(\omega) = \frac{1}{2} \omega' M^{-1} \omega$. The second term on the right hand side in the potential energy equation results from the normalization factor. The deterministic proposal for the position variable is obtained by solving the Hamiltonian equations for the momentum and position variables, respectively, given by

$$\frac{d\theta}{d\tau} = \frac{\partial H}{\partial \omega} = M^{-1} \omega \quad \text{and} \quad \frac{d\omega}{d\tau} = -\frac{\partial H}{\partial \theta} = \nabla_\theta L(\theta).$$

These equations determine how $\theta$ and $\omega$ change over a fictitious time $\tau$. Starting with the current state $(\theta, \omega)$, the proposed state $(\theta^*, \omega^*)$ is then accepted as the next state in the Markov chain with probability

$$P(\theta, \omega; \theta^*, \omega^*) = \min\{1, \exp\{-H(\theta^*, \omega^*) + H(\theta, \omega)\}\}.$$ 

Recently, Girolami and Calderhead (2011) proposed a new HMC method, the “RMHMC”, for improving the convergence and mixing of the chain. In their study, the covariance matrix $M$ depends on the variable $\theta$ and can be any positive definite matrix. $M(\theta)$ is chosen to be the metric tensor, i.e.,

$$M(\theta) = \text{cov} \left[ \frac{\partial}{\partial \theta} L(\theta) \right] = -E_{y|\theta} \left[ \frac{\partial^2}{\partial \theta^2} L(\theta) \right]$$

which is the sum of the expected Fisher information matrix and the negative Hessian of the log prior. Therefore, the Hamiltonian equations for the momentum
and position variables, respectively, are now defined by
\[
\frac{d\theta}{d\tau} = \frac{\partial H}{\partial \omega} = M(\theta)^{-1}\omega
\]
and
\[
\frac{d\omega}{d\tau} = -\frac{\partial H}{\partial \theta} = \nabla_\theta L(\theta) - \frac{1}{2} \text{tr} \left[ M(\theta)^{-1} \frac{\partial M(\theta)}{\partial \theta} \right] + \frac{1}{2} \omega' M(\theta)^{-1} \frac{\partial M(\theta)}{\partial \theta} M(\theta)^{-1} \omega.
\]

The above position variable equation requires calculation of the second- and third-order derivatives of \( L \). This adds to the computational complexity of the algorithm, making it infeasible in many applications.

In practice, the differential equations of Hamiltonian dynamics are often simulated in a finite number of steps using the leapfrog scheme. Neal (2011) showed that the leapfrog scheme yields better results than the standard and modified Euler’s method. To ensure that the leapfrog algorithm is performing well, the step size and number of leapfrog steps must be tuned, which can usually be achieved with some experimentation. The “generalized leapfrog algorithm” operates as follows (for a chosen step size \( \Delta\tau \) and simulation length \( N_L \)):

(i) update the momentum variable in the first half step using the equation
\[
\omega_{\tau+(1/2)\Delta\tau} = \omega_{\tau} - \frac{1}{2} \Delta\tau \frac{\partial H(\theta_{\tau}, \omega_{\tau+(1/2)\Delta\tau})}{\partial \theta},
\]
(ii) update the parameter \( \theta \) over a full time step using the equation
\[
\theta_{\tau+\Delta\tau} = \theta_{\tau} + \frac{\Delta\tau}{2} \left\{ \frac{\partial H(\theta_{\tau}, \omega_{\tau+(1/2)\Delta\tau})}{\partial \omega} + \frac{\partial H(\theta_{\tau}+\Delta\tau, \omega_{\tau+(1/2)\Delta\tau})}{\partial \omega} \right\},
\]
and
(iii) update the momentum variable in the second half step using the equation
\[
\omega_{\tau+\Delta\tau} = \omega_{\tau+(1/2)\Delta\tau} - \frac{1}{2} \Delta\tau \frac{\partial H(\theta_{\tau+\Delta\tau}, \omega_{\tau+(1/2)\Delta\tau})}{\partial \theta}.
\]

The full algorithm for the HMC or RMHMC methods can be summarized in the following three steps (for covariance matrix \( M \)).

1. Randomly draw a sample momentum vector \( \omega \sim \mathcal{N}(\omega | 0, M) \).
2. Starting with the current state \((\theta, \omega)\), run the leapfrog algorithm for \( N_L \) steps with step size \( \Delta\tau \) to generate the proposal \((\theta^*, \omega^*)\). At every leapfrog step, especially for the RMHMC algorithm, the values of \( \omega_{\tau+(1/2)\Delta\tau} \) and \( \theta_{\tau+\Delta\tau} \) are determined numerically by a fixed-point iteration method.
3. Accept \((\theta^*, \omega^*)\) with probability \( P(\theta, \omega; \theta^*, \omega^*) \), otherwise retain \((\theta, \omega)\) as the next Markov chain draw.
3.2. MCMC simulation in the R-NASV models

Consider two observation vectors \( R = \{R_i\}_{i=1}^T \) and \( Y = \{Y_i\}_{i=1}^T \), two latent variable vectors \( h = \{h_i\}_{i=1}^T \) and \( z = \{z_i\}_{i=1}^T \), and the parameter vectors \( \theta_1 = (\xi_0, \xi_1, \sigma_y) \), \( \theta_2 = (\lambda, \alpha, \phi, \varphi, \psi) \), and \( \theta_3 = (\kappa, \nu) \), where \( \kappa \) is either \( \mu \) or \( \beta \). The analysis presented here aims to obtain the estimates and other inferences of the proposed model parameters. Using Bayes’ theorem, the joint posterior density of the parameters and latent unobservable variables conditional on the observations is given as

\[
(3.1) \quad p(h, z, \theta \mid R, Y) = \pi(\theta) \times p(R \mid h, z, \theta_3) \times p(z \mid \nu) \times p(Y \mid \theta_1, h) \times p(h \mid R, \theta_2, \theta_3)
\]

\[
\propto \pi(\theta) \times \prod_{t=1}^T e^{-(1/2)h_t z_t^{-1/2}} \exp \left\{ -\frac{1}{2} \tilde{R}_t^2 \right\} \times \left( \nu \right)^{(1/2)\nu T} \Gamma \left( \frac{\nu}{2} \right) \prod_{t=1}^T (z_t)^{-(1/2)\nu-1} \exp \left\{ -\frac{\nu}{2z_t} \right\} \times (\sigma_y^2)^{-1/2} \prod_{t=1}^T [(Y_t - \xi_0 - \xi_1 h_t)^2] \times (\psi^2)^{-1/2} \prod_{t=2}^T \exp \left\{ -\frac{1}{2} \tilde{z}_t^2 \right\},
\]

where \( \pi(\theta) \) is a prior density on \( \theta \), with \( \theta = (\theta_1, \theta_2, \theta_3) \), and

\[
\tilde{h}_t = h_t - \alpha - \phi (P_{t-1} - \alpha) - \varphi \tilde{R}_{t-1}, \quad \text{for} \quad t = 2, \ldots, T,
\]

\[
\tilde{R}_t = \begin{cases} R_t e^{-(1/2)h_t z_t^{-1/2}} - \mu, & \text{for R-NASV-NCT model;} \\ R_t e^{-(1/2)h_t z_t^{-1/2} - \beta (z_t - \mu_z)} z_t^{-1/2}, & \text{for R-NASV-SKT model.} \end{cases}
\]

By convention, we assume the following priors for the unknown structural parameters:

\[
\begin{align*}
\lambda & \sim \mathcal{N}(m_\lambda, v_\lambda), \quad \kappa \sim \mathcal{N}(m_\kappa, v_\kappa), \quad \nu \sim \mathcal{G}(a_\nu, b_\nu), \\
\xi_0 & \sim \mathcal{N}(m_{\xi_0}, v_{\xi_0}), \quad \xi_1 \sim \mathcal{N}(m_{\xi_1}, v_{\xi_1}), \quad \sigma_y^2 \sim \mathcal{IG}(a_{\sigma_y}/2, b_{\sigma_y}/2), \\
\alpha & \sim \mathcal{N}(m_\alpha, v_\alpha), \quad \phi \sim \mathcal{B}(A, B), \quad \varphi \sim \mathcal{N}(m_\varphi, v_\varphi), \quad \psi^2 \sim \mathcal{IG}(a_\psi/2, b_{\psi}/2),
\end{align*}
\]

where \( \mathcal{B} \) and \( \mathcal{G} \) represent the beta and gamma densities, respectively. This choice of priors ensures that all parameters have the right support; in particular, the beta prior for \( \phi \) would produce \(-1 < \phi < 1\) that ensures stationary conditions for the volatility process.

