



Bayesian Inference for Mixed Gaussian GARCH-Type Model by Hamiltonian Monte Carlo Algorithm

Rubing Liang¹ · Binbin Qin¹ · Qiang Xia¹

Accepted: 30 September 2022 / Published online: 30 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

MCMC algorithm is widely used in parameters' estimation of GARCH-type models. However, the existing algorithms are either not easy to implement or not fast to run. In this paper, Hamiltonian Monte Carlo (HMC) algorithm, which is easy to perform and also efficient to draw samples from posterior distributions, is firstly proposed to estimate for the Gaussian mixed GARCH-type models. And then, based on the estimation of HMC algorithm, the forecasting of volatility prediction is investigated. Through the simulation experiments, the HMC algorithm is more efficient and flexible than the Griddy-Gibbs sampler, and the credibility interval of forecasting for volatility prediction is also more accurate. A real application is given to support the usefulness of the proposed HMC algorithm well.

Keywords Mixed Gaussian · GARCH-type models · HMC algorithm · Forecasting · Bayesian inference

1 Introduction

Financial time series, such as exchange rates and stock returns, often have exhibited time-varying volatility, excess kurtosis and volatility clustering reported by Mandelbrot (1963) and Fama (1965). The family of autoregressive conditional heteroskedastic (ARCH) model of Engle (1982) and the generalized ARCH (GARCH) model of Bollerslev (1986) provide effective techniques to fit the volatility of the

✉ Qiang Xia
xiaqiang@scau.edu.cn

Rubing Liang
liang_ru_bing@163.com

Binbin Qin
3252014058@qq.com

¹ College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, People's Republic of China

financial time series. Since then, huge research works for the GARCH model and its extensions can be found in the literature for the past three decades, such as the GJR-GARCH model of Glosten et al. (1993), the threshold GARCH (TGARCH) model of Zakoian (1994), the exponential GARCH (EGARCH) model of Nelson (1991), the integrated GARCH (IGARCH) model of Engle and Bollerslev (1986), the power-transformed and threshold GARCH model (PTTGARCH) of Pan et al. (2008). For the GARCH type models, Westerfield (1977) and McFarland et al. (1982) have shown that the assumption of the GARCH model with normal errors can not provide an appropriate framework for some return series with excessive kurtosis and volatility clustering. Bollerslev (1987) suggested that the GARCH models with t -student innovations should be considered to describe the conditional distributions of the stock returns. However, these models can not capture volatility clustering, high kurtosis, heavy-tailed distributions and the phenomenon of extreme events. A mixture normal GARCH-type model with zero mean and different variances is generated by a normal density with a small variance, while a small number of innovations are generated by a normal density with a large variance. Therefore it becomes very popular to model the financial return data, and provide a better fitting and forecasting than the GARCH-type models with normal or t -student innovations. See Bauwens et al. (1999), McLachlan and Peel (2000), Bai et al. (2003), Wong and Li (2001), Haas et al. (2004), Zhang et al. (2006) and Alexander and Lazar (2006) among others.

In the literature, maximum likelihood, quasi-maximum likelihood, the generalized method of moments and the least absolute deviations approach are traditionally carried out to infer the GARCH-type models, see Bollerslev and Wooldridge (1992) and Pan et al. (2008) for details. It is well known that the Bayesian inference offers a natural way to overcome computing problems and to avoid analytical difficulties in the estimation of volatilities. Some authors have applied Markov chain Monte Carlo (MCMC) algorithm to approximate the posterior distributions of the parameters for the GARCH-type models. For example, Geweke (1994) suggested the importance sampling to provide an efficient and generic method for updating posterior distributions. The Griddy-Gibbs sampler suggested by Ritter and Tanner (1992) has been used by Bauwens and Lubrano (1998) and Xia et al. (2017) for a GARCH-type model with normal or t -distributed errors, and Ausín and Galeano (2007) for a GARCH model with Gaussian mixture errors. In fact, finding an appropriate proposal distribution in the Metropolis-Hastings algorithm suggested by Metropolis et al. (1953) and Hastings (1970) or an importance function is not easy. Although the Griddy-Gibbs sampler is easier to implement than other methods, it takes much computation time. Therefore, designing an algorithm that is easier to implement and less heavy to run, is very important for inferring the mixed normal distribution GARCH-type models in practise.

In recent years, attention has been paid for Hamiltonian Monte Carlo (HMC) algorithm proposed by Neal (2011) in the literature, because people realized that it can search the typical set of parameters effectively by using the gradient information of the target distribution. Different from the Metropolis-Hastings algorithm, the HMC sampler may not lead to random walk. Hence, the HMC algorithm has been introduced for inferring the time series models. For instance,

Paixão and Ehlers (2017) used the HMC algorithm to estimate the GJR-GARCH model proposed by Glosten et al. (1993) with normal and *t*-Student errors. Burda and Bélisle (2019) applied HMC algorithm to overcome the difficulty of inferring the Copula-GARCH model, of which the distribution of parameters is skewness, asymmetry and truncation. Kreuzer and Czado (2021) proposed Bayesian inference for a single factor copula stochastic volatility model. Their related results show that the HMC algorithm can be implemented more easily in practice. Stan suggested by Carpenter et al. (2016) can provide HMC and No-U-Turn [NUTS, Hoffman and Gelman (2014)] methods to carry out Bayesian inference, but it aim at the models with continuous-variable. In the GARCH-type models with mixed normal errors, the latent variable is regarded as discrete parameter to define the likelihood function, thus we can not implement HMC procedure directly by Stan. Therefore, the main objective of this article is to propose a procedure for Bayesian inference and prediction of the general GARCH-type models with the Gaussian mixture innovations based on the HMC algorithm.

The arrangements of this paper are as follows. Section 2 presents the general GARCH-type models with mixed normal innovations and describes a Bayesian inference of the model. A Hamiltonian Monte Carlo algorithm for sampling the posterior density of volatilities and VaR forecasts is also addressed. Section 3 provides some simulation experiments and a real data application, which illustrate the accuracy in the estimation of the parameters and the prediction of volatilities and VaR. Section 4 is our conclusions.

2 Gaussian Mixture GARCH-Type Models and Bayesian Inference

2.1 GARCH-type Models with Mixed Gaussian Innovations

A series $\{y_t\}$ is said to follow the normal mixture GARCH-type models given by

$$\begin{aligned} y_t &= \sqrt{h_t} \epsilon_t, \\ h_t &= \text{var}(y_t | F_{t-1}) = f(\Theta, F_{t-1}), \end{aligned} \tag{2.1}$$

where $f(\Theta, F_{t-1})$ is a differentiable function of Θ and given the previous information $F_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$. The error term ϵ_t is taken from independent mixed normal distributions as follows,

$$\varphi(\epsilon_t) = \begin{cases} \varphi_1(\epsilon_t), & \text{with probability } \rho, \\ \varphi_2(\epsilon_t), & \text{with probability } 1 - \rho, \end{cases} \tag{2.2}$$

where $\varphi_1(\epsilon_t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\epsilon_t^2}{2\sigma^2})$, $\varphi_2(\epsilon_t) = \frac{\sqrt{\lambda}}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\lambda\epsilon_t^2}{2\sigma^2})$ and $\sigma^2 = \frac{\lambda}{1+(\lambda-1)\rho}$ with $0 < \lambda < 1$. Then $E(\epsilon_t) = 0$, $\text{var}(\epsilon_t) = 1$.

As we can see, model (2.1) includes the following conditional heteroscedasticity models:

- (1) If $f(\Theta, F_{t-1}) = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}$, then (2.1) becomes the standard GARCH model [Bollerslev (1986)];
- (2) If $f(\Theta, F_{t-1}) = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i}^2$, then (2.1) is the ARCH model [Engle (1982)], which is a special case of GARCH;
- (3) If $f(\Theta, F_{t-1}) = \alpha_0 + \sum_{i=1}^p (\alpha_i + \gamma_i N_{t-i}) y_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}$, where $\gamma_i > 0, i = 1, \dots, p$ and $N_{t-i} = 1$ if $y_{t-i} < 0$ or $N_{t-i} = 0$ if $y_{t-i} \geq 0$, then (2.1) is the GJR-GARCH model [Glosten et al. (1993)];
- (4) If $f(\Theta, F_{t-1}) = \alpha_0 + \beta_1 h_{t-1} + (1 - \beta_1) y_{t-1}^2$, then (2.1) is the IGARCH(1,1) model [Engle and Bollerslev (1986)];
- (5) If $\ln f(\Theta, F_{t-1}) = \alpha_0 + \sum_{i=1}^p \alpha_i g(\epsilon_{t-i}) + \sum_{j=1}^q \beta_j \ln h_{t-j}$, then (2.1) becomes the EGARCH model [Nelson (1991)], where $\alpha_1 = 1, g(\epsilon_t) = \theta \epsilon_t + \gamma [|\epsilon_t| - E(|\epsilon_t|)]$.
- (6) If $f(\Theta, F_{t-1}) = \left[\alpha_0 + \sum_{i=1}^p \alpha_{1i} (y_{t-i}^+)^{2\delta} + \sum_{i=1}^p \alpha_{2i} (y_{t-i}^-)^{2\delta} + \sum_{j=1}^q \beta_j h_{t-j}^\delta \right]^{1/\delta}$, (2.1) represents the PTTGARCH model [Pan et al. (2008)], where δ is a known positive number.

Remark 2.1 Diebolt and Robert (1994) defined the latent variables $\mathbf{z} = \{z_t, 1 \leq t \leq n\}$, where z_t is a Bernoulli random variable with probability ρ , and $I_{\{\cdot\}}$ is the indicator function. Then (2.2) can be interpreted as

$$\epsilon_t = \epsilon_t^{(1)} I_{\{z_t=1\}} + \epsilon_t^{(2)} I_{\{z_t=0\}}, \tag{2.3}$$

$$\epsilon_t^{(1)} \sim N(0, \sigma^2), \epsilon_t^{(2)} \sim N(0, \sigma^2/\lambda). \tag{2.4}$$

Thus, if z_t can be identified, then the distribution of ϵ_t can be determined.

