WILEY

Estimating expected shortfall using a quantile function model

Yuzhi Cai

School of Management, Swansea University, Swansea, UK

Correspondence
Yuzhi Cai, School of Management, Swansea University, Swansea SA1 8EN, UK.
Email: y.cai@swansea.ac.uk

Abstract

Distribution of financial returns defined by the existing GARCH models usually focus on the overall features such as the location, scale, skewness and kurtosis of the distribution. When using such GARCH models for expected shortfall (ES) estimation, it is difficult to consider specific information about the tails (such as the shape of the tails of the distribution), resulting in possible bias in ES estimation. We propose a quantile function threshold GARCH model to overcome some of the limitations of existing models. The model allows us to use the information including the skewness and tail shape of the distribution and the structure changes in the volatility of financial returns to obtain ES estimates. Our results show that the proposed model outperforms the benchmark models, confirming that tail shape of the distribution also plays an important role in ES estimation.

KEYWORDS

expected shortfall, financial returns, generalized Lambda distribution, quantile function, tail shape of distribution

1 | INTRODUCTION

There are two commonly used risk measures in finance: Value-at-Risk (VaR) and expected shortfall (ES). By definition, VaR measures the conditional quantile of the distribution of financial returns, while ES measures the conditional expectation of loss given that the loss is beyond the VaR level. Both measures have some limitations. For example, VaR ignores any loss beyond the VaR level, it is not a coherence risk measure; it may provide mislead information to rational investors who wish to maximize expected utility, and it is hard to use when investors want to optimize their portfolios, etc.; while ES needs very large history to see extreme events and it is hard to be back-tested by using actual breaches etc. See, for example, Artzner et al. (1997,1999), Acerbi and

Tasche (2002) and Yamai and Yoshiba (2002) and references therein.

Since the work of Artzner et al. (1997,1999), ES has become a more and more popular financial risk measure and, like VaR, it has also been used widely nowadays in practice. In this paper, we focus on ES. Various estimation methods for conditional ES can be found in the literature. Recently, Nadarajah et al. (2014) gave an excellent review on the estimation methods for ES, in which more than 150 references were discussed.

The majority of the estimation methods for ES are parametric in the sense that they are estimated by using a parametric model that defines the conditional distribution of returns. For example, Standard GARCH (denoted by S-GARCH) models of Engle (1982) and Bollerslev (1986) are not only popular for estimating volatility but also often

used for studying VaR of returns, see, for example, Poon and Granger (2003). ES can then be calculated easily from the distributions defined by the S-GARCH models.

Recently, in addition to the normal distribution, other standard distributions have also been used for the innovation term of the S-GARCH models in order to capture various important features of financial data, which include the skewed t-distribution, normal-inverse Gaussian distribution (NIG), skewed generalized error distribution (GED), generalized hyperbolic distribution (GHYP) and Johnson's SU distribution (JSU), see, for example, Socgnia and Wilcox (2014), Necula (2009), Önalan (2010), Theodossiou (2015), Corlu and Corlu (2015) and references therein. For convenience, we call the skewed t-, NIG, skewed GED, GHYP and JSU distributions as standard distributions in order to distinguish them from the distribution that we will introduce later in the paper.

S-GARCH models have some limitations. For example, they assume that only the magnitude and not the positivity or negativity of unanticipated excess returns affect the volatilities of financial returns. To overcome the limitations of S-GARCH models, alternative models such as threshold GARCH (T-GARCH) models have been developed in the literature. For example, Glosten et al. (1993) proposed a T-GARCH model, the so-called GJR-GARCH model, Zakoian (1994) also proposed a threshold GARCH model, denoted by T-GARCH, and Nelson (1991) developed the exponential GARCH model, denoted by E-GARCH. More recently, Yang and Chang (2008) considered a double-threshold GARCH model and Yu et al. (2010) extended the CAViaR idea (Engle and Manganelli, 2004) to T-GARCH and mixture-GARCH models in order to take into account possible nonlinearity and structural changes in the VaR process. These T-GARCH models can also be defined by using the standard distributions. An excellent review on threshold time series models in finance was given by Chen et al. (2011).