3.2.1. Updating parameters \((\theta_1, \alpha, \varphi, \psi, \kappa)\)

The MCMC method simulates a new value for each parameter from its conditional posterior density, assuming that the current values for the other parameters
are true values. We implement the MCMC scheme by first simulating random
draws that can be sampled directly from their conditional posteriors:

\[
\begin{align*}
\xi_0 \mid \xi_1, \sigma_y, h, Y & \sim \mathcal{N}(M_{\xi_0}, V_{\xi_0}), \\
\xi_1 \mid \xi_0, \sigma_y, h, Y & \sim \mathcal{N}(M_{\xi_1}, V_{\xi_1}), \\
\sigma_y^2 \mid \xi_0, \xi_1, h, Y & \sim \mathcal{IG}(d_{\sigma_y}/2, D_{\sigma_y}/2), \\
\alpha \mid \lambda, \phi, \varphi, \psi, \theta_3, h, z, R, Y & \sim \mathcal{N}(M_{\alpha}, V_{\alpha}), \\
\varphi \mid \lambda, \alpha, \phi, \psi, \theta_3, h, z, R, Y & \sim \mathcal{N}(M_{\varphi}, V_{\varphi}), \\
\psi^2 \mid \lambda, \alpha, \phi, \varphi, h, z, R, Y & \sim \mathcal{IG}(d_\psi/2, D_\psi/2), \\
\kappa \mid \theta_2, \nu, h, z, R & \sim \mathcal{N}(M_\kappa, V_\kappa),
\end{align*}
\]

where

\[
\begin{align*}
V_{\xi_0} & = [v_{\xi_0}^{-1} + T\sigma_y^{-2}]^{-1}, \\
M_{\xi_0} & = V_{\xi_0} \left[ m_{\xi_0}v_{\xi_0}^{-1} + \sigma_y^{-2} \sum_{t=1}^{T}(Y_t - \xi_t h_t) \right], \\
V_{\xi_1} & = \left[ v_{\xi_1}^{-1} + \sigma_y^{-2} \sum_{t=1}^{T} h_t^2 \right]^{-1}, \\
M_{\xi_1} & = V_{\xi_1} \left[ m_{\xi_1}v_{\xi_1}^{-1} + \sigma_y^{-2} \sum_{t=1}^{T} h_t(Y_t - \xi_0) \right], \\
d_{\sigma_y} & = a_{\sigma_y} + T, \\
D_{\sigma_y} & = b_{\sigma_y} + \sum_{t=1}^{T}(Y_t - \xi_0 - \xi_1 h_t)^2, \\
V_{\alpha} & = [v_{\alpha}^{-1} + ((1 - \phi^2) + (T - 1)(1 - \phi)^2)\psi^{-2}]^{-1}, \\
M_{\alpha} & = V_{\alpha} \left[ m_{\alpha}v_{\alpha}^{-1} + (1 - \phi^2)h_1 + (1 - \phi) \sum_{t=2}^{T}(h_t - \phi:\tilde{P}_{t-1} - \varphi:\tilde{R}_{t-1}) \right], \\
V_{\varphi} & = \left[ v_{\varphi}^{-1} + \psi^{-2} \sum_{t=2}^{T} \tilde{R}_{t-1} \right]^{-1}, \\
M_{\varphi} & = V_{\varphi} \left[ m_{\varphi}v_{\varphi}^{-1} + \psi^{-2} \sum_{t=2}^{T} (h_t - \alpha - \phi(\tilde{P}_{t-1} - \alpha))\tilde{R}_{t-1} \right], \\
d_{\psi} & = a_\varphi + T, \\
D_{\psi} & = b_\varphi + (1 - \phi^2)(h_1 - \alpha)^2 + \sum_{t=2}^{T} \tilde{h}_t^2, \\
V_{\kappa} & = \left[ v_{\kappa}^{-1} + \left(1 + \frac{\varphi^2}{\psi^2}\right) \sum_{t=1}^{T-1} f^2(z_t) + f^2(z_T) \right]^{-1}, \\
M_{\kappa} & = V_{\kappa} \left[ m_\kappa v_\kappa^{-1} + \sum_{t=1}^{T} R_t e^{-(1/2)h_t z_t^{-1/2}} f(z_t) \right].
\end{align*}
\]
\[-\frac{\phi}{\psi^2} \sum_{t=2}^{T} (h_t - \alpha - \phi (P_{t-1} - \alpha) - \varphi R_{t-1} e^{-(1/2)h_{t-1}z_{t-1}^{-1/2}}) f(z_{t-1}) \],

where \( f(z_t) = 1 \) for the R-NASV-NCT model and \( f(z_t) = (z_t - \mu_z) z_t^{-1/2} \) for the R-NASV-SKT model.

In the following, we study ways to update the other parameters and latent variables that are unable to be sampled directly.

### 3.2.2. Updating parameter \( \lambda \)

On the basis of the joint density (3.1), the logarithm of the full conditional posterior density of \( \lambda \) is represented by

\[ \mathcal{L}(\lambda) \propto -\frac{1}{2} v_\lambda^2 (\lambda - m_\lambda)^2 - \frac{1}{2} \psi^2 \sum_{t=2}^{T} \tilde{h}_t^2, \]

which is not of standard form, and therefore we cannot sample from it directly. The RMHMC sampling scheme is not applicable to sample \( \lambda \) because the metric tensor required to implement the RMHMC sampling scheme cannot be explicitly derived from the above density. Therefore, we use the HMC algorithm for estimating the power parameter \( \lambda \). This, specially in the leapfrog algorithm, requires evaluation of only the first partial derivative of the log posterior with respect to \( \lambda \):

\[ \nabla_\lambda \mathcal{L}(\lambda) = -\frac{\lambda - m_\lambda}{v_\lambda^2} + \frac{\phi}{\psi^2} \sum_{t=2}^{T} [\tilde{h}_t - \tilde{R}_{t-1}] \frac{\partial \mathcal{P}(h_{t-1}, \lambda)}{\partial \lambda}. \]

### 3.2.3. Updating parameter \( \phi \)

An inspection of the joint density reveals that the logarithm of the full conditional posterior of \( \phi \), which is given by

\[ \mathcal{L}(\phi) \propto \frac{1}{2} \ln(1 - \phi^2) - \frac{1}{2} \psi^2 (1 - \phi^2)(h_1 - \alpha)^2 - \frac{1}{2} \psi^2 \sum_{t=2}^{T} \tilde{h}_t^2 + (A - 1) \ln \phi + (B - 1) \ln(1 - \phi), \]

and is in non-standard form; thus, so it is not straightforward to sample \( \phi \) from this posterior. To sample \( \phi \), we must be able to employ the RMHMC algorithm, and it is necessary to implement the transformation \( \phi = \tanh(\hat{\phi}) \) for dealing with the constraint \(-1 < \phi < 1\) so that it is restricted to the stationary region. The important aspect is that the leapfrog algorithm requires evaluations of the gradient vector and metric tensor of the log posterior. The partial derivative of the above log posterior with respect to \( \hat{\phi} \) is as follows:

\[ \nabla_{\phi} \mathcal{L}(\phi) = -\phi + \frac{\phi}{\psi^2} (h_1 - \alpha)^2 (1 - \phi^2) + \frac{1}{\psi^2} \sum_{t=2}^{T} \tilde{h}_t (P_{t-1} - \alpha)(1 - \phi^2) \]

\[ + \frac{(A - 1)(1 - \phi^2)}{\phi} - (B - 1)(1 + \phi). \]
Then, the metric tensor and its partial derivative, respectively, are

\[ M(\phi) = 2\phi^2 - (T - 1)(\phi^2 - 1) - \left[ (1 - A) \left( 1 + \frac{1}{\phi^2} \right) - (B - 1) \right] (1 - \phi^2) \]

and

\[ \frac{\partial M(\phi)}{\partial \phi} = [4\phi - 2(T - 1)\phi](1 - \phi^2) - \left[ \frac{2(A - 1)}{\phi^3} + 2\phi(B - 1) \right] (1 - \phi^2). \]