2.2 The Posterior Distribution

Let Θ denote the parameter vector of the Gaussian mixed GARCH-type models (2.1), and $\mathbf{y} = \{y_t, t = 1, 2, \dots, n\}$ is observed series with n sample size. Then the posterior density can be written as

$$p(\Theta|\mathbf{y}) \propto \pi(\Theta)l(\Theta;\mathbf{y}, \mathbf{z}), \tag{2.5}$$

where $\pi(\Theta)$ is the prior and $l(\Theta;\mathbf{y}, \mathbf{z})$ is the likelihood function, which can be derived in terms of latent variables $\mathbf{z} = \{z_t, t = 1, 2, \dots, n\}$, i.e.,

$$l(\Theta;\mathbf{y}, \mathbf{z}) \propto \prod_{t:z_t=1} \rho \varphi_1 \left[\frac{y_t}{f(\Theta, F_{t-1})} \right] \prod_{t:z_t=0} (1 - \rho) \varphi_2 \left[\frac{y_t}{f(\Theta, F_{t-1})} \right]. \tag{2.6}$$

The parameter ρ controls the proportion of the two normal distributions. When ρ goes to 1, the innovations ϵ_t will almost come from the first normal distribution; when ρ tends to 0, ϵ_t will almost come from the second normal distribution. It would

be assumed that $0.5 < \rho < 1$, so that most of the innovation ϵ_t comes from the first part.

To implement the Bayesian inference about the parameters Θ in model (2.1), we need the joint posterior distribution $P(\Theta|\mathbf{y})$, which can be obtained by using the conditional posterior distribution in a HMC process. Therefore, we need to choose priors to derive the conditional posterior distribution for the unknown parameters.

For any priors $\pi(\Theta)$, as h_t is a function of Θ in (2.1), the conditional posterior densities of most parameters in Θ will contain h_t in (2.5). Consequently, they cannot be a normal or any other well known density, from which random numbers could be easily generated. Also, they are less likely to have the property of conjugacy. Therefore, the uninformative prior distributions will be preferred, and the relevant posterior distribution will be obtained.

If model (2.1) is the standard GARCH model, the uniform prior of $\Theta = (\alpha_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q, \rho, \lambda)'$ can be chosen as follows:

$$\begin{aligned} \alpha_0 &\sim U(0, +\infty); & \alpha_i &\sim U(0, 1), i = 1, \dots, p; \\ \beta_j &\sim U(0, 1), j = 1, \dots, q; & \rho &\sim U(0.5, 1); & \lambda &\sim U(0, 1). \end{aligned} \tag{2.7}$$

Therefore, according to the uninformative prior distribution, the joint posterior distribution function of Θ is,

$$p(\Theta|\mathbf{y}, \mathbf{z}) \propto \prod_{t:z_t=1} \rho \frac{1}{\sqrt{\sigma^2 h_t}} \exp \left[\frac{-y_t^2}{2\sigma^2 h_t} \right] \prod_{t:z_t=0} (1 - \rho) \sqrt{\frac{\lambda}{\sigma^2 h_t}} \exp \left[\frac{-\lambda y_t^2}{2\sigma^2 h_t} \right]. \tag{2.8}$$

2.3 Sampling Scheme Using HMC

Since Neal (2011) introduced the HMC algorithm, which originated from the algorithm when Duane et al. (1987) studied molecular dynamics simulation, the statistical inference method based on Hamiltonian dynamics became popular. According to Betancourt (2017), Girolami and Calderhead (2011) and Neal (2011), the HMC algorithm has an excellent performance in solving some difficult high-dimensional inference problems and pathological behaviors of distribution functions.

In Hamiltonian dynamics, there are two parameters with the same dimension, the *position* vector Θ and the *momentum* vector Φ , to describe the motion process. The system is described by a function $H(\Theta, \Phi)$ defined on the phase space (Θ, Φ) . At the point (Θ, Φ) , $H(\Theta, \Phi)$ is known as the Hamiltonian, and can be decomposed into two parts, i.e.,

$$H(\Theta, \Phi) = U(\Theta) + K(\Phi), \tag{2.9}$$

where $U(\Theta)$ and $K(\Phi)$ are called the potential energy and the kinetic energy, respectively. In particular, the canonical distribution of $H(\Theta, \Phi)$ has the form $\pi(\Theta, \Phi) = \exp(-H(\Theta, \Phi))$. From Hamilton's equations (2.10), we can know how Θ and Φ change over time t ,

$$\begin{aligned}\frac{d\Theta_i}{dt} &= + \frac{\partial H}{\partial \Phi_i}, \\ \frac{d\Phi_i}{dt} &= - \frac{\partial H}{\partial \Theta_i}.\end{aligned}\tag{2.10}$$

In a non-physical context, the *position* parameter corresponds to the parameter of interest with the density $\pi(\Theta)$, and the parameter *momentum* is assumed to be a normal distribution random vector with the density $\pi(\Phi)$ independent of the parameter *position*. The joint probability distribution is written as the product of two densities,

$$\pi(\Theta, \Phi) = \pi(\Theta)\pi(\Phi).\tag{2.11}$$

Since the joint distribution is regarded as a canonical distribution, then the Hamiltonian function can be written as follows,

$$H(\Theta, \Phi) = -\ln \pi(\Theta)\pi(\Phi) = -\ln \pi(\Theta) - \ln \pi(\Phi) = U(\Theta) + K(\Phi).\tag{2.12}$$

In the mixed normal GARCH-type models, the parameter Θ corresponds to the *position* in Hamiltonian function $H(\Theta, \Phi)$, and the *momentum* variable is denoted as Φ with the same dimension as Θ . In order to calculate the Hamiltonian $H(\Theta, \Phi)$ of the Gaussian mixed GARCH type models, we adopt Φ follows a normal distribution with mean zero and covariance Σ , then the Hamiltonian function can be further written as,

$$H(\Theta, \Phi) = U(\Theta) + K(\Phi) = -\ln p(\Theta|\mathbf{y}, \mathbf{z}) + \frac{1}{2}\Phi'\Sigma^{-1}\Phi.$$

The normalising constant omitted in the HMC iteration will not affect the program. Using the above Hamiltonian function, we perform the HMC algorithm to update the parameter Θ .

Each iteration of the HMC algorithm has two steps. The first changes only the momentum, while the second can change both position and momentum. Both steps leave the canonical joint distribution of (Θ, Φ) invariant, and hence their combination also maintains the distribution invariant. In the first step, new value of Φ is randomly drawn from its Gaussian distribution, independently of the current value of Θ . Because Θ is not changed, and Φ is drawn from its correct conditional distribution. Thus, this step obviously keeps the canonical joint distribution invariant. In the second step, a Metropolis update is performed to propose a new state by using Hamiltonian dynamics. Starting with the current state (Θ, Φ) , Hamiltonian dynamics is simulated for L steps using the leapfrog method with a stepsize of Δs . Here, L and Δs are parameters of the algorithm, which need to be tuned to obtain good performance. The momentum variables at the end of this L -step trajectory are then negated, giving a proposed state (Θ_L, Φ_L) . It is easy to obtain $P(\Theta, \Phi)/P(\Theta_L, \Phi_L) = \exp\{-H(\Theta, \Phi) + H(\Theta_L, \Phi_L)\}$. Then this proposed state is accepted as the next state of the Markov chain with probability

$$\min [1, \pi(\Theta_L, \Phi_L)/\pi(\Theta, \Phi)] = \min [1, \exp(-H(\Theta_L, \Phi_L) + H(\Theta, \Phi))].\tag{2.13}$$

If the proposed state is not accepted, the next state is the same as the current state. The negation of the momentum variables at the end of the trajectory makes the Metropolis proposal symmetrical, as needed for the acceptance probability (2.13) to be valid.

For the mixed normal GARCH-type models, the gradient of the kinetic function $\nabla K(\Phi)$ is $\Sigma^{-1}\Phi$, and the potential function $U(\Theta) = U(\Theta|\mathbf{y}, \mathbf{z})$ is first-order differentiable. Then the gradient $\nabla U(\Theta)$ can be calculated. Note that ρ, λ is different from the other parameters in $U(\Theta)$. The partial derivatives of U with respect to ρ and λ can be calculated directly.

However, the other parameters in $U(\Theta)$ can not be expressed as the explicit formula, and can be calculated according to the chain rule as follows:

$$\begin{aligned} \frac{\partial U}{\partial \Theta_i} &= \sum_t (U(\Theta|\mathbf{y}, \mathbf{z}))'_{f(\Theta, F_{t-1})} (f(\Theta, F_{t-1}))'_{\Theta_i} \\ &= \sum_t (U(\Theta|\mathbf{y}, \mathbf{z}))'_{h_t} (h_t)'_{\Theta_i}. \end{aligned} \tag{2.14}$$

Example 2.1 Notice that the partial derivative $(h_t)'_{\Theta_i}$ can be obtained indirectly. For example, for the standard GARCH model, the parameters in $f(\Theta, F_{t-1})$ are $\alpha_0, \alpha_i, \beta_j, i = 1, 2, \dots, p, j = 1, 2, \dots, q$. The derivative of α_0 can be written as follows,

$$(h_t)'_{\alpha_0} = 1 + \beta_1(h_{t-1})'_{\alpha_0} + \dots + \beta_q(h_{t-q})'_{\alpha_0}, \tag{2.15}$$

$$\partial U / \partial \alpha_0 = \sum_{t=l}^n (h_t)'_{\alpha_0} u_t, \tag{2.16}$$

where $l = \max(p + 1, q + 1)$, $u_t = \frac{1}{2}(\frac{1}{h_t} - \frac{y_t^2}{\sigma^2 h_t^2})I_{\{t:z_t=1\}} + \frac{1}{2}(\frac{1}{h_t} - \frac{\lambda y_t^2}{\sigma^2 h_t^2})I_{\{t:z_t=0\}}$. $(h_{t-k})'_{\beta_j}$ and h_k , for $l - q \leq k \leq l - 1, i = 0, 1, 2, \dots, p, j = 1, 2, \dots, q$ are assumed to be known. The derivatives of other parameters α_i, β_j have similar expressions, for $i = 1, 2, \dots, p, j = 1, 2, \dots, q$. Thus using the above formula, the gradient of $U(\Theta)$ can be calculated.

If $f(\Theta, F_{t-1})$ is the case 5 in (2.1), one can treat $\ln h_t$ as h_t to calculate the partial derivatives.