It is worth noting that GARCH models (including S-GARCH and T-GARCH) were originally developed for the conditional volatility of financial returns, which measures the stability of financial returns with respect to the average returns. The distribution of returns defined by GARCH models often mainly focuses on the overall features such as location, scale, skewness and kurtosis of the distribution. When using GARCH models for ES estimations, specific information about the tails such as tail shape of the distribution cannot be taken into account, leading to possible bias in ES estimation. Hence, we may ask: is it safe to use such GARCH models for the estimation of conditional ES? An answer to this question is obviously very important for financial risk management, but to our best knowledge, there exists a gap in the literature because not much work can be found in the literature to address this research question.

Our main contribution to the literature is that we propose a novel quantile function model in order to address this research question. We illustrate our approach by focusing on the T-GARCH model of Yu et al. (2010) because it gives a more general formulation for T-GARCH models, it includes the S-GARCH models as a special case, and it also allows us to take account of structural changes in the volatility of financial returns. One of the advantages of the proposed model is that it allows us to obtain ES estimates by using the information not only on the overall features but also on the tail shape of the distribution, which provides a means for filling the gap in the literature.

On the other hand, for model estimations, apart from the maximum likelihood estimation (MLE) method, Bayesian approach for GARCH models has often been used recently in the literature. For example, Bauwens and Lubrano (2002) studied option pricing with asymmetric GARCH models by using a Bayesian method; Ausín and Galeano (2007) considered a Bayesian estimation of the Gaussian mixture-GARCH models; a Bayesian method for estimating volatility asymmetries with a class of tree structured multivariate GARCH models was proposed by Dellaportas and Vrontos (2007); and Jensen and Maheu (2013) proposed a Bayesian approach to semiparametric multivariate GARCH modelling. More recently, Virbickaite et al. (2015) gave an excellent review on Bayesian inference methods for GARCH models. It is worth noting that the above-mentioned Bayesian estimation methods are generally based on the likelihood functions defined by density functions.

For our proposed model, two main difficulties are involved in the model estimation. One is that the delay parameter and multiple thresholds also need to be estimated, and the other is that the likelihood function of the model parameters is defined by quantile functions. It is worth noting that estimating a T-GARCH model with multiple thresholds and a delay parameter is difficult if the MLE method is used. Hence, in this paper, we use a Bayesian approach and propose a MCMC estimation method for parameter estimation in order to facilitate the use of the proposed model for ES estimation. As the likelihood function of the model parameters is defined by quantile functions, our MCMC algorithm also illustrates how a quantile function model could be estimated by using a Bayesian approach, which could be viewed as our another contribution to the literature.

We further conducted an empirical study on six financial return series in this paper, which allows us to gain some insight into the performance of our model and other models regarding the estimation of ES. In fact, our results show that the proposed model outperforms all benchmark models considered in this paper, which confirm that the

information about the tail shape of the conditional distribution of returns also plays a crucial role in the estimation of ES. Hence, such information should not be ignored when estimating ES of financial returns.

The paper is organized as follows. Section 2 introduces our quantile function T-GARCH model and Section 3 discusses the estimation method. Section 4 provides some details about the benchmark models and evaluation methods for ES estimates. Results on the empirical study are given in Section 5. Finally, some concluding discussions are given in Section 6. Moreover, proofs of Propositions are given in Appendices A–D, G, and H, the prior density functions and detailed MCMC algorithm are given in Appendices 5 and 6, and some tables about the BIC values of the estimated models, the results about the coverage probabilities of the VaR estimates and the evaluation results about the ES estimates are given in Appendix I.

2 | QUANTILE FUNCTION T-GARCH MODEL

Consider the general -regime T-GARCH model

$$
\begin{aligned}
y_t &= \varepsilon_t \sqrt{h_t}, \\
h_t &= \sum_{j=1}^{J} \left(\alpha_{0j} + \sum_{i=1}^{p_j} \alpha_{ij} y_{t-i}^2 + \sum_{\ell=1}^{q_j} \beta_{\ell j} h_{t-\ell} \right) I\left(\gamma_{j-1} \leq y_{t-d} < \gamma_j \right),
\end{aligned}
$$

(1)

where $I(\cdot)$ is the indicator function, the regime number J and the delay parameter d are positive integers, the γ_j's are real numbers (thresholds) such that $-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_{J-1} < \gamma_J = \infty$, and $p_j \geq 0$ and $q_j \geq 0$ define the order of the model. Moreover, $\alpha_{0j} > 0$, $\alpha_i^{(j)} \geq 0$, $\beta_{\ell j} \geq 0$. The ε_t's are assumed to be i.i.d. with mean 0 and variance 1. Liu et al. (1997) proved that, under some regularity conditions, there exist stationary and ergodic solutions satisfying model (1). It is worth noting that in this paper we assume the delay parameter $d \in D = \{1, ..., d_0\}$, where, by following Yu et al. (2010), we let $d_0 = 3$.