### 3.2.4. Updating parameter \( \nu \)

The RMHMC algorithm is applicable to sample \( \nu \). Evaluations of gradient vector and metric tensor of the log posterior required in the leapfrog algorithm are as follows. The log posterior of \( \nu \) is given by

\[ L(\nu) \propto \frac{1}{2} T \nu \ln \left( \frac{1}{\nu^2} \right) - T \ln \Gamma \left( \frac{1}{2} \nu \right) - \frac{1}{2} \nu \sum_{t=1}^{T} [\ln(z_t) + z_t^{-1}] \]

\[ + (a_\nu - 1) \ln(\nu) - b_\nu \nu - q(\nu), \]

where

\[ q(\nu) = \begin{cases} 
0, & \text{for R-NASV-NCT model;} \\
\frac{1}{2} \sum_{t=1}^{T} \tilde{R}_t^2 + \frac{1}{2\psi^2} \sum_{t=2}^{T} \tilde{h}_t^2, & \text{for R-NASV-SKT model,} 
\end{cases} \]

and the required gradient is given by

\[ \nabla_\nu L(\nu) = \frac{1}{2} T \left[ \ln \left( \frac{\nu}{2} \right) + 1 \right] - \frac{1}{2} T \Psi \left( \frac{\nu}{2} \right) - \frac{1}{2} \sum_{t=1}^{T} [\ln(z_t) + z_t^{-1}] \]

\[ + \frac{a_\nu - 1}{\nu} - b_\nu - \frac{\partial q(\nu)}{\partial \nu}, \]

where \( \Psi(x) \) is a digamma function defined by \( \Psi(x) = d \ln \Gamma(x) / dx \) and

\[ \frac{\partial q(\nu)}{\partial \nu} = \begin{cases} 
0, & \text{for R-NASV-NCT model;} \\
- \frac{2\beta}{(\nu - 2)^2} \left( \sum_{t=1}^{T} \tilde{R}_t \tilde{z}_t^{-1/2} - \frac{\varphi}{\psi^2} \sum_{t=2}^{T} \tilde{h}_t \tilde{z}_t^{-1/2} \right), & \text{for R-NASV-SKT model.} 
\end{cases} \]

We then obtain the metric tensor \( M(\nu) \) and its partial derivative with respect to \( \nu \), respectively, as

\[ M(\nu) = -\frac{T}{2\nu} + \frac{T}{4} \Psi' \left( \frac{\nu}{2} \right) + \frac{a_\nu - 1}{\nu^2} + S(\nu), \]

\[ \frac{\partial M}{\partial \nu} = \frac{T}{2\nu^2} + \frac{T}{8} \Psi'' \left( \frac{\nu}{2} \right) - \frac{2(a_\nu - 1)}{\nu^3} - \frac{4}{\nu - 2} S(\nu). \]
where $\Psi'(x)$ is a trigamma function defined by $\Psi'(x) = d\Psi(x)/dx$, and $\Psi''(x)$ is a tetragamma function defined by $\Psi''(x) = d\Psi'(x)/dx$. The function $S(\nu)$ is defined by

$$S(\nu) = \begin{cases} 0, & \text{for R-NASV-NCT model;} \\ \frac{4\beta^2}{(\nu - 2)^4} \left( T + (T - 1) \frac{\varphi^2}{\psi^2} \right), & \text{for R-NASV-SKT model.} \end{cases}$$

### 3.2.5. Updating the latent variable $\mathbf{z}$

The logarithm of the full conditional posterior density of the latent variable $\mathbf{z}$ is

$$L(z) \propto \sum_{t=1}^{T} \left[ -\frac{\nu + 3}{2} \ln(z_t) - \frac{1}{2} \left( \tilde{R}_t^2 - \frac{\nu}{z_t} \right) - \frac{1}{2\psi^2} \sum_{t=2}^{T} \tilde{h}_t^2 \right].$$

From this conditional density, draws can be obtained using the RMHMC algorithm and implementing the transformation $z_t = \exp(\hat{z}_t)$ to ensure constrained sampling for $z_t > 0$. The gradient vector and metric tensor of the log posterior required in the leapfrog algorithm are then evaluated as follows. The first partial derivative of the log posterior with respect to $\hat{z}_t$ is as follows:

$$\nabla_{\hat{z}_t} L(z_t) = -\frac{\nu + 3}{2} + \frac{\nu}{2z_t} - \tilde{R}_t \frac{\partial \tilde{R}_t}{\partial \hat{z}_t} + \frac{\varphi}{\psi} \tilde{h}_{t+1} \frac{\partial \tilde{R}_t}{\partial \hat{z}_t} I_{t<T},$$

where $I$ is an indicator function and

$$\frac{\partial \tilde{R}_t}{\partial \hat{z}_t} = \begin{cases} -\frac{1}{2} R_t e^{-(1/2)h_t} z_t^{-1/2}, & \text{for R-NASV-NCT model;} \\ -\beta z_t^{1/2} - \frac{1}{2} \tilde{R}_t, & \text{for R-NASV-SKT model.} \end{cases}$$

Furthermore, the metric tensor $M(z)$ is a diagonal matrix whose diagonal entries are given by

$$M(t,t) = \frac{\nu + 1}{2} + \begin{cases} \frac{\mu^2}{4} + \frac{\varphi^2}{4\psi^2} (1 + \mu^2) I_{t<T}, & \text{for R-NASV-NCT model;} \\ \frac{\beta^2 \nu}{\nu - 2} + \frac{\varphi^2}{4\psi^2} \left( 1 + \frac{4\beta^2 \nu}{\nu - 2} \right) I_{t<T}, & \text{for R-NASV-SKT model.} \end{cases}$$

Since the above metric tensor is not a function of the latent variable $\mathbf{z}$, the associated partial derivatives with respect to the transformed latent variable are zero.

### 3.2.6. Updating the log volatilities $h$

Obviously, the posterior density of the latent volatility $h_t$ is in the non-standard form or well-known density. To sample the latent variable $h$, we employ the HMC sampling scheme since the metric tensor cannot be explicitly derived from the log posterior. Therefore, the leapfrog algorithm requires evaluation
of only the first partial derivative of the log posterior with respect to \( h_t \). The logarithm of the full conditional posterior of \( h \) is expressed as

\[
L(h) \propto -\frac{1}{2} \sum_{t=1}^{T} h_t^2 - \frac{1}{2} \sum_{t=1}^{T} \tilde{R}_t^2 - \frac{1}{2\sigma_y^2} \sum_{t=1}^{T} (Y_t - \xi_0 - \xi_1 h_t)^2 - \frac{1 - \phi^2}{2\psi^2} (h_1 - \alpha)^2 - \frac{1}{2\psi^2} \sum_{t=2}^{T} \tilde{h}_t^2,
\]

and its partial derivatives with respect to \( h_t \) have the following expressions:

\[
\nabla_{h_t} L(h) = \nabla_{h_t} L_R + \nabla_{h_t} L_Y - \frac{1 - \phi^2}{\psi^2} (h_1 - \alpha) + \frac{1}{\psi^2} \left( \phi \frac{\partial P(h_1, \lambda)}{\partial h_1} + \varphi \frac{\partial \tilde{R}_1}{\partial h_1} \right) \tilde{h}_2
\]

\[
\nabla_{h_t} L(h) = \nabla_{h_t} L_R + \nabla_{h_t} L_Y + \frac{1}{\psi^2} \left( \phi \frac{\partial P(h_i, \lambda)}{\partial h_i} + \alpha \frac{\partial \tilde{R}_i}{\partial h_i} \right) \tilde{h}_{i+1} - \frac{1}{\psi^2} \tilde{h}_i,
\]

\[
\nabla_{h_T} L(h) = \nabla_{h_T} L_R + \nabla_{h_T} L_Y - \frac{1}{\psi^2} \tilde{h}_T,
\]

for \( 1 < i < T \). Derivatives \( \frac{\partial \tilde{R}_t}{\partial h_t} \), \( \nabla_{h_t} L_R \), and \( \nabla_{h_t} L_Y \), for \( t = 1, \ldots, T \), are defined by

\[
\frac{\partial \tilde{R}_t}{\partial h_t} = -\frac{1}{2} R_t e^{-(1/2) h_t^2 z_t^{-1/2}}, \quad \nabla_{h_t} L_R = -\frac{1}{2} - \tilde{R}_t \frac{\partial \tilde{R}_t}{\partial h_t}, \quad \text{and}
\]

\[
\nabla_{h_t} L_Y = \frac{\beta_1}{\sigma_y^2} (Y_t - \xi_0 - \xi_1 h_t),
\]

respectively.

4. Marginal likelihood and Bayes factors

The fundamental quantity in the Bayesian model comparison is the marginal density of the observed data (also known as the integrated likelihood or evidence or marginal likelihood). A higher marginal likelihood for a given model indicates a better fit of the data by that model. For certain types of posterior simulators, several approximating methods for estimating the marginal likelihood from the MCMC output have been proposed, including Geweke’s estimator for importance sampling, Chib’s estimator for Gibbs sampling, Chib-Jeliazkov’s estimator for the Metropolis-Hastings algorithm, and Meng-Wong’s estimator for a general theoretical perspective (Geweke and Whiteman (2006)). Another estimator, which is simpler, faster, and general, was proposed by Gelfand and Dey (1994).