In the mixed normal GARCH-type models, generating the latent data \mathbf{z} is very important for inferring the interested parameter. There are several methods can be found in the literature. For example, Tanner and Wong (1987) led to the posterior of the interested parameter by combining the observed data with latent data. Diebolt and Robert (1994) presented a Bayesian method to evaluate the interested mixture distribution in terms of the missing data scheme. Denote $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ to be the observed data. Referring to Tanner and Wong (1987) and Diebolt and Robert (1994), we first generate the missing values $\mathbf{z} = (z_1, z_2, \dots, z_n)'$, and then sample the parameters from the posterior function

$p(\Theta|\mathbf{y}, \mathbf{z})$ based on the complete data $(y_t, z_t), t = 1, 2, \dots, n$. The algorithm works at the step m for $m = 1, \dots, N$ as follows:

- (1) Use $\Theta^{(m)}$ and \mathbf{y} to generate $z_t^{(m)}$ from posterior $p(z_t|\Theta^{(m)}, y_t)$, where

$$p(z_t|\Theta^{(m)}, y_t) = B\left(1, \frac{pr_1}{pr_1 + pr_2}\right), \tag{2.17}$$

$B(1, \frac{pr_1}{pr_1+pr_2})$ is the Bernoulli distribution with probability $\frac{pr_1}{pr_1+pr_2}$ and $pr_1 = \frac{\rho}{\sqrt{\sigma^2 h_t}} \exp(-\frac{y_t^2}{2\sigma^2 h_t}), pr_2 = \frac{(1-\rho)}{\sqrt{\sigma^2 h_t/\lambda}} \exp(-\frac{\lambda y_t^2}{2\sigma^2 h_t})$.

- (2) Draw momentum variable Φ from a zero-mean Gaussian distribution with covariance matrix $\Sigma = \text{diag}(1, 1, \dots, 1), \Phi \sim N(\mathbf{0}, \Sigma)$.
- (3) Use the Algorithm 1 to propose a new state (Θ^*, Φ^*) .
- (4) Accept the proposal (Θ^*, Φ^*) with probability Pr and reject the proposal with probability $1 - Pr$, where

$$Pr = \min [1, \exp(-H(\Theta^*, \Phi^*) + H(\Theta, \Phi))]. \tag{2.18}$$

Algorithm 1 Hamiltonian Monte Carlo

Input: initial (Θ, Φ) , step size Δs and steps L

Output: final (Θ^*, Φ^*)

$\Phi = \Phi - \Delta s \nabla U(\Theta|\mathbf{y}, \mathbf{z})/2$

for $i=1$ to L **do**

$\Theta = \Theta + \Delta s \nabla K(\Phi)$

while Θ_i violated constraints **do**

if $(\Theta_i > UB_i) \Theta_i = UB_i - (\Theta_i - UB_i), \phi_i = -\phi_i$

if $(\Theta_i < LB_i) \Theta_i = LB_i + (LB_i - \Theta_i), \phi_i = -\phi_i$

end while

end for

$\Phi = \Phi - \Delta s \nabla U(\Theta|\mathbf{y}, \mathbf{z})/2$

Remark 2.2 In fact, Algorithm 1 is the leapfrog method. Here, UB_i and LB_i represent the upper and lower bounds of each parameter respectively.

The simplest method of discretization equations (2.10) is leapfrog method with order $(\Delta s)^3$ local error, while Euler’s method and its modified version have order $(\Delta s)^2$ local error [See Leimkuhler and Reich (2004) for details]. Because a big step size will lead to a poor acceptance rate and a small step size will waste computation time. As Neal’s comments, a relatively suitable step size Δs should be considered based on computational efficiency. The adjustment of parameter Δs is necessary as well as misleading. Hence, it is essential to run multiple Markov chains with different initial values to ensure that the parameter Δs is sufficient to keep the algorithm stable.

According to Beskos et al. Beskos et al. (2013) and Hoffman and Gelman Hoffman and Gelman (2014), tuning of parameters Δs and L determines the performance of the HMC algorithm. An inappropriate step size Δs will destroy the stability of simulated trajectory, and then affects the acceptance rate. The steps L means the total run in the leapfrog trajectory. Care must be taken when we choose the step size of HMC algorithm, because too small step size can maintain the stability of the trajectory. After a leapfrog iteration we can only get proposal point, which is close to the previous one. Moreover, an overlong step size will destroy the stability. A selected critical step size can keep the algorithm stable and efficient. The initial L is assumed to be 1. Given the initial Θ_0, Φ_0 and s , then one use the leapfrog method to get Θ_L, Φ_L . If $\exp\{-H(\Theta_L, \Phi_L) + H(\Theta_0, \Phi_0)\} > 0.5$, then $s = 1/2s$; otherwise, $s = 2s$. In simulation experiments and empirical analysis, taking into account the acceptance rate of these two stages, we set a threshold 0.5 and run the Metropolis update multiple times. According to Algorithm 1, repeating this process until the algorithm remains stable. The final acceptance rate for the proposed points is about 0.85. After getting the stable step size, the number of steps L can be increased according to the correlation of the sampling results. One can find an appropriate step size through sufficient preliminary runs of several chains. Furthermore, it's necessary to choose an appropriate L that avoids producing a dependent point.

In the previous works, the HMC algorithm performs well in sampling unconstrained parameters. As for handling distributions with the constraints on the variables, we can reparameterize the variables. Taking the GARCH model as an example, because the coefficients of autoregressive term α_i is on interval (0,1), we can take logarithm of α_i . Under the transformation $\alpha_i^* = \ln(\alpha_i/(1 - \alpha_i))$, the HMC algorithm also performs well. However, the reparameterization will take more arithmetic operations. Another way for the posterior with restricted variables can produce similar results. In order to deal with the constraints, once the variable violates any constraints, we set the value of the potential energy and obtain an acceptance rate of almost zero immediately. Thus the parameters will fall in the limited interval, see e.g. Algorithm 1. In each leapfrog iteration, the boundaries of variables can be seen as "walls". If Θ_i is beyond the boundaries, the state (Θ_i, Φ_i) bounces off the "walls" perpendicularly, see Betancourt (2011) for details.

2.4 Bayesian Forecasting

Volatility and VaR forecasting are important in analyzing derivative pricing, yielding a good portfolio for investments and managing market risks in financial markets. In this section, the estimation of in-sample volatilities and the prediction of future volatilities are also investigated. In order to predict the volatility and VaR, based on the HMC iterations, a simulation-based approach is used to obtain the posterior samples and distributions of volatilities $h_t, t = 1, 2, \dots, n$, where $h_t = f(\Theta, F_{t-1})$ is a function of parameter Θ . For each HMC iteration of $\Theta^{(m)}, m = 1, 2, \dots, N$, we estimate the in-sample volatilities $h_t^{(1)}, \dots, h_t^{(N)}$,

$$\hat{h}_t \simeq E(h_t | \mathbf{y}_n) \simeq \frac{1}{N} \sum_{m=1}^N h_t^{(m)}, \quad t = 1, 2, \dots, n, \tag{2.19}$$

where $\mathbf{y}_n = (y_1, y_2, \dots, y_n)'$.

For one-step ahead prediction of y_{n+1} and h_{n+1} , one can obtain samples from the predictive distributions $f(y_{n+1} | \mathbf{y}_n)$ and $f(h_{n+1} | \mathbf{y}_n)$, respectively. Note that $h_{n+1}^{(m)}$ is the function of parameter $\Theta^{(m)}$ and \mathbf{y}_n . Then $h_{n+1}^{(1)}, h_{n+1}^{(2)}, \dots, h_{n+1}^{(N)}$ can be obtained and form a sample of the predictive distribution $f(h_{n+1} | \mathbf{y}_n)$. The posterior distribution $f(y_{n+1} | \mathbf{y}_n, \Theta^{(m)})$ has a simple form, so $y_{n+1}^{(m)}$ can be drawn from a zero mean mixed normal distribution with variance $h_{n+1}^{(m)}$. Hence, $y_{n+1}^{(1)}, y_{n+1}^{(2)}, \dots, y_{n+1}^{(N)}$ is a sample of the predictive distribution $f(y_{n+1} | \mathbf{y}_n)$. From this perspective, the distribution of y_{n+1} can be calculated by means of the posterior sample mean, that is

$$f(y_{k+1} | \mathbf{y}_n) \simeq \frac{1}{N} \sum_{m=1}^N f(y_{k+1} | \mathbf{y}_n, \Theta^{(m)}). \tag{2.20}$$

For the prediction of two-step ahead volatility, we can obtain the predictive distribution of h_{n+2} through the calculated y_{n+1} and h_{n+1} . According to the $\Theta^{(m)}$ and $h_{n+1}^{(m)}$, the predictive $y_{n+1}^{(m)}$ and $h_{n+2}^{(m)}$ can be obtained in the same way. Similarly, $y_{n+2}^{(m)}$ also follows the mixed normal distribution given $h_{n+2}^{(m)}$. Then it can be generated from the mixed normal density with known $\Theta^{(m)}$ and $h_{n+1}^{(m)}$. For $t = n + j, j \geq 2$, the prediction of the volatilities h_t and y_t can be carried out through the above process. If we need to obtain the predicted distributions $f(h_t | \mathbf{y}_{n+j-1})$ and $f(y_t | \mathbf{y}_{n+j-1})$, we can repeat the prediction process above. The final prediction distribution can help us compute the prediction interval and the prediction mean at time t .

Due to the common presence of extreme events in financial time series, VaR has become a widely used measure of market risk. It is usually defined as the loss of a financial asset or securities portfolio, which is exceeded with a predetermined probability α over a time horizon of d periods,

$$Pr(y[d] \leq -VaR) = \alpha, \tag{2.21}$$

where $y[d] = y_{n+1} + y_{n+2} + \dots + y_{n+d}$. In fact, VaR is the α -quantile of distribution of $y[d]$ at a given confidence level α . To forecast VaR, we need to estimate the extreme α th percentiles. In this paper, the probability α of interest is 0.01 and 0.05. Given $y_{n+1}^{(m)}, y_{n+2}^{(m)}, \dots, y_{n+d}^{(m)}, m = 1, 2, \dots, N$, the d -period $\alpha\%$ VaR can be evaluated from distribution of $f(y[d] | \mathbf{y})$, see Jorion (2000) for details.

3 Simulated Example and Real Data Example

In this section, we first use the simulated experiment to illustrate the performance of the HMC method for the standard GARCH model with mixed normal errors, and then apply the proposed method to a real data analysis. The program is written based on R.