As discussed in Section 1, if the innovation term ε_t follows a standard distribution, then the conditional ES estimates obtained from Model (1) may be biased. Hence, our approach here is to identify a new distribution for the innovation term so that not only the overall shape of the distribution but also specific features about the tails can be taken into account.

Freimer et al. (1988) and Fournier et al. (2007) discussed the generalized Lambda distributions, denoted by GLD, that can be a good candidate for our model because GLD can accurately approximate many commonly used distributions such as normal, log-normal, Weibull, t-, F- and skewed t-distributions as well as many others. Moreover, it contains two parameters that control the skewness of the distribution and that also determine the relative weights of the tails. In other words, the skewness of the distribution is modelled as a result of tail shape and not as an independent feature of the distribution (Gilchrist, 2000). Therefore, a model defined by the GLD is more robust to model specification errors and it is certainly worth considering this distribution for the estimation of ES of financial returns.

It is worth noting that the GLD has been used in the finance literature. For example, Corrado (2001) used the GLD for option pricing, Lee (2003) showed that the GLD provides a good model for spot exchange rates, Tarsitano (2004) used the GLD to model income data, and Corlu and Corlu (2015) studied the performance of GLD in capturing the leptokurtic and skewed behaviour of exchange rate returns. To our best knowledge, the GLD has not been used to define a T-GARCH model for the estimation of ES in the finance literature.

Hence, to make use of the GLD for our model, we first let $\varepsilon_t = \{e_t - \mu(\theta)\}/\sigma(\theta)$, where e_t follows the GLD with mean $\mu(\theta)$ and variance $\sigma(\theta)^2$, and θ is a vector of parameters of e_t. It is seen that ε_t is a standardized GLD. Due to the general popularity of order (1, 1) for GARCH models in finance, we let $p_j = q_j = 1$ in the rest of this paper. Then in this case, model (1) becomes

$$
\begin{aligned}
y_t &= \sqrt{h_t}\varepsilon_t = \sqrt{h_t}\{(e_t - \mu(\theta))/\sigma(\theta)\}, \\
h_t &= \sum_{j=1}^{J} \left(\alpha_{0j} + \alpha_{1j} y_{t-1}^2 + \beta_{1j} h_{t-1} \right) I\left(\gamma_{j-1} \leq y_{t-d} < \gamma_j \right).
\end{aligned}
$$

(2)

Model (2) says that the conditional distribution of y_t is a scaled standardized GLD and the conditional volatility of y_t is again defined by $\sqrt{h_t}$.

It is worth noting that the GLD is defined by a quantile function, which can be expressed by $Q_{gld}(\tau, \theta) = \mu_{gld} + s_{gld}Q(\tau, \eta)$, where $\tau \in (0, 1)$,

$$
Q(\tau, \eta) = \frac{\tau^{\eta_1} - 1}{\eta_1} - \frac{(1-\tau)^{\eta_2} - 1}{\eta_2}, \quad \eta_1 < 0, \eta_2 < 0, \quad (3)
$$

and $\theta = (\mu_{gld}, s_{gld}, \eta)$ is the parameter vector of the GLD, in which $\eta = (\eta_1, \eta_2)$, μ_{gld} is the location of the GLD and s_{gld} is the scale of the GLD. Hence, function (3) defines the GLD with location 0 and scale 1. Moreover, η controls the skewness of the distribution: if $\eta_1 = \eta_2$ then the distribution is symmetric, otherwise it is skewed. Furthermore, $1/(-\eta_1)$ and $1/(-\eta_2)$ also determine the relative weights of left and right tails respectively. Hence, both skewness and tail shape of the distribution are controlled by η.

As the GLD is used for the innovation term of our model, it is reasonable to consider the GLD with location $\mu_{gld} = 0$ and scale $s_{gld} = 1$ for e_t. Hence, we let e_t follows the GLD defined by $Q(\tau, \boldsymbol{\eta})$ given by (3). For model (2) to be well defined, we need to show that e_t has finite mean and variance, which can be seen from the following result.