In our model framework, the marginal likelihood of the Gelfand-Dey’s (GD) method is given by

\[
m_{GD}(X) = \left[ \int_{\theta} f(\theta, H) \frac{L(X | \theta, H) \pi(\theta, H) p(\theta, H) d(\theta, H)}{L(X | \theta, H) \pi(\theta, H) p(\theta, H) d(\theta, H)} \right]^{-1},
\]
where $X$ is the matrix of the data, $H$ is the matrix of the latent variables, $f(\cdot)$ can be any probability density function with the domain contained in the posterior probability density $\Theta$, and $p(\theta, H)$ is the prior density for $(\theta, H)$ with respect to the Lebesgue measure. For computational convenience, we set $f(\theta, H) = f(\theta)f(H)$, where $f(H) = \pi(H)$ because the latent volatility $H$ is high-dimensional. Then the marginal likelihood can be estimated by

$$
\hat{m}_{GD} \approx \left[ \frac{1}{N} \sum_{j=1}^{N} \frac{f(\theta^{(j)})}{L(X | \theta^{(j)}, H^{(j)})\pi(\theta^{(j)})} \right]^{-1},
$$

where $\theta^{(j)}$ and $H^{(j)}$ are the draws from the posterior density and we have used the fact that $p(\theta)$ and $p(H^{(j)})$ are independent.

As explained by Geweke (1999), if $f(\theta)$ is thin tailed relative to the likelihood function, then $f(\theta)/L(X | \theta, H)$ is bounded above and the estimator is consistent. Therefore, following Geweke’s (1999) suggestion, we choose $f(\theta)$ as a thin tailed truncated normal distribution $\mathcal{N}(\theta^*, \Sigma^*)$, where $\theta^*$ and $\Sigma^*$ are the posterior mean and covariance matrix of the $\theta$ draws, respectively. The domain of the truncated normal, $\Theta$, is then constructed as

$$
\Theta = \{ \theta : (\theta^{(j)} - \theta^*)'(\Sigma^*)^{-1}(\theta^{(j)} - \theta^*) \leq \chi^2_{.99}(D) \},
$$

where $D$ is the dimension of the parameter vector and $\chi^2_{.99}(D)$ is the 99th percentile of the chi-squared distribution with $D$ degrees of freedom. According to $\Theta$, the normalizing constant of $f(\theta)$ is $1/\chi_{.99}(D)$ (Koop et al. (2007)).

Next, the prior density $\pi(\theta^{(j)})$ can be directly evaluated and $L(X | \theta^{(j)}, H^{(j)})$ is calculated by substituting $\theta^{(j)}$ and $H^{(j)}$ into the likelihood function. The log likelihood function of our model can be expressed as

$$
\mathcal{L}(R, Y | \theta, h, z) = -T \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln(\hat{\Sigma}_t) - \frac{1}{2} \sum_{t=1}^{T-1} (\hat{R}_t - \hat{\hat{R}}_t)^2 \hat{\Sigma}_t^{-1}
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quan
model $\mathcal{M}_1$ with respect to the null model $\mathcal{M}_0$, we consider the ratio of their marginal likelihoods:

$$BF_{10} = \frac{m_{GD}(X \mid \mathcal{M}_1)}{m_{GD}(X \mid \mathcal{M}_0)}.$$ 

On the basis of their similarity to the likelihood ratio statistic, general guidelines for the interpretation of Bayes factors were suggested by Kass and Raftery (1995). When the log Bayes factor is greater than 1 and less than 3, the null model is positively favored; when the log Bayes factor is greater than 3 and less than 5, the null model is strongly favored; and when the log Bayes factor is greater than 5, the null model is very strongly favored.

5. **Empirical results on real data sets**

This section applies the R-NASV models and the MCMC algorithm discussed in the previous section using TOPIX data over 4-year and 8-year periods. The data consist of intra-day high frequency observations from January 5, 2004 to December 30, 2011, excluding weekends and holidays.

5.1. **Data in the evaluation period**

The TOPIX data analysed in this study were divided into two periods: from January 2004 to December 2007 (984 trading days) and from January 2004 to December 2011 (1962 trading days). The asset price data were sampled at a frequency of 1 min when the market was open. The $t$th percentage continuously compounded daily returns $R_t$ is calculated as the difference between the logarithm of the $t$th day’s closing price $P_t$ and the logarithm of the $(t-1)$th day’s closing price $P_{t-1}$, which is

$$R_t = 100 \times \left[ \ln(P_t) - \ln(P_{t-1}) \right].$$

For extending the sampled volatility to a full-day volatility measure, Hansen and Lunde (2005) defined

$$RV_{t}^{HL} = c \cdot RV_{t}^{(open)}, \quad c = \frac{\sum_{t=1}^{T}(R_t - \bar{R})}{\sum_{t=1}^{T} RV_{t}^{(open)}},$$

as a measure of the volatility on day $t$. We apply this adjustment to the four classes of RV, which are denoted as $RV1^{HL}$ for a 1-min sub-sampled RV, $RV5^{HL}$ for a 5-min sub-sampled RV, $BV1^{HL}$ for a skip-one BV, and $TSRV5^{HL}$ for a 5-min sub-sampled TSRV.

Figure 3 displays the time series plots of return and log RV data. The descriptive statistics such as mean, standard deviation (SD), skewness, kurtosis, the Jarque-Bera (JB) normality test, and the Ljung-Box (LB) correlation test are listed in Table 1. From the JB and LB tests, the daily returns of TOPIX 2004–2007 and 2004–2011 are neither normally distributed nor serially correlated up to order 8, whereas the log RVs are significantly autocorrelated and some are not normally distributed. Both returns series are skewed left. The kurtosis value
Figure 3. Time series plots of percentage daily returns and RVs of the TOPIX data from January 2004 to December 2011.

of the returns for both series is significantly greater than 3, indicating that the returns distribution is peaked relative to the normal (leptokurtic). Therefore, we fit the generalized Student’s $t$-distributions to the error distribution in our model.

5.2. MCMC setup and efficiency of simulators

In the application of the MCMC algorithm, a set of prior densities required in the joint prior density is specified as in Table 2. In the SV model with an SKT distribution, Nakajima and Omori (2012) found that the posterior estimates of $\beta$ and $\nu$ are more sensitive to the choice of prior distribution for $\nu$ than other parameters. The prior distribution of $\nu$ with a higher mean value results in its higher posterior means, and this would lead to an even lower posterior mean of $\beta$ so as to retain some skewness and heavy tailedness in the empirical return distribution, as shown in Fig. 1. For the transformation parameter $\lambda$, Tsiotas
Table 1. Descriptive statistics of daily returns and the logarithm of realized volatilities in the TOPIX data sets.