3.1 Simulation Experiment

Here, the mixed Gaussian GARCH(1,1) model is adopted as the simulated example, that is

$$\begin{aligned}
 y_t &= \sqrt{h_t} \epsilon_t, \\
 h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1},
 \end{aligned}
 \tag{4.1}$$

where ϵ_t follows a mixture Gaussian distribution (2.3) and the true values are set as follows,

$$\alpha_0 = 0.1, \alpha_1 = 0.2, \beta_1 = 0.5, \rho = 0.8, \lambda = 0.15.
 \tag{4.2}$$

We first generated three series with sample size 800, 1000, 2000, which are denoted as (s1), (s2) and (s3) respectively. Then 10,000 HMC iterations are carried out for all

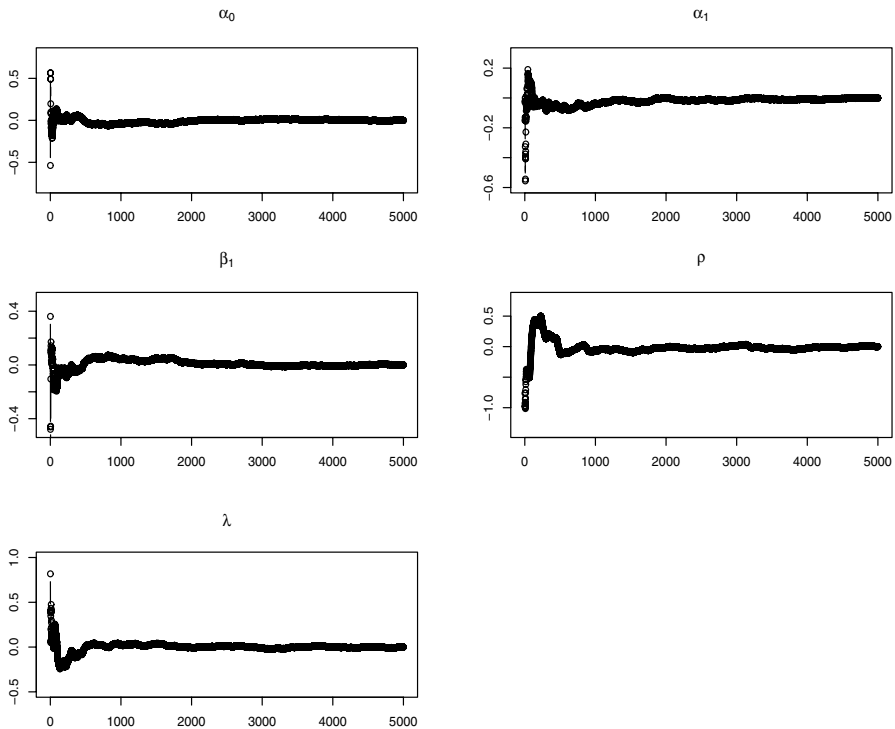


Fig. 1 The CUMSUM plots of posterior mean estimates for s2 of the model 4.1

series with the same initial value $\Theta^{(0)} = (0.5, 0.4, 0.4, 0.65, 0.5)$. The first 5000 iterations considered as burn-in, the trace plots and histograms of all series are shown in Figs 9–14 in Appendix. The trace plots indicate that the HMC algorithm is successful in exploring the posterior density of the parameters. The CUMSUM plots for simulated s2 are displayed in Fig. 1, which shows that the estimates of parameters have good convergence. Meanwhile, convergence diagnosis was investigated by the Geweke test [Geweke (1992)], and the results are more convinced that the chains have converged from Table 8.

Remark 3.1 In Algorithm 1, as Neal (2011) discussed, the stability and periodicity of the trajectory needs to be considered when the HMC algorithm is executed. The leapfrog step $L = 59$ and multiple step sizes Δs are 0.003, 0.002, 0.002, 0.012 and 0.003 are reasonable. They ensure the stability as well as take into account the excellent performance of the program. The multiple step sizes used here can reduce the correlation of the samples. As discussed above, when discretizing the Hamilton equations, the discretization error can be kept equal to 0 theoretically. However, if the discretization error exists, then the acceptance rate is less than 1, that is round 85%.

Remark 3.2 Here two tests are used to test the convergence of the Markov chain. One is the visual inspection of CUMSUM statistics proposed by Yu and Mykland (1994), and defined by

$$CS_t = \left(\frac{1}{t} \sum_{i=1}^t \theta^i - \mu_\theta \right) / \sigma_\theta, \tag{4.3}$$

where μ_θ and σ_θ are the empirical mean and standard deviation of the N draws. Another is the Geweke test, which can be obtained by the R package “coda” [see Plummer et al. (2008) for details].

To verify the usefulness of our method a bit further, we conduct an comparison between the HMC algorithm and the Griddy-Gibbs (GG) sampler through 100 simulations. The summary of the results for the two methods is recorded in Table 1, which includes the posterior means, medians and standard deviations (SD) in parentheses.

From Table 1, we can see that the estimated results of two methods are very well. The average of 100 samples of all estimated parameters is very close to the true

Table 1 Comparison results for the mixture normal GARCH(1,1) model using HMC algorithm and GG sampler based on 100 replicates samples

Parameter	$\alpha_0(0.1)$	$\alpha_1(0.2)$	$\beta_1(0.5)$	$\rho(0.8)$	$\lambda(0.15)$	Time
HMC-mean	0.1167	0.2155	0.4478	0.7922	0.1486	
HMC-median	0.1187 (0.0269)	0.2139 (0.0621)	0.4356 (0.0866)	0.7970 (0.0449)	0.1486 (0.0210)	12.23 mins
GG-mean	0.1163	0.2339	0.4611	0.7929	0.1406	
GG-median	0.1128 (0.0247)	0.2244 (0.0604)	0.4636 (0.0917)	0.7961 (0.0467)	0.1416 (0.0200)	56.22 mins

Table 2 Comparisons of HMC algorithm and GG sampler in terms of MSE, ESS/min and MAD for each parameter

	α_0	α_1	β_1	ρ	λ
<i>HMC</i>					
MSE	0.0015	0.0304	0.0193	0.0018	0.0006
ESS/min	411	496	381	28	159
MAD	0.0318	0.0498	0.1212	0.0413	0.0230
<i>GG</i>					
MSE	0.0036	0.0223	0.0203	0.0022	0.0030
MAD	0.04360	0.0917	0.1288	0.0388	0.0335
ESS/min	8	27	5	12	43

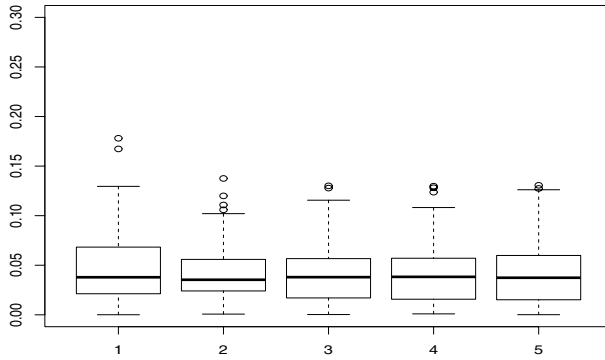


Fig. 2 The boxplot of AE of predictive volatilities for future times $T = 1001, \dots, 1005$

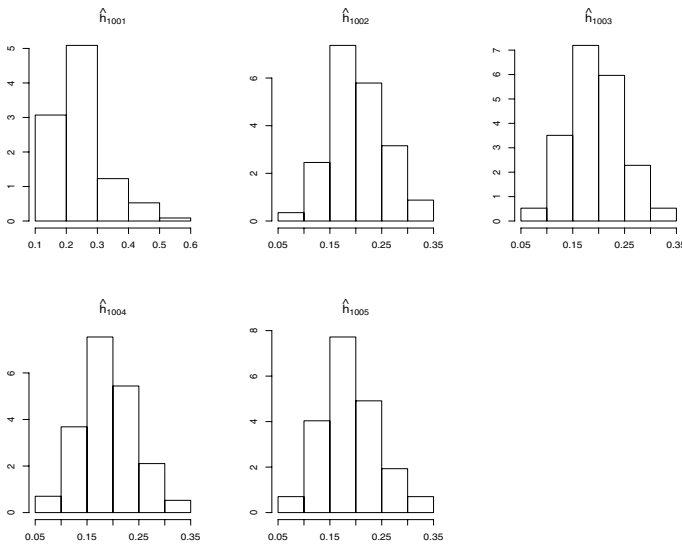


Fig. 3 The histograms of predictive distribution of volatilities for simulated series at times $T = 1001, \dots, 1005$ with sample size 1000

parameters, excluding a little bias for β_1 . In particular, the posterior estimate of ρ is good and equal to 0.7922, its SD is small. Meanwhile, the median and mean of each parameter are very close, which indicate that the distributions are approximately symmetrical.

The posterior estimates of the HMC algorithm are close to the GG sampler. However, the HMC algorithm runs more quickly than the Griddy-Gibbs sampler. As shown in the last column of Table 1, the computation time of the HMC algorithm for the mixed Gaussian GARCH(1,1) model is 12.23 minutes (mins), while the GG sampler is 56.22 mins. In other words, with the same sample size, the time consumption of the GG sampler is almost 5 times that of the HMC algorithm.

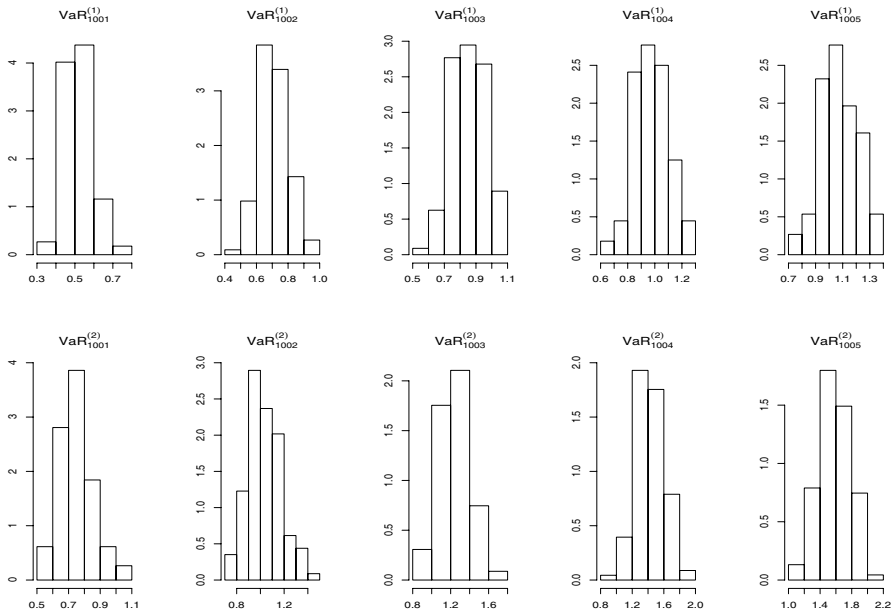


Fig. 4 The predictive distributions of $VaR_{1001} \sim VaR_{1005}$ at 5% (top) and 1% (bottom) level

Table 3 The predictive means, SD's and 95% credibility intervals for volatilities at times $T = 1001, \dots, 1005$ based on 100 replicates

	h_{1000+1}	h_{1000+2}	h_{1000+3}	h_{1000+4}	h_{1000+5}
True	0.2304	0.1984	0.1898	0.1872	0.1823
Mean	0.2402	0.2051	0.1948	0.1915	0.1905
	(0.2254,0.2550)	(0.1953,0.2149)	(0.1851,0.2046)	(0.1817,0.2014)	(0.1805,0.2005)
SD	0.0813	0.0534	0.0530	0.0538	0.0542

Table 4 The predictive means, SD's and 95% credibility intervals of VaR at times $T = 1001, \dots, 1005$ with probabilities $\alpha=0.05$ and 0.01 based on 100 replicates

	VaR_{1000+1}	VaR_{1000+2}	VaR_{1000+3}	VaR_{1000+4}	VaR_{1000+5}
Mean ¹	0.5165	0.7085	0.8519	0.9718	1.0783
	(0.5029,0.5302)	(0.6911,0.7260)	(0.8315,0.8723)	(0.9486,0.9950)	(1.0524,1.1043)
SD ¹	0.0729	0.0933	0.1090	0.1241	0.1385
Mean ²	0.7458	1.0332	1.2477	1.4287	1.5866
	(0.7256,0.7660)	(1.0073,1.0591)	(1.2176,1.2778)	(1.3946,1.4628)	(1.5486,1.6246)
SD ²	0.1089	0.1397	0.1621	0.1836	0.2047

¹ VaR with probability 0.05

² VaR with probability 0.01

In addition, Table 2 shows the detailed comparisons of the two methods, including mean square error (MSE), mean absolute deviation (MAD) for each parameter and effective sample size per minute (ESS/min). We can see that compared with GG sampler, HMC algorithm has a lower MAD and a lower MSE for most parameters. However, ESS/min of HMC algorithm is much higher than GG sampler.