Proposition 1 *Suppose e_t follows the GLD defined by (3). Then*

$$
\begin{aligned}
\mu(\boldsymbol{\eta}) &= E(e_t) = \frac{1}{\eta_2 + 1} - \frac{1}{\eta_1 + 1}, \\
\sigma(\boldsymbol{\eta})^2 &= E\left(e_t^2\right) - \mu(\boldsymbol{\eta})^2 \\
&= \frac{1}{\eta_1^2}\left(\frac{1}{2\eta_1 + 1} - \frac{1}{(\eta_1 + 1)^2}\right) + \frac{1}{\eta_2^2}\left(\frac{1}{2\eta_2 + 1} - \frac{1}{(\eta_2 + 1)^2}\right) \\
&\quad - \frac{2}{\eta_1\eta_2}\left(B(\eta_1, \eta_2) - \frac{1}{(\eta_1 + 1)(\eta_2 + 1)}\right),
\end{aligned}
\tag{4}
$$

in which $B(\eta_1, \eta_2) = \int_0^1 \tau^{\eta_1}(1 - \tau)^{\eta_2} d\tau$. Moreover, let $\Omega_1 = [-1/2 + m_0, -m_0]$, $\Omega_2 = [-1 + m_0, -1/2 - m_0]$ and $\Omega_3 = [-M, -1 - m_0]$, where M is a fixed very large positive number and m_0 is a fixed very small positive number, then both $\mu(\boldsymbol{\eta})$ and $\sigma(\boldsymbol{\eta})$ are finite on $\cup_{i,j=1}^3 (\Omega_i \times \Omega_j)$.

See Appendix A for a proof. Note that we let $M = 10^{20}$ and $m_0 = 10^{-20}$ in this paper to make the parameter space as large as possible and that $B(\eta_1, \eta_2)$ can be calculated numerically. As both $\mu(\boldsymbol{\eta})$ and $\sigma(\boldsymbol{\eta})$ are finite on $\cup_{i,j=1}^3 (\Omega_i \times \Omega_j)$, model (2) is well defined. Furthermore, we have the following result.

Proposition 2 *Suppose y_t follows model (2), where e_t follows the GLD defined by (3). Let $Q_{y_t}(\tau|\mathbf{y}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\eta}, \gamma_1, d)$ be the conditional quantile function of y_t and let $Q_{\varepsilon_t}(\tau, \boldsymbol{\eta})$ be the quantile function of ε_t. Then we have*

$$
\begin{aligned}
Q_{y_t}(\tau|\mathbf{y}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\eta}, \gamma_1, d) &= \sqrt{h_t} Q_{\varepsilon_t}(\tau, \boldsymbol{\eta}), \\
h_t &= \sum_{j=1}^J \left(\alpha_{0j} + \alpha_{1j}y_{t-1}^2 + \beta_{1j}h_{t-1}\right)I\left(\gamma_{j-1} \le y_{t-d} < \gamma_j\right),
\end{aligned}
\tag{5}
$$

where $Q_{\varepsilon_t}(\tau, \boldsymbol{\eta}) = \{Q(\tau, \boldsymbol{\eta}) - \mu(\boldsymbol{\eta})\}/\sigma(\boldsymbol{\eta})$, $\boldsymbol{\beta} = (\alpha_{0j}, \alpha_{1j}, \beta_{1j}, j = 1, \dots, J)$ and $\mathbf{y}_t = (y_t, \dots, y_1)$.

See Appendix B for a proof. Hence, it follows from Proposition 2 that Model (5) can be expressed by

$$
\begin{aligned}
Q_{y_t}(\tau|\mathbf{y}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\eta}, \gamma, d) &= \sqrt{h_t}\left\{\frac{Q(\tau, \boldsymbol{\eta}) - \mu(\boldsymbol{\eta})}{\sigma(\boldsymbol{\eta})}\right\}, \\
h_t &= \sum_{j=1}^J \left(\alpha_{0j} + \alpha_{1j}y_{t-1}^2 + \beta_{1j}h_{t-1}\right)I\left(\gamma_{j-1} \le y_{t-d} < \gamma_j\right),
\end{aligned}
\tag{6}
$$

which is well defined on the parameter space $\bar{\Omega} = \{\alpha_{ij}, \beta_{1j}, \gamma_j, \eta_v, d \mid 0 < \alpha_{0j} < M, 0 \le \alpha_{1j}, \beta_{1j} \le a < 1, 0 \le \alpha_{1j} + \beta_{1j} \le a < 1, \gamma_j \in R, \boldsymbol{\eta} \in \cup_{\ell