<table>
<thead>
<tr>
<th>Period</th>
<th>Data</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>JB (Normality)</th>
<th>LB(8) (Autocorr.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004/1/6–2011/12/30</td>
<td><em>R</em></td>
<td>-0.02</td>
<td>1.48</td>
<td>-0.41</td>
<td>11.25</td>
<td>5611.14 (No)</td>
<td>9.03 (No)</td>
</tr>
<tr>
<td></td>
<td>ln(RV1&lt;sup&gt;HL&lt;/sup&gt;)</td>
<td>0.32</td>
<td>0.83</td>
<td>0.83</td>
<td>4.55</td>
<td>420.13 (No)</td>
<td>8634.09 (Yes)</td>
</tr>
<tr>
<td></td>
<td>ln(RV5&lt;sup&gt;HL&lt;/sup&gt;)</td>
<td>0.25</td>
<td>0.93</td>
<td>0.48</td>
<td>4.04</td>
<td>162.71 (No)</td>
<td>8332.54 (Yes)</td>
</tr>
<tr>
<td></td>
<td>ln(BV1&lt;sup&gt;HL&lt;/sup&gt;)</td>
<td>0.40</td>
<td>0.76</td>
<td>0.81</td>
<td>4.66</td>
<td>437.01 (No)</td>
<td>9094.33 (Yes)</td>
</tr>
<tr>
<td></td>
<td>ln(TSRV5&lt;sup&gt;HL&lt;/sup&gt;)</td>
<td>0.23</td>
<td>0.95</td>
<td>0.43</td>
<td>4.07</td>
<td>153.01 (No)</td>
<td>8354.23 (Yes)</td>
</tr>
<tr>
<td>2004/1/6–2007/12/28</td>
<td><em>R</em></td>
<td>0.03</td>
<td>1.08</td>
<td>-0.47</td>
<td>5.28</td>
<td>248.51 (No)</td>
<td>11.55 (No)</td>
</tr>
<tr>
<td></td>
<td>ln(RV1&lt;sup&gt;HL&lt;/sup&gt;)</td>
<td>0.05</td>
<td>0.65</td>
<td>0.35</td>
<td>3.64</td>
<td>36.82 (No)</td>
<td>3685.25 (Yes)</td>
</tr>
<tr>
<td></td>
<td>ln(RV5&lt;sup&gt;HL&lt;/sup&gt;)</td>
<td>-0.08</td>
<td>0.81</td>
<td>0.00</td>
<td>3.22</td>
<td>1.98 (Yes)</td>
<td>3371.64 (Yes)</td>
</tr>
<tr>
<td></td>
<td>ln(BV1&lt;sup&gt;HL&lt;/sup&gt;)</td>
<td>0.23</td>
<td>0.62</td>
<td>0.14</td>
<td>3.18</td>
<td>4.41 (Yes)</td>
<td>4041.57 (Yes)</td>
</tr>
<tr>
<td></td>
<td>ln(TSRV5&lt;sup&gt;HL&lt;/sup&gt;)</td>
<td>-0.08</td>
<td>0.83</td>
<td>-0.08</td>
<td>3.25</td>
<td>3.50 (Yes)</td>
<td>3341.18 (Yes)</td>
</tr>
</tbody>
</table>

Table 2. Prior densities, means, and standard deviations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Density</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>α, ξ₀, ξ₁, φ</td>
<td><em>N</em></td>
<td>0</td>
<td>√10</td>
</tr>
<tr>
<td>λ, κ</td>
<td><em>N</em></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>φ</td>
<td><em>B</em>(30,1.5)</td>
<td>0.952</td>
<td>0.037</td>
</tr>
<tr>
<td>ν</td>
<td><em>G</em>(16,0.8)</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>σ&lt;sub&gt;y&lt;/sub&gt;, ψ&lt;sup&gt;2&lt;/sup&gt;</td>
<td><em>IG</em>(5,0.2)</td>
<td>0.05</td>
<td>0.029</td>
</tr>
</tbody>
</table>

(2009) showed that in the case of the YJ transformation family, the performance of the λ posterior simulations appears to be quite robust.

Next, the MCMC algorithm is run for 15,000 iterations using code written in Matlab, where the first 5,000 draws are discarded as a burn-in period. From the resulting N = 10,000 draws for each parameter, we calculate the posterior means, SDs, 95% and 92.5% highest posterior density (HPD) interval, and Geweke’s (1992) convergence diagnostic (G-CD) statistics on the initial 10% and the last 50% of the draws. To find out how fast our Markov chain converges to its stationary distribution regardless of the initial state, we estimate a quantity so called the inefficiency factor (IF). IF can be roughly interpreted as the number of MCMC iterations required to produce independent draws. When the IF is equal to m, we need to draw MCMC samples m times as many as the number of uncorrelated samples. A value of one indicates that the draws are uncorrelated while large values indicate a slow mixing. In this study, the IF is particularly estimated as the numerical variance (square of numerical standard error) of the sample mean from the MCMC sampling scheme divided by the variance of the posterior sample mean, where the numerical standard error is computed using a Parzen window (see Kim et al. (1998) for details) with a bandwidth of 10% of the simulated draws.

The performance of the HMC and RMHMC samplers is highly sensitive to two user-specified parameters, i.e., a step size Δ<sub>τ</sub> and a desired number of
leapfrog iterations $N_L$. In addition, the RMHMC implementation requires a number of fixed point iterations. A bad choice of these parameters may result in slow mixing or incur a high computational cost in the algorithm. The selection of parameter values is particularly problematic, and there is no general guidance on how these values might best be chosen. Therefore, we tune our choices based on their acceptance rate and IF estimate from preliminary MCMC runs. Table 3 presents the parameter values used in our HMC and RMHMC implementations, in which optimal acceptance probabilities are achieved.

With the above MCMC setup, we checked the mixing performance of the samplers. In Table 4, we report the range of minimum, mean, and maximum of IF values for the estimated $h_t$ series in all cases. The results show that the IFs are quite small, typically less than 50, suggesting that the sampler is highly efficient. For example, the IF plot for the latent volatility series of the R-MTASV-SKT model obtained using RV_1^{HL} 2004–2011 is displayed in Fig. 4. Thus, the HMC sampler can reliably estimate all latent variables in the R-NASV models.

For the other parameters (for example, see Tables 5 and 6), we found that the samplers are also efficient, producing small IFs with values less than 105. Our results particularly show that the highest IF value for parameter $\nu$ in the RMHMC sampler does not exceed the smallest IF value in the multi-move sampler reported in Takahashi et al. (2014) for estimating the similar models, that is the models with SKT distribution and general non-linear bias correction. It indicates that the use of the RMHMC sampler to sample $\nu$ is more efficient than the multi-move sampler. In contrast, the RMHMC sampler for sampling $\phi$ is less efficient than the multi-move sampler.

### Table 3. Tuning parameters for the HMC and RMHMC implementations in the R-NASV models. $N_{FP1}$ and AcR denote the number of fixed point iterations and acceptance rate, respectively, which has been measured every 100 iterations for 15,000 iterations.

<table>
<thead>
<tr>
<th>Parameter of algorithm</th>
<th>Parameter of model</th>
<th>$\lambda$</th>
<th>$\phi$</th>
<th>$\nu$</th>
<th>$z$</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_L$</td>
<td></td>
<td>100</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_t$</td>
<td></td>
<td>0.001</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.01</td>
</tr>
<tr>
<td>$N_{FP1}$</td>
<td></td>
<td>—</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>—</td>
</tr>
<tr>
<td>AcR</td>
<td></td>
<td>0.97–1.00</td>
<td>0.77–0.99</td>
<td>0.93–1.00</td>
<td>1.00</td>
<td>0.92–1.00</td>
</tr>
</tbody>
</table>

### Table 4. The range of IF values for latent volatilities $h_t$ in the HMC sampler on the R-NASV models using all data sets. The statistics have been measured for applying all transformations families, RV estimators, and data sets.

<table>
<thead>
<tr>
<th>Model</th>
<th>IF value of $h_t$</th>
<th>min.</th>
<th>mean</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-NASV</td>
<td>1.1–1.5</td>
<td>2.2–5.0</td>
<td>12.5–31.3</td>
<td></td>
</tr>
<tr>
<td>R-NASV-T</td>
<td>1.2–2.0</td>
<td>2.5–6.9</td>
<td>15.6–34.3</td>
<td></td>
</tr>
<tr>
<td>R-NASV-NCT</td>
<td>1.2–2.1</td>
<td>2.7–7.8</td>
<td>12.9–37.5</td>
<td></td>
</tr>
<tr>
<td>R-NASV-SKT</td>
<td>1.8–3.7</td>
<td>3.2–9.1</td>
<td>16.9–42.1</td>
<td></td>
</tr>
</tbody>
</table>
The autocorrelation and trace plots of the posterior samples for the R-MTASV-SKT model adopting RV1^{HL} over the time period 2004–2011 are displayed in Figs. 5 and 6, respectively. The autocorrelation plots show quick decay of the autocorrelation as time lag between samples increases, indicating that the process is stationary. Trace plots of samples indicate that the chains fluctuate to be around their means, indicating that chains could have reached the right distribution. We can conclude that the mixing of chains is quite good here. Thus, the HMC and RMHMC samplers can reliably estimate parameters in the R-NASV models.

In addition, all empirical results were obtained via implementation of code in MATLAB 2011b (running in Microsoft Windows 7), on a desktop computer incorporating an Intel Xeon 3.47GHz hexa-core CPU with 16GB RAM. The real computational time for the MCMC is approximately two, nine, and eight minutes for the models with ET, MT, and YJ specifications, respectively, when using the 2004–2007 data and approximately five, seventeen, and fifteen minutes for the models with ET, MT, and YJ specifications, respectively, when using the 2004–2011 data.

5.3. Parameter estimates

In this subsection, we concentrate on the key parameters that build the extension models. Tables 5 and 6 summarize the posterior simulation results of parameters in the R-NASV-NCT and R-NASV-SKT models for the 1-minute RV data set. These are derived from the models with a value of \( \lambda \) corresponding to no transformation that does not lie in 92.5% HPD interval. The posterior results for the other RV data sets are not presented because of space constraints.

Because the volatility process has been specified by a power transformation family encompassing both non-linear and linear cases, we observe the posterior
mean and HPD interval of $\lambda$ for helping determine which transformation is suitable for the data. First, the posterior mean of $\lambda$ appears to suggest that the assumption of no transformation in the lagged log volatility series is firmly rejected for all R-NASV models in each period and RV data. This result indicates that the proposed transformations are more suitable than a logarithmic transformation to transform the lagged log volatility series using a certain specified $\lambda$.

Second, in terms of the HPD interval, the 92.5% intervals of $\lambda$ computed from the 2004–2007 data suggest to transform the lagged log volatility series
using the non-linear specification of exponential ($\lambda \neq 0$) and YJ ($\lambda \neq 1$) transformations. Meanwhile, in the MT specification, the 90% HPD interval of $\lambda$ includes 1, suggesting that the lagged log volatility series is left untransformed. When the 95% HPD interval of $\lambda$ is considered, the result is statistically significant for the R-YJASV model only. The results of the 2004–2011 data have a different pattern from the RV estimator. Considering the 92.5% HPD interval of $\lambda$, several facts emerge as follows. The RV1$^{HL}$ estimator favors the ET and MT specifications in all models with any distribution and the YJ specification in models with an SKT distribution only. When applying the BV1$^{HL}$ estimator, all
non-linear specifications seem to underestimate the HPD intervals. In the model adopting $RV5^{HL}$ and $TSRV5^{HL}$, the result is statistically significant for the MT specification only. Furthermore, we found that the 95% HPD intervals computed from the MT specification are slightly overestimated. Thus, the TOPIX data provide significant evidence in support of some power transformations when the 92.5% and 95% HPD intervals are considered.

In addition, we observe that the HPD interval of $\lambda$ is relatively narrow (width less than 0.15) with low variability, which is similar to results obtained by Tsiotas (2009). The least and largest variations are provided by the ET and MT specifications, respectively. In particular, the variance value of $\lambda$ in the 2004–2011
Table 6. Summary of the posterior sample of parameters in the R-NASV models with generalized Student’s t-distributions adopting RV1^{HL} 2004–2011 data.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>$\lambda$</th>
<th>$\mu$</th>
<th>$\nu$</th>
<th>$\xi_0$</th>
<th>$\xi_1$</th>
<th>$\sigma_{y}$</th>
<th>$\alpha$</th>
<th>$\phi$</th>
<th>$\sigma_{h}$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-ETASV-NCT model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.016</td>
<td>0.026</td>
<td>28.97</td>
<td>0.161</td>
<td>0.898</td>
<td>0.302</td>
<td>0.046</td>
<td>0.947</td>
<td>0.233</td>
<td>−0.365</td>
</tr>
<tr>
<td>SD</td>
<td>0.008</td>
<td>0.023</td>
<td>4.71</td>
<td>0.034</td>
<td>0.037</td>
<td>0.008</td>
<td>0.121</td>
<td>0.008</td>
<td>0.014</td>
<td>0.040</td>
</tr>
<tr>
<td>95% LB</td>
<td>−0.001</td>
<td>−0.016</td>
<td>19.81</td>
<td>0.098</td>
<td>0.827</td>
<td>0.287</td>
<td>−0.189</td>
<td>0.931</td>
<td>0.206</td>
<td>−0.446</td>
</tr>
<tr>
<td>95% UB</td>
<td>0.001</td>
<td>−0.014</td>
<td>20.70</td>
<td>0.102</td>
<td>0.835</td>
<td>0.288</td>
<td>−0.173</td>
<td>0.932</td>
<td>0.208</td>
<td>−0.440</td>
</tr>
<tr>
<td>IF</td>
<td>1.37</td>
<td>1.4</td>
<td>94.8</td>
<td>66.8</td>
<td>65.4</td>
<td>13.0</td>
<td>3.9</td>
<td>6.6</td>
<td>59.3</td>
<td>13.5</td>
</tr>
<tr>
<td>R-MTASV-NCT model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.054</td>
<td>0.027</td>
<td>28.60</td>
<td>0.169</td>
<td>0.893</td>
<td>0.302</td>
<td>0.083</td>
<td>0.927</td>
<td>0.235</td>
<td>−0.365</td>
</tr>
<tr>
<td>SD</td>
<td>0.028</td>
<td>0.022</td>
<td>5.18</td>
<td>0.030</td>
<td>0.034</td>
<td>0.007</td>
<td>0.086</td>
<td>0.016</td>
<td>0.014</td>
<td>0.040</td>
</tr>
<tr>
<td>95% LB</td>
<td>0.998</td>
<td>−0.017</td>
<td>20.20</td>
<td>0.103</td>
<td>0.831</td>
<td>0.288</td>
<td>−0.081</td>
<td>0.897</td>
<td>0.208</td>
<td>−0.443</td>
</tr>
<tr>
<td>95% UB</td>
<td>1.005</td>
<td>−0.015</td>
<td>20.65</td>
<td>0.106</td>
<td>0.834</td>
<td>0.289</td>
<td>−0.064</td>
<td>0.901</td>
<td>0.209</td>
<td>−0.436</td>
</tr>
<tr>
<td>IF</td>
<td>1.102</td>
<td>0.065</td>
<td>38.07</td>
<td>0.218</td>
<td>0.953</td>
<td>0.317</td>
<td>0.249</td>
<td>0.958</td>
<td>0.258</td>
<td>−0.297</td>
</tr>
<tr>
<td>95% UB</td>
<td>1.104</td>
<td>0.072</td>
<td>39.58</td>
<td>0.225</td>
<td>0.964</td>
<td>0.319</td>
<td>0.267</td>
<td>0.960</td>
<td>0.262</td>
<td>−0.290</td>
</tr>
<tr>
<td>R-ETASV-SKT model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.016</td>
<td>−0.513</td>
<td>29.34</td>
<td>0.185</td>
<td>0.907</td>
<td>0.304</td>
<td>0.059</td>
<td>0.948</td>
<td>0.230</td>
<td>−0.383</td>
</tr>
<tr>
<td>SD</td>
<td>0.008</td>
<td>0.321</td>
<td>4.78</td>
<td>0.036</td>
<td>0.037</td>
<td>0.008</td>
<td>0.117</td>
<td>0.007</td>
<td>0.013</td>
<td>0.041</td>
</tr>
<tr>
<td>95% LB</td>
<td>−0.001</td>
<td>−1.149</td>
<td>20.00</td>
<td>0.117</td>
<td>0.833</td>
<td>0.288</td>
<td>−0.159</td>
<td>0.934</td>
<td>0.202</td>
<td>−0.464</td>
</tr>
<tr>
<td>95% UB</td>
<td>0.000</td>
<td>−1.052</td>
<td>20.82</td>
<td>0.119</td>
<td>0.843</td>
<td>0.290</td>
<td>−0.150</td>
<td>0.935</td>
<td>0.205</td>
<td>−0.458</td>
</tr>
<tr>
<td>IF</td>
<td>1.112</td>
<td>0.113</td>
<td>38.36</td>
<td>0.256</td>
<td>0.987</td>
<td>0.320</td>
<td>0.267</td>
<td>0.959</td>
<td>0.255</td>
<td>−0.309</td>
</tr>
<tr>
<td>95% UB</td>
<td>1.111</td>
<td>0.038</td>
<td>40.38</td>
<td>0.253</td>
<td>0.979</td>
<td>0.319</td>
<td>0.300</td>
<td>0.964</td>
<td>0.256</td>
<td>−0.305</td>
</tr>
<tr>
<td>R-MTASV-SKT model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.056</td>
<td>−0.474</td>
<td>29.00</td>
<td>0.177</td>
<td>0.910</td>
<td>0.304</td>
<td>0.102</td>
<td>0.927</td>
<td>0.228</td>
<td>−0.389</td>
</tr>
<tr>
<td>SD</td>
<td>0.027</td>
<td>0.283</td>
<td>4.70</td>
<td>0.035</td>
<td>0.035</td>
<td>0.007</td>
<td>0.082</td>
<td>0.016</td>
<td>0.014</td>
<td>0.041</td>
</tr>
<tr>
<td>95% LB</td>
<td>1.105</td>
<td>−1.900</td>
<td>20.50</td>
<td>0.106</td>
<td>0.845</td>
<td>0.289</td>
<td>−0.057</td>
<td>0.897</td>
<td>0.201</td>
<td>−0.469</td>
</tr>
<tr>
<td>95% UB</td>
<td>1.111</td>
<td>0.113</td>
<td>38.36</td>
<td>0.256</td>
<td>0.987</td>
<td>0.320</td>
<td>0.267</td>
<td>0.959</td>
<td>0.255</td>
<td>−0.309</td>
</tr>
<tr>
<td>IF</td>
<td>13.0</td>
<td>37.6</td>
<td>102.9</td>
<td>54.9</td>
<td>49.6</td>
<td>9.3</td>
<td>10.9</td>
<td>12.6</td>
<td>48.4</td>
<td>15.3</td>
</tr>
<tr>
<td>R-YJASV-LSKT model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.026</td>
<td>−0.486</td>
<td>29.86</td>
<td>0.178</td>
<td>0.901</td>
<td>0.304</td>
<td>0.053</td>
<td>0.949</td>
<td>0.231</td>
<td>−0.383</td>
</tr>
<tr>
<td>SD</td>
<td>0.014</td>
<td>0.303</td>
<td>4.91</td>
<td>0.035</td>
<td>0.035</td>
<td>0.007</td>
<td>0.123</td>
<td>0.007</td>
<td>0.013</td>
<td>0.041</td>
</tr>
<tr>
<td>95% LB</td>
<td>0.998</td>
<td>−1.103</td>
<td>20.31</td>
<td>0.108</td>
<td>0.832</td>
<td>0.289</td>
<td>−0.184</td>
<td>0.933</td>
<td>0.206</td>
<td>−0.464</td>
</tr>
<tr>
<td>95% UB</td>
<td>1.000</td>
<td>−1.022</td>
<td>21.57</td>
<td>0.115</td>
<td>0.839</td>
<td>0.290</td>
<td>−0.162</td>
<td>0.935</td>
<td>0.208</td>
<td>−0.463</td>
</tr>
<tr>
<td>IF</td>
<td>1.054</td>
<td>0.091</td>
<td>39.48</td>
<td>0.248</td>
<td>0.970</td>
<td>0.320</td>
<td>0.304</td>
<td>0.964</td>
<td>0.258</td>
<td>−0.300</td>
</tr>
<tr>
<td>95% UB</td>
<td>1.054</td>
<td>0.091</td>
<td>39.48</td>
<td>0.248</td>
<td>0.970</td>
<td>0.320</td>
<td>0.304</td>
<td>0.964</td>
<td>0.258</td>
<td>−0.300</td>
</tr>
</tbody>
</table>
data is smaller than the corresponding value in the 2004–2007 data.

Regarding the non-linear bias correction parameter of the RV equation, $\xi_1$ deviated from the assumption of Takahashi et al. (2009) that $\xi_1 = 1$ when applying the RV1$^{HL}$ and BV1$^{HL}$ estimators. Even the upper limit of the 98% HPD interval of $\xi_1$ was less than one (result not shown). We note that the deviation from 1 tends to be larger as the variance decreases from approximately 0.12. So the proposed models provide a log RV persistence of less than one using the RV estimators based on a sampling at very high frequency. This finding is consistent with the empirical evidence found by Hansen et al. (2011) and Takahashi et al. (2014) when applying a realized kernel estimator.

We next consider the posterior evidence regarding parameters of generalized Student’s $t$-error distributions. Deviation of returns from the normality assumption is expressed by the $\nu$, $\mu$, and $\beta$ parameters. The posterior means of the degrees of freedom $\nu$ are considerably higher than 8 (between 23 and 26 for the 2004–2007 data and between 27 and 31 for the 2004–2011 data), indicating skewness and kurtosis (see Aas and Haff (2006) for explanation). In the NCT specification, the results show that the 2004–2007 data favor the NCT distribution since the 95% HPD interval of $\mu$ excludes 0. We even found that the posterior probability that $\mu$ is positive is greater than 0.98 (result not shown). Meanwhile, although the 90% HPD intervals of $\mu$ in the model with NCT specification over the time period 2004–2011 include 0 (result not shown), their posterior distributions are largely positive as shown in Fig. 7 (first row) for the model with the MT specification. Considering the SKT specification, the measure of skewness expressed by $\beta$ shows that all data do not favor the SKT specification since the HPD interval includes 0 when the 90% HPD interval of $\beta$ is considered (result not shown). The only exception is the model adopting RV1$^{HL}$ over the time pe-

![Figure 7](image_url)

Figure 7. Histograms of the posterior distribution of parameter $\mu$ in the R-MTASV-NCT model (first row) and skewness parameter $\beta$ in the R-MTASV-SKT model (second row) adopting the 2004–2011 data.
period 2004–2011. In contrast to the empirical results from Takahashi et al. (2014), our returns data provide strong evidence in support of skewness over both time periods, where the deviation of $\beta$ from zero is relatively large. In fact, posterior distributions of $\beta$ in the SKT specification are largely negative as shown in Fig. 7 (second row) for the model with the MT specification. Those results present evidence of generalized Student’s $t$-error distributions in both data sets using all four RVs.

Finally, we observe the performance of persistence $\phi$. Since the assumption of stationary has been made in the prior selection of $\phi$, its posterior mean as well as its confidence interval deviate from the non-stationary assumption. Furthermore, we found that the highest persistence in the non-linear volatility process is provided by the MT specification in the 2004–2007 data and by the YJ specification in the 2004–2011 data. Compared with the linear version, our results show that the non-linear volatility process with ET and YJ specifications in each period and with the MT specification in the 2004–2011 data is less persistent.

5.4. Model selection

This section investigates whether the data better support the linear or non-linear volatility processes to improve the model fit. To determine the best transformation, we compute marginal likelihoods using the Gelfand and Dey (1994) and Geweke (1999) numerical procedures, as explained in Section 4. Tables 7 and 8 present the values of the Bayes factor for the R-NASV models against the R-ASV models and the models with SKT distribution against competing models.

From Table 7, we found that the log Bayes factors for the R-NASV models against the R-ASV models are greater than 3 in all cases, indicating that the R-NASV models provide the best fit. In fact, the log Bayes factors provide strong

<table>
<thead>
<tr>
<th>RV</th>
<th>Transformation</th>
<th>R-ANSV vs R-ASV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SKT</td>
<td>NCT</td>
</tr>
<tr>
<td>RV1</td>
<td>ET</td>
<td>7.78</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>YJ</td>
<td>10.12</td>
</tr>
<tr>
<td>RV5</td>
<td>ET</td>
<td>13.65</td>
</tr>
<tr>
<td></td>
<td>YJ</td>
<td>11.64</td>
</tr>
<tr>
<td>BV1</td>
<td>ET</td>
<td>22.88</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>19.96</td>
</tr>
<tr>
<td></td>
<td>YJ</td>
<td>38.47</td>
</tr>
<tr>
<td>TSRV5</td>
<td>ET</td>
<td>11.76</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>3.92</td>
</tr>
<tr>
<td></td>
<td>YJ</td>
<td>11.98</td>
</tr>
</tbody>
</table>
and very strong evidence in support of power transformations of the lagged log volatility rather than no transformation in the R-ASV models.

When the three competing R-NASV models are compared over the time period 2004–2007, there is no clear pattern to determine which transformation is the best in all returns distribution specifications (indicated by the largest value of log Bayes factor for any RV data sets) or in all RV data sets (indicated by the largest value of log Bayes factor for any returns distributions). We note that the R-ETASV and R-YJASV models are very competitive and outperform the R-MTASV model. The YJ and ET specifications are suggested in the models adopting RV1^{HL} and RV5^{HL}, respectively. In both models adopting BV1^{HL} and TSRV5^{HL}, the YJ and ET specifications rank first for the model accommodating generalized Student’s t-distributions and other distributions, respectively. Over the 2004–2011 data having a very high kurtosis, the R-MTASV models provide the best fit for any returns distributions and RV data sets. Those results are consistent with previously reported results in terms of a credible interval.

Comparing the results of the model using the four different returns distributions in any transformation specification, see Table 8, indicates that the R-NSV model with the SKT distribution specification for returns error is the most favored in each period. The log Bayes factors in favor of this specification compared with the second best fitting specification are greater than 11, which is very strong margin. Therefore, the following discussion is focused on the models with the SKT distribution.

Table 8. Logarithmic Bayes factors of the models with SKT distribution against competing models on the same transformation for lag volatility evaluated in the TOPIX data set.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NCT</td>
<td>T</td>
<td>normal</td>
</tr>
<tr>
<td>RV1</td>
<td>no</td>
<td>26.56</td>
<td>32.09</td>
<td>39.86</td>
</tr>
<tr>
<td></td>
<td>ET</td>
<td>23.40</td>
<td>32.99</td>
<td>42.14</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>26.04</td>
<td>30.33</td>
<td>38.63</td>
</tr>
<tr>
<td></td>
<td>YJ</td>
<td>24.86</td>
<td>34.49</td>
<td>41.85</td>
</tr>
<tr>
<td>RV5</td>
<td>no</td>
<td>14.66</td>
<td>20.37</td>
<td>30.22</td>
</tr>
<tr>
<td></td>
<td>ET</td>
<td>17.34</td>
<td>21.72</td>
<td>31.32</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>21.40</td>
<td>26.96</td>
<td>35.20</td>
</tr>
<tr>
<td></td>
<td>YJ</td>
<td>18.27</td>
<td>22.78</td>
<td>32.04</td>
</tr>
<tr>
<td>BV1</td>
<td>no</td>
<td>13.48</td>
<td>23.33</td>
<td>41.25</td>
</tr>
<tr>
<td></td>
<td>ET</td>
<td>15.21</td>
<td>22.38</td>
<td>30.38</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>17.62</td>
<td>23.48</td>
<td>35.95</td>
</tr>
<tr>
<td></td>
<td>YJ</td>
<td>29.49</td>
<td>41.30</td>
<td>52.45</td>
</tr>
<tr>
<td>TSRV5</td>
<td>no</td>
<td>19.11</td>
<td>33.37</td>
<td>44.47</td>
</tr>
<tr>
<td></td>
<td>ET</td>
<td>25.57</td>
<td>29.54</td>
<td>37.14</td>
</tr>
<tr>
<td></td>
<td>MT</td>
<td>18.08</td>
<td>30.83</td>
<td>34.32</td>
</tr>
<tr>
<td></td>
<td>YJ</td>
<td>32.04</td>
<td>52.45</td>
<td>38.06</td>
</tr>
</tbody>
</table>
For the 2004–2007 data, the R-YJASV-SKT and R-ETASV-SKT models, respectively, provide the first and second best fit to the RV1HL, BV1HL, and TSRV5HL data. Using the RV5HL data, the ranking between the first and second best is reversed with the R-ETASV-SKT model becoming first. Clearly, with regard to its log Bayes factors, the R-YJASV-SKT model is very strongly favored for the BV1HL data, positively favored for the RV1HL data, and fairly insignificant for the TSRV5HL data. With the RV5HL data, the R-ETASV-SKT model against the R-YJASV-SKT model is positively favored. Furthermore, the evidence in favor of the poorest fitting model of the R-NASV-SKT models, the one having the MT specification, compares strongly with the R-ASV-SKT model in the RV1HL and TSRV5HL cases and is also very strong in the RV5HL and BV1HL cases.

For the 2004–2011 data, compared with the second best models, the R-MTASV-SKT model is strongly for the RV1HL and BV1HL data and very strongly favored for the RV5HL and TSRV5HL data. The R-ETASV and R-YJASV models are very competitive. Furthermore, the poorest fitting model of the R-NASV-SKT models compared with the R-ASV-SKT model is strongly favored in the RV1HL case and very strongly favored in the others.

### 5.5. Sensitivity of priors

Concerning the sensitivity of MCMC output to prior choices, let us recall from the previous analysis that the R-MTASV-SKT model is the best performing model for the 2004–2011 data. This section applies a sensitivity test for this model on the power parameter $\lambda$. Our intention is to show that posterior sim-

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Prior</th>
<th>Prior</th>
<th>Prior</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mathcal{N}(0,1)$</td>
<td>$\mathcal{N}(0,10)$</td>
<td>$\mathcal{U}(-10,10)$</td>
<td>$\mathcal{U}(-100,100)$</td>
</tr>
<tr>
<td>model adopting RV1HL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.056 (0.027)</td>
<td>1.055 (0.025)</td>
<td>1.056 (0.028)</td>
<td>1.055 (0.027)</td>
</tr>
<tr>
<td>90% HPD</td>
<td>[1.012,1.102]</td>
<td>[1.012,1.105]</td>
<td>[1.009,1.102]</td>
<td>[1.009,1.100]</td>
</tr>
<tr>
<td>IF</td>
<td>12.8</td>
<td>13.5</td>
<td>13.5</td>
<td>12.3</td>
</tr>
<tr>
<td>model adopting RV5HL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.055 (0.030)</td>
<td>1.054 (0.030)</td>
<td>1.057 (0.030)</td>
<td>1.055 (0.030)</td>
</tr>
<tr>
<td>90% HPD</td>
<td>[1.006,1.107]</td>
<td>[1.003,1.103]</td>
<td>[1.007,1.106]</td>
<td>[1.004,1.102]</td>
</tr>
<tr>
<td>IF</td>
<td>18.2</td>
<td>15.4</td>
<td>15.4</td>
<td>15.6</td>
</tr>
<tr>
<td>model adopting BV1HL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.044 (0.028)</td>
<td>1.045 (0.029)</td>
<td>1.044 (0.029)</td>
<td>1.046 (0.028)</td>
</tr>
<tr>
<td>90% HPD</td>
<td>[0.999,1.095]</td>
<td>[0.995,1.090]</td>
<td>[0.997,1.092]</td>
<td>[0.998,1.091]</td>
</tr>
<tr>
<td>IF</td>
<td>14.1</td>
<td>13.0</td>
<td>13.0</td>
<td>13.3</td>
</tr>
<tr>
<td>model adopting TSRV5HL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>1.061 (0.030)</td>
<td>1.058 (0.030)</td>
<td>1.059 (0.029)</td>
<td>1.059 (0.030)</td>
</tr>
<tr>
<td>90% HPD</td>
<td>[1.011,1.109]</td>
<td>[1.007,1.107]</td>
<td>[1.008,1.107]</td>
<td>[1.010,1.109]</td>
</tr>
<tr>
<td>IF</td>
<td>14.4</td>
<td>15.6</td>
<td>15.6</td>
<td>13.9</td>
</tr>
</tbody>
</table>
Figure 8. The posterior distributions for the $\lambda$ parameter using various prior assumptions. These are obtained from the last 10,000 iterations of the R-MTASV-SKT model adopting RV1HL data (2004–2011).

These estimations give similar characteristics in terms of the mean, standard deviation, credible interval, and sampling efficiency estimates under alternative priors of $\lambda$.

In the previous estimation, we used a normal prior for $\lambda$ with variance 1. To study the sensitivity of the estimation results to the diffuse prior on $\lambda$, we first choose the same normal prior but we increase the variance to a value of 10. We then choose the uniform priors on $[-10, 10]$ and $[-100, 100]$. Notice that the standard deviation for the last two consecutive priors are 33.33 and 833.33. The estimates for all parameters are found to be almost the same under all priors.

In particular, the parameter estimates and sampling efficiency for $\lambda$ are reported in Table 9. It shows that the estimates for all parameters in the model are not affected by changing the prior for $\lambda$. To graphically illustrate our results, Fig. 8 displays the sample paths and posterior distributions for the $\lambda$ using all four prior assumptions.

6. Conclusions and extensions

This study proposed a class of non-linear RSV models with asymmetric effect and generalized Student’s $t$-error distributions by applying three families of power transformation—exponential, modulus, and Yeo-Johnson—to lagged log volatility. We developed an efficient MCMC algorithm including Gibbs, HMC, and RMHMC steps for sampling from the posterior distribution of the models. Empirical results using TOPIX data show that the posterior mean accepts all proposed models and the HPD intervals only accept some of the proposed models.

Importantly, the RV data sampled at very high frequency (say 1-minute) strongly supports that the non-linear bias correction of the log RV may deviate from 1 (unlike the assumption of Takahashi et al. (2009)) in terms of HPD inter-
val although Takahashi et al. (2014) showed that this additional bias correction does not improve the model fit. The performance of competing models was quantified by the logarithm of estimated marginal likelihood. The estimation results demonstrated that the model with the SKT distribution best fitted the TOPIX data, although the skewness parameter in some of the models is not fully guaranteed by the 90% HPD interval. Furthermore, the marginal likelihood and Bayes factor criterion indicate that the proposed R-NASV model outperforms the RSV model, where the three competing R-NASV models are very competitive.

The proposed R-NASV models could also be extended by considering a class of non-linear transformations for RV. Gonçalves and Meddahi (2011) particularly showed that the log transformation of RV can be improved upon by choosing values for the BC parameter other than zero.

Acknowledgements

The authors wish to thank Dr. Shuichi Nagata for some Matlab codes and helpful discussions.

References


REALIZED NON-LINEAR STOCHASTIC VOLATILITY MODELS