Furthermore, we consider the forecasting by the proposed HMC algorithm. 1005 time series data by model (4.1) is used for predictions, while the first 1000 data are used for the posterior estimation of the model. The HMC algorithm is also performed 10000 iterations with the initial value $\Theta^{(0)} = (0.5, 0.4, 0.4, 0.65, 0.5)'$, and the first 5000 iterations are as burn-in values. We repeat the process to obtain the posterior means, SD's, 95% credibility intervals and predictive distributions of volatilities and VaR. Moreover, in order to ensure the validity of the prediction, we calculate the results of 100 replicates and obtain the absolute error (AE) of volatilities, which are $|\hat{h}_{1000+m} - h_{1000+m}|, m = 1, 2, \dots, 5$ plotted in Fig. 2.

Remark 3.3 Ausín and Galeano (2007) and Xia et al. (2017) suggested choosing fixed grids with 40 points to compute the value of GG sampler. For the GG sampler, we choose 40 gridly points and the same initial value $\Theta^{(0)}$, which can explore the parameter space sufficiently. In addition, linear interpolation is used to implement the approximation of the cumulative distribution function.

Using the steps described above, one can obtain the estimation of predictive distribution. The histograms of the predictive distributions of volatilities and VaR are shown in Figs. 3 and 4. The estimated means, SD's and 95% credibility intervals are summarized in Tables 3 and 4. From the histograms of predictive volatilities in Fig. 3, we believe that the distributions of future volatilities are nearly symmetric. For the forecasting of

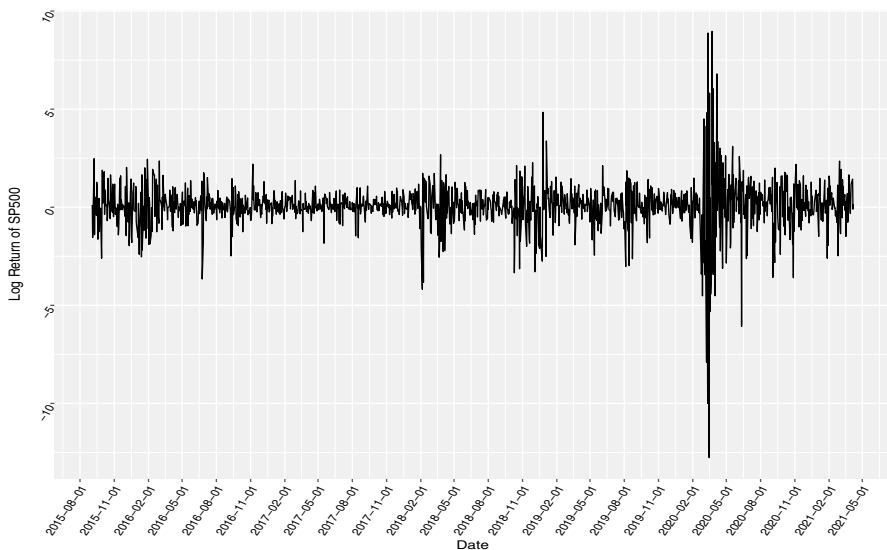


Fig. 5 The the log return plot of SP 500

Table 5 The posterior estimations for SP500 by using HMC algorithm and GG sampler

Parameter	α_0	α_1	β_1	ρ	λ	Time
HMC-mean	0.0239	0.1936	0.8032	0.8873	0.1631	14.52 mins
HMC-median	0.0231 (0.0068)	0.1915 (0.0307)	0.8046 (0.0273)	0.8964 (0.0426)	0.1618 (0.0323)	
GG-mean	0.0453	0.2351	0.7649	0.8593	0.1560	55.65 mins
GG-median	0.0431 (0.0191)	0.2370 (0.0626)	0.7630 (0.0489)	0.8644 (0.0732)	0.1609 (0.0504)	

volatility, the predictive mean is very close to the true value, and the SD's are very small. For the boxplot of predictive volatilities in Fig. 2, we can see that AE plots are pretty small, which indicates that our suggested method provides an accurate estimation. Figure 4 shows that most of the predictive distributions of VaR are symmetric, while a small part of them are skewed.

Compared with GG sampler, the simulation results show that HMC algorithm not only provides more accurate estimations and predictions, but also performs higher efficiency for the mixture normal GARCH model.

3.2 Real Data Example

In order to illustrate the good performance of the mixed GARCH type models in practice, we demonstrate the feasibility of the model with the real data analysis. The daily closing prices of SP500 index from Sep./3/2015 to Apr./7/2021 are selected with $T = 1407$ observed data. The sample mean, variance, skewness and kurtosis of log return series are 0.0525, 1.196, -1.0865 and 24.1190 respectively. The return of the observation series is defined by $y_t = 100(\log P_t - \log P_{t-1})$ shown in Fig. 5, where P_t is the closing price at time t . The Dickey-Fuller test is used to verify the stationarity of this time series, and the results show that the log return series is stationary. P_t is used to study the estimation problem of the volatility model. The volatility means the movement of the stock price. For example, the bigger volatility indicates that the stock will increase in price at a future point in time. Then, the prediction of volatilities in the future time can help investors to judge whether to hold the stock or not. Meanwhile, the predicted distribution of VaR is also carried out, which is concerned with a forecast of the possible losses of the investment portfolio over a given time interval. Thus, using the forecast results, financial institutions or individuals can respond in time to reduce losses when dealing with market risks.

The analysis of the real data is based on model (4.1). We use the HMC algorithm to estimate parameters of the model, and GG sampler is also used for the comparison. 10,000 iterations are carried out for the HMC algorithm and the GG sampler with discarding the first 5000 iterations. The posterior means, medians and SD's in parentheses are shown in Table 5.

Table 5 shows that the estimation for posterior mean of the two methods are very close, indicating that the HMC algorithm is as much reliable as the GG sampler. Although the posterior results of the parameters are generally close, the SD's of the

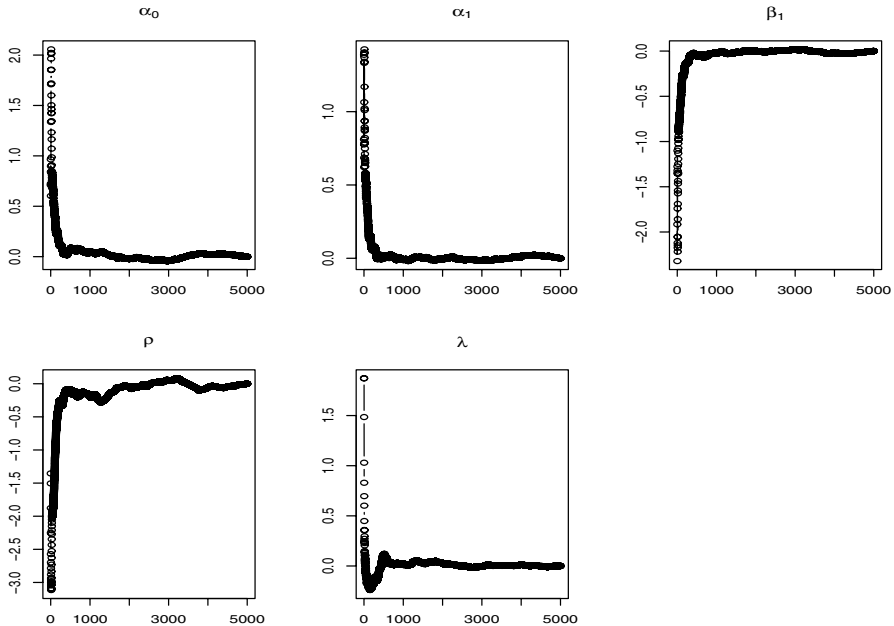


Fig. 6 The CUMSUM plots of posterior mean estimates for mixed normal GARCH model

estimated parameters of the HMC method is smaller, and HMC algorithm consumes less time in terms of running time. Moreover, the posterior mean of parameter ρ equals to 0.8873 and SD is 0.0426. That is to say, about 89% of the data may come from a normal distribution with a small variance 0.64, and the other 11% of the data may come from a normal distribution with a big variance 3.96. In reality, this finding result is very close to the features of the log return series. For example, there were many trade disputes around the world in 2018, which caused the daily closing price of stocks to fall. Another notable example is the global epidemic of COVID-19 in 2020. The epidemic has caused unprecedented trauma to the economies of countries around the world, and the stock market has also experienced severe shocks. Many stock prices, including the SP500 index, have been strongly affected by the epidemic. The observed data in Fig. 5 has a little volatility for most of the period before 2020, and large fluctuations in the short term after 2020. The excess kurtosis of the observed series is 24.12, which implies that capturing the extreme events using t or normal distribution is not a good choice. Under the framework of the mixture normal GARCH model, the heavy-tailed distributions of the returns can be well fitted.

Table 6 The predictive means, SD's and 95% credibility intervals for volatilities h_T at times $T = 1408, \dots, 1412$ of SP 500 using HMC method

	h_{1407+1}	h_{1407+2}	h_{1407+3}	h_{1407+4}	h_{1407+5}
Mean	1.1798 (1.1795,1.1801)	1.1027 (1.1019,1.1035)	1.0324 (1.0313,1.0335)	0.9683 (0.9669,0.9697)	0.9101 (0.9086,0.9117)
SD	0.0014	0.0041	0.0058	0.0069	0.0078

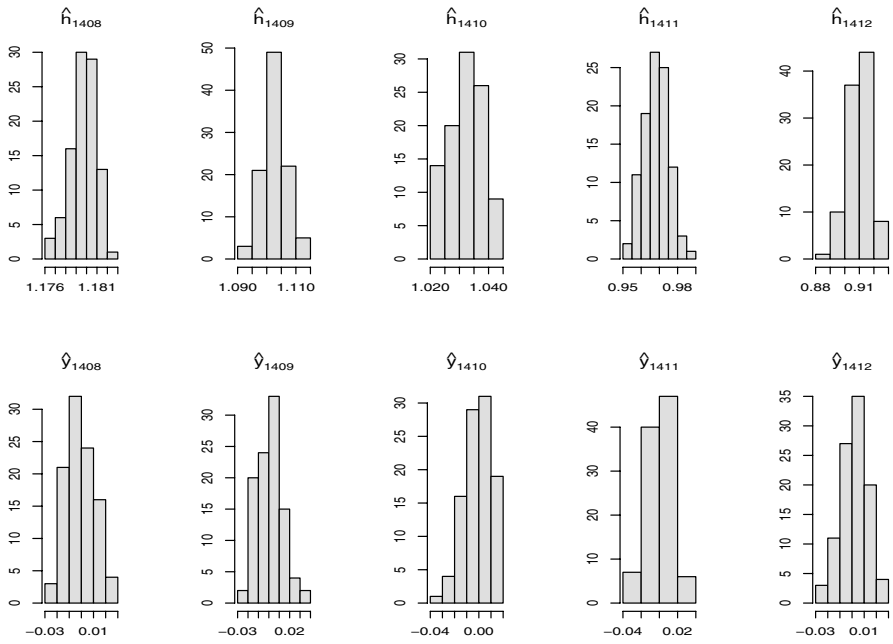


Fig. 7 The histograms of predictive $h_{1408} \sim h_{1412}$ and $y_{1408} \sim y_{1412}$ for SP 500

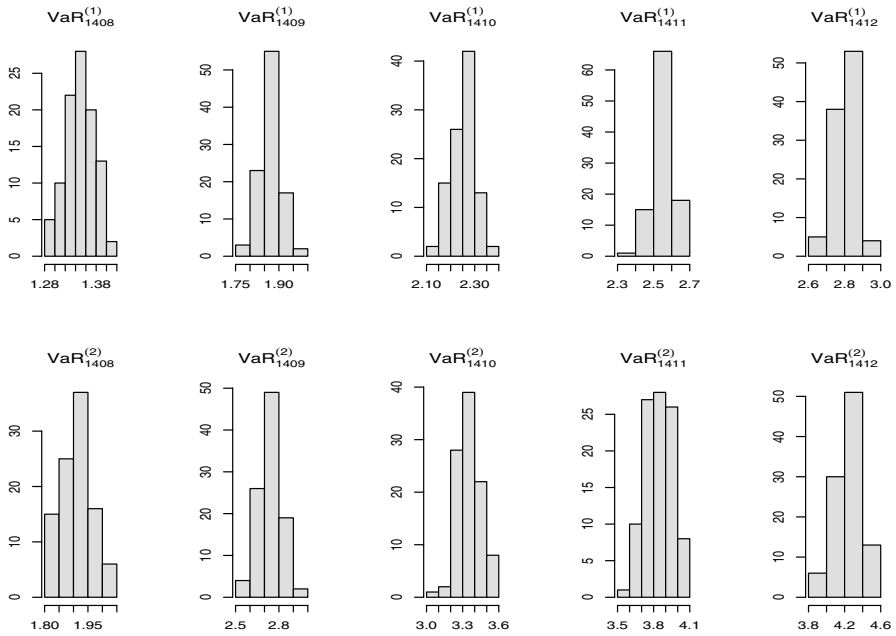


Fig. 8 The predictive distributions of $VaR_{1408} \sim VaR_{1412}$ for SP 500 at 5%(top) and 1%(bottom) level

In addition, the diagrams and histograms of the posterior estimation for each parameter are shown in Figs. 15, 16 of Appendix. It can be seen from trace plots of each parameter that the HMC algorithm explores the parameter space well and does not fall into the local region in Fig. 15. According to the histograms in Fig. 16, except that the histogram of ρ is slightly right skew, the marginal posterior distribution of the other parameter is almost symmetrical.

We use the visual inspection of CUMSUM statistics to check convergence of the HMC algorithm, and the plots based on 5000 draws with discarding the first 5000 draws are reported in Fig. 6. In the CUMSUM plots, we can see that the convergence of parameter ρ is slower than other parameters, and the convergence is relatively good for the parameters $\alpha_0, \alpha_1, \beta_1$ and λ . The Geweke statistics of $\alpha_0, \alpha_1, \beta_1, \rho$ and λ are 0.6575, $-0.3713, -0.1276, -0.2118$ and 0.1332, respectively, indicating that all estimated results are convergent. For the parameters of the HMC algorithm, we run several Markov chains to find a suitable step size. In fact, we can refer to Beskos et al. (2013), which provided theoretical analysis of optimal step sizes for HMC algorithm. It's worth noting that the multiple step size Δs selected to keep a good performance are $\{0.0025, 0.0017, 0.0025, 0.0067, 0.0042\}$, and the leapfrog step is 55. Considering the discretization error of the leapfrog, the selection of Δs and L results in an acceptance rate of 77%.

For the future prediction, we use the same approach described in Sect. 2.4, which is to perform 10000 iterations in each HMC algorithm. We consider 5-step predictions of 1407 observations. Table 6 displays the estimations and confidence intervals of volatilities. As can be seen from Table 6, the length of confidence intervals and the SD's of volatilities are quite small, although the SD's of volatilities increases over time. The results of credibility intervals show that HMC algorithm provides an accurate forecasting intervals. The predictive histogram of y_t and forecasts of h_t are shown in Fig. 7, which can be found that the predictive distributions of y_t and h_t are nearly symmetric. The mean of y_t in the next 5 steps is about zero, which indicates that the future log returns are unlikely to occur large volatilities.

The predictive distributions of VaR at 1% and 5% level can be seen from Fig. 8 that they are almost symmetric. The posterior estimations of VaR are shown to be well in Table 7. From the results shown in Table 7 and Fig. 8, the predictive distributions and

Table 7 The predictive means, SD's and 95% credibility intervals of VaR at times $T = 1408.., 1412$ with probabilities $\alpha=0.05$ and 0.01 for SP 500

	VaR_{1407+1}	VaR_{1407+2}	VaR_{1407+3}	VaR_{1407+4}	VaR_{1407+5}
Mean ¹	1.3472 (1.3417,1.3527)	1.8703 (1.8627,1.8778)	2.2513 (2.2416,2.2611)	2.555 (2.5442, 2.5662)	2.8108 (2.7989,2.8228)
SD ¹	0.0279	0.0377	0.0491	0.0556	0.0603
Mean ²	1.9090 (1.8982,1.9189)	2.7373 (2.7223,2.7523)	3.3527 (3.3356,3.3718)	3.8446 (3.8236,3.8657)	4.2381 (4.2086,4.2677)
SD ²	0.0545	0.0756	0.0963	0.1061	0.1489

¹ VaR with probability 0.05

² VaR with probability 0.01

credibility intervals for VaR are fairly symmetric. From the Table 7, we can see that the estimated values of VaR increase over time, which imply that the maximum possible loss increases as the future period $[d]$ increases. VaR is related to the market risk, investors should choose a reasonable trading time to prevent from loss.

The log return of the SP500 index exhibits heavy-tailed distributions and volatility cluster. The GARCH model with mixed normal distribution can capture the extreme events of market risk as well as the heavy-tailed distribution of the real data well. Meanwhile, the final estimated results illustrate that HMC algorithm performs more accurate and faster than GG sampler. Therefore, we can conclude that the HMC algorithm provides a good estimate and forecast for the log return of the SP500 index using GARCH model with mixed normal errors.

4 Conclusion

In this article, we carry out Bayesian inference and prediction for the GARCH-type models with mixed normal errors by the HMC algorithm. This new approach can be simply constructed to capture the GARCH effect in the heavy-tailed behavior of the distributions. Comparing with the GG sampler, the HMC algorithm performs higher sampling efficiency and accuracy for the mixture normal GARCH model in the simulation experiments as well as the empirical analysis.

Because the step size and the number of steps will affect the performance of the HMC algorithm, attention should be paid to design the scheme for adjusting the

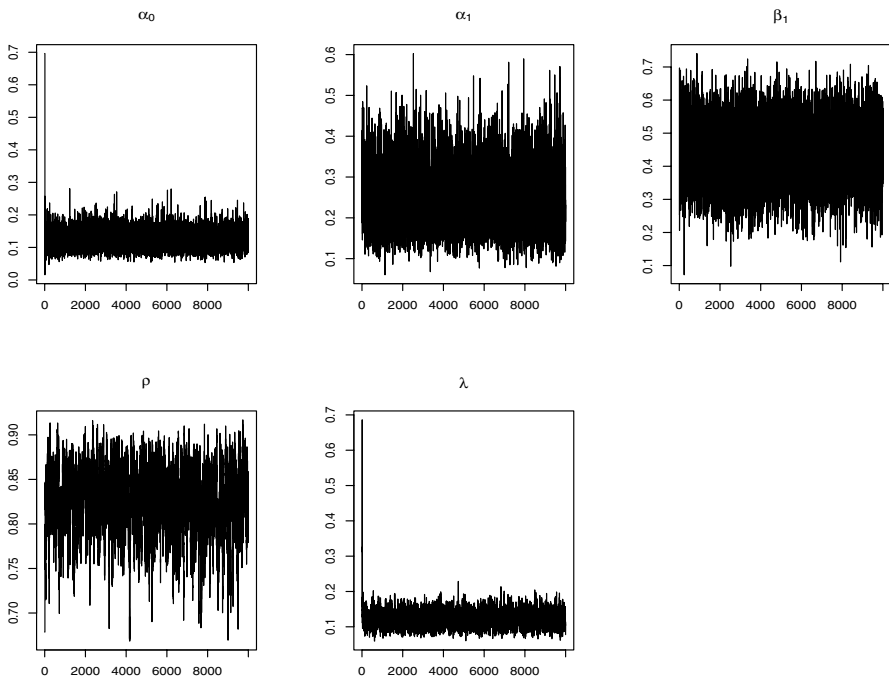


Fig. 9 The trace plots of parameters for s1

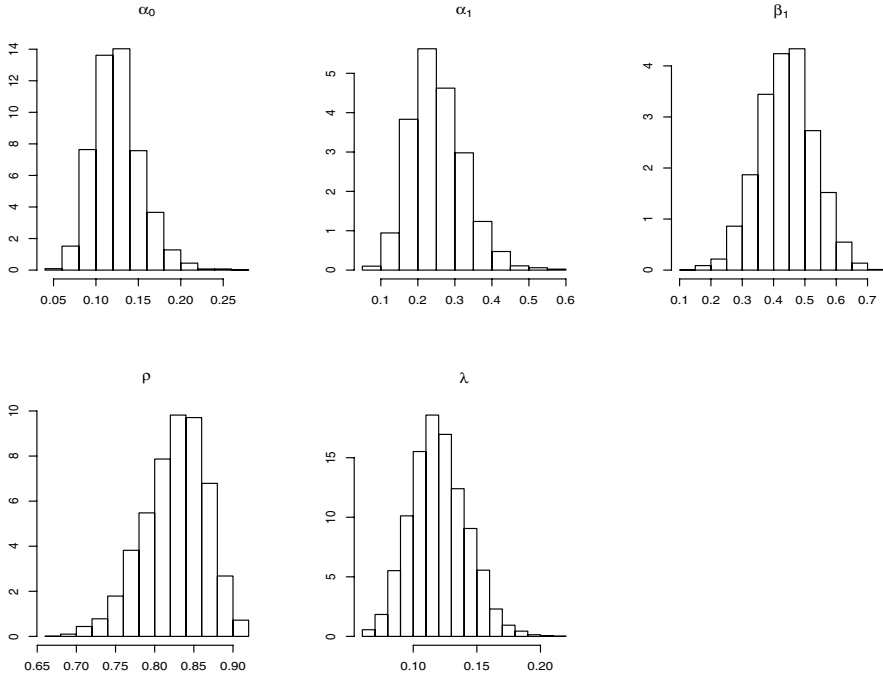


Fig. 10 The histograms of parameters for s1

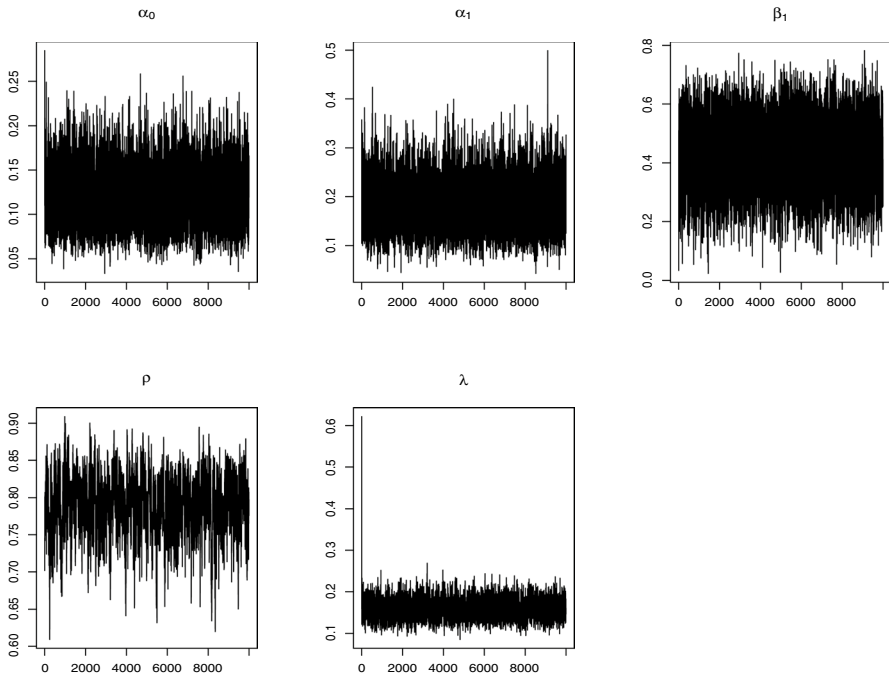


Fig. 11 The trace plots of parameters for s2

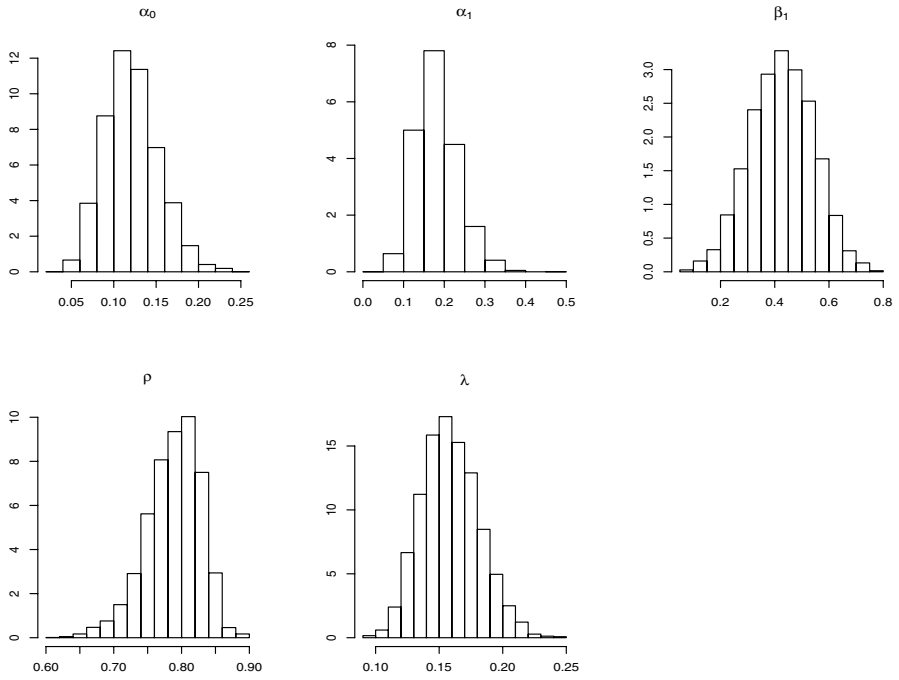


Fig. 12 The histograms of parameters for s2

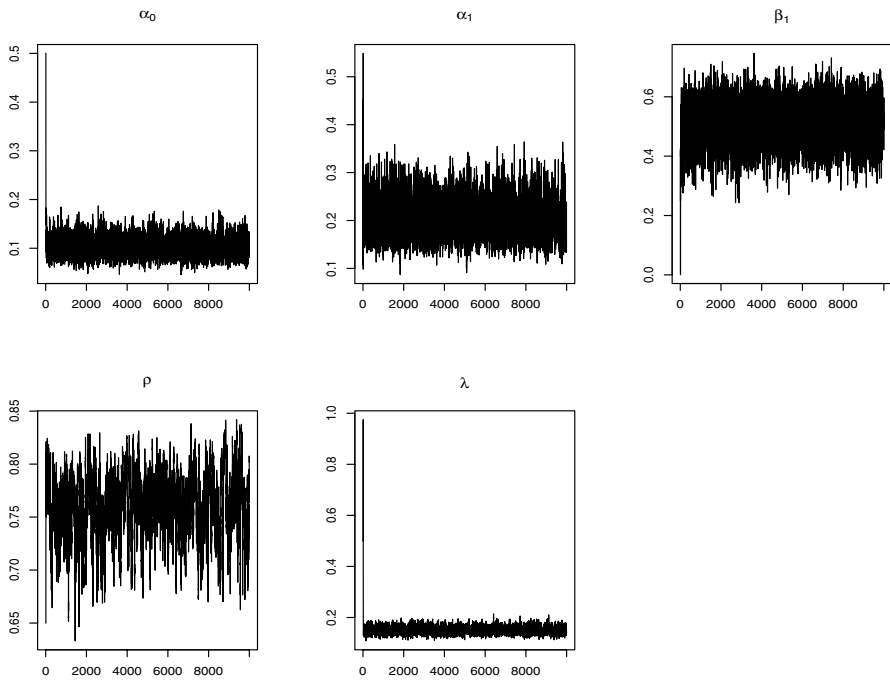


Fig. 13 The trace plots of parameters for s3

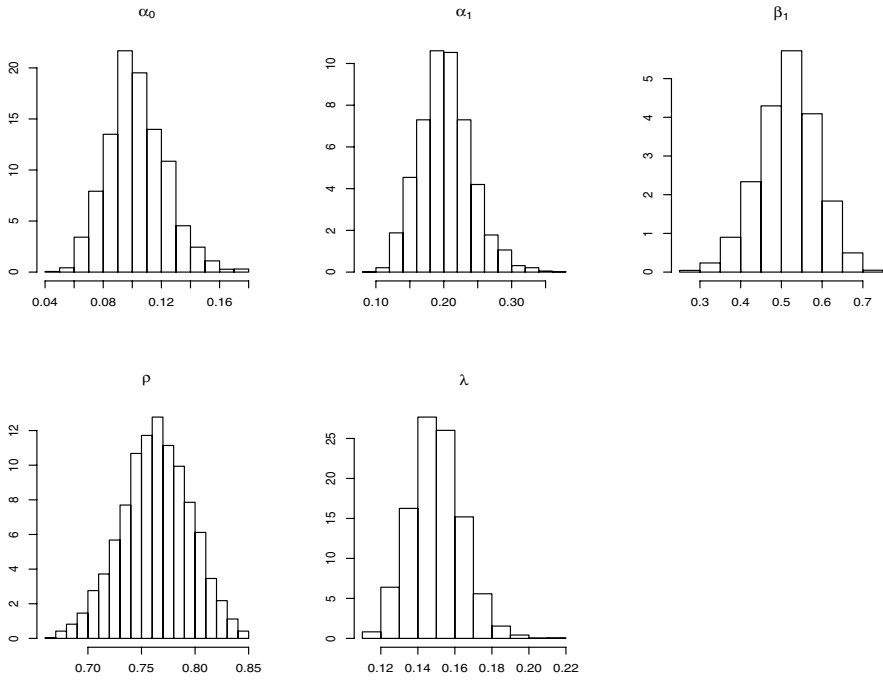


Fig. 14 The histograms of parameters for s

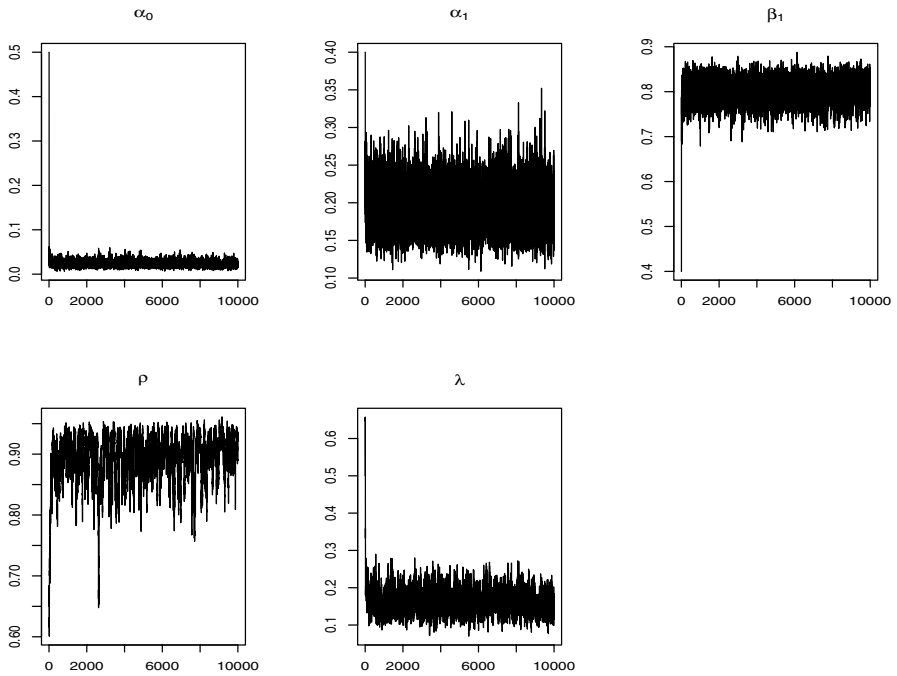


Fig. 15 The trace plots of parameters for SP 500

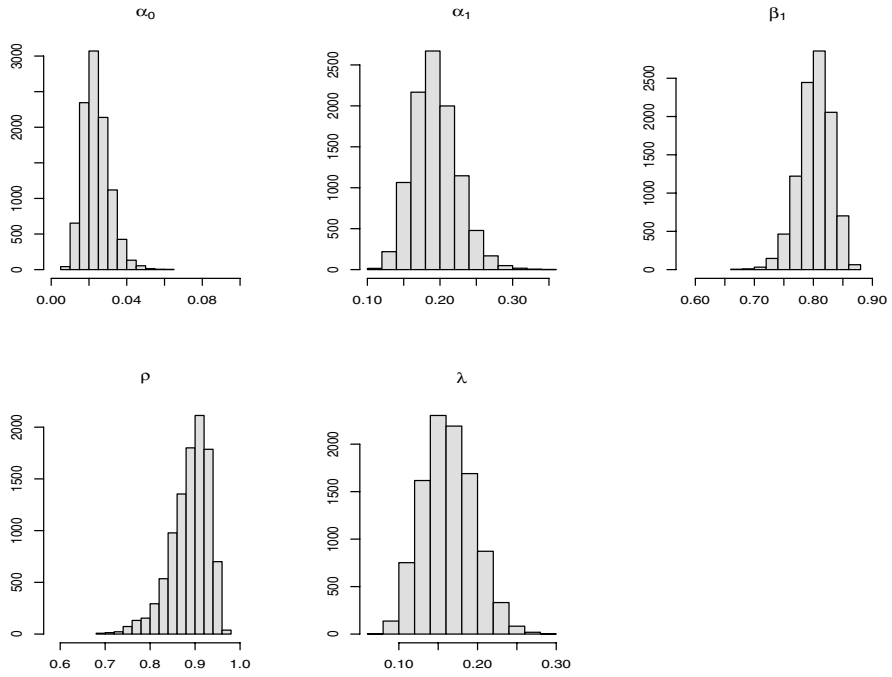


Fig. 16 The histograms of parameters for SP 500

Table 8 Convergence diagnosis by Geweke test for each parameter

Parameter	α_0	α_1	β_1	ρ	λ
s1	0.95	-0.16	0.07	-0.45	0.58
s2	2.42	0.86	-2.12	-0.59	0.17
s3	1.03	1.22	-1.77	-0.95	1.25

Table 9 Statistics summary of three simulation series

Series	Mean	Variance	Skewness	Kurtosis
s1	-0.0184	0.3556	0.4520	6.2986
s2	0.0161	0.3007	0.2480	7.2510
s3	-0.001	0.3614	-0.698	7.6601

Table 10 Posterior means for the simulated s1, s2 and s3 using HMC sampler

Parameter	α_0	α_1	β_1	ρ	λ	Time
	0.1	0.2	0.5	0.8	0.15	
s1	0.1265 (0.0335)	0.2528 (0.0720)	0.4396 (0.0891)	0.8276 (0.0386)	0.1197 (0.0228)	10.61 mins
s2	0.1217 (0.0318)	0.1820 (0.0513)	0.4238 (0.1161)	0.7896 (0.0414)	0.1584 (0.0235)	12.94 mins
s3	0.1034 (0.0228)	0.2032 (0.0390)	0.5158 (0.0716)	0.7615 (0.0320)	0.1507 (0.0222)	20.07 mins

parameters to obtain best performance in the HMC algorithm. Meanwhile, the extension of the multiple mixture distribution is a challenging problem, it is not easy to identify mixture components. These issues are still worthy of further study and explore.

Appendix

See Figures 9, 10, 11, 12, 13, 14, 15 and 16, Tables 8, 9 and 10.

Acknowledgements The research of Qiang Xia was supported by the National Natural Science Foundation of China (No.12171161, 91746102), the Natural Science Foundation of Guangdong Province of China (No. 2022A1515011754), and Ministry of Education in China Project of Humanities and Social Sciences (No.17YJA910002).

Funding The authors have not disclosed any funding.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

References

- Alexander, C., & Lazar, E. (2006). Normal mixture garch(1,1): Applications to exchange rate modelling. *Journal of Applied Econometrics*, 21(3), 307–336.
- Ausín, M. C., & Galeano, P. (2007). Bayesian estimation of the gaussian mixture garch model. *Computational Statistics & Data Analysis*, 51(5), 2636–2652.
- Bai, X., Russell, J. R., & Tiao, G. C. (2003). Kurtosis of garch and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, 114, 349–360.
- Bauwens, L., Bos, C. S. & Van Dijk, H. K. (1999) . Adaptive polar sampling with an application to a bayes measure of value-at-risk, Working paper, CORE, Universite Catholique de Louvain .
- Bauwens, L., & Lubrano, M. (1998). Bayesian inference on garch models using gibbs sampler. *Econometrics Journal*, 1(1), 23–46.
- Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J. M., & Stuart, A. (2013). Optimal tuning of the hybrid monte carlo algorithm. *Bernoulli*, 19(5A), 1501–1534.
- Betancourt, M. (2011) . Nested sampling with constrained hamiltonian monte carlo, AIP Conference Proceedings, Vol. 1305, American Institute of Physics, pp. 165–172.

- Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo, arXiv preprint [arXiv: 1701.02434](https://arxiv.org/abs/1701.02434).
- Bollerslev, T. (1986). Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31(3), 307–327.
- Bollerslev, T. (1987). A conditionally heteroscedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69(3), 524–547.
- Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, 11(2), 143–172.
- Burda, M., & Bélisle, L. (2019). Copula multivariate garch model with constrained hamiltonian monte carlo. *Dependence Modeling*, 7(1), 133–149.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., et al. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32.
- Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society: Series B*, 56(2), 363–375.
- Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B*, 195, 216–222.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50, 987–1008.
- Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1), 1–50.
- Fama, E. F. (1965). The behaviour of stock market prices. *Journal of Business*, 38(1), 34–105.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian Statistics*, 4, 641–649.
- Geweke, J. (1994). Bayesian comparison of econometric models, Working Paper, Federal Reserve Bank of Minneapolis (35).
- Girolami, M., & Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2), 123–214.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5), 1779–1801.
- Haas, M., Mittnik, S., & Paolella, M. S. (2004). Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics*, 2, 211–250.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1), 97–109.
- Hoffman, M., & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15, 1593–1623.
- Jorion, P. (2000). Value at Risk: The New Benchmark for Managing Financial Risk(second ed.).
- Kreuzer, A., & Czado, C. (2021). Bayesian inference for a single factor copula stochastic volatility model using hamiltonian monte carlo. *Econometrics and Statistics*, 19, 130–150.
- Leimkuhler, B. & Reich, S. (2004). *Simulating hamiltonian Dynamics*, Cambridge University Press.
- Mandelbrot, B. (1963). New method in statistical economics. *Journal of Political Economy*, 71(5), 421–440.
- McFarland, J. W., Pettit, R. R., & Sung, S. K. (1982). The distribution of foreign exchange price changes: Trading day effects and risk measurement. *Journal of Finance*, 37(3), 693–715.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6), 1087–1092.
- Neal, R. M. (2011). MCMC Using Hamiltonian Dynamics, Handbook of Markov Chain Monte Carlo(S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds.) CRC Press, New York.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*: Journal of the Econometric Society pp. 347–370.
- Paixão, R. S. & Ehlers, R. S. (2017). Zero variance and hamiltonian monte carlo methods in garch models, arXiv preprint [arXiv: 1710.07693](https://arxiv.org/abs/1710.07693).
- Pan, J., Wang, H., & Tong, H. (2008). Estimation and tests for power-transformed and threshold garch models. *Journal of Econometrics*, 142, 352–378.
- Plummer, M., Best, N., Cowles, K. & Vines, K. (2008). Coda: Output analysis and convergence diagnosis for mcmc. <https://CRAN.R-project.org/package=coda>

- Ritter, C., & Tanner, M. A. (1992). Facilitating the gibbs sampler: The gibbs stopper and the griddy-gibbs sampler. *Journal of the American Statistical Association*, 87, 861–868.
- Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), 528–540.
- Westerfield, R. (1977). The distribution of common stock price changes: An application of transactions time and subordinated stochastic models, *Journal of Financial and Quantitative Analysis* pp. 743–765.
- Wong, C. S., & Li, W. K. (2001). On a mixture autoregressive conditional heteroscedastic model. *Journal of the American Statistical Association*, 96(455), 982–995.
- Xia, Q., Wong, H., Liu, J., & Liang, R. (2017). Bayesian analysis of power-transformed and threshold garch models: A griddy-gibbs sampler approach. *Computational Economics*, 50(3), 353–372.
- Yu, B., & Mykland, P. (1994). *Looking at markov samplers through cumsum path plots: A simple diagnostic idea*, Technical Report 413. Department of Statistics: University of California at Berkeley.
- Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5), 931–955.
- Zhang, Z., Li, W. K., & Yuen, K. C. (2006). On a mixture garch time-series model. *Journal of Time Series Analysis*, 27(4), 577–597.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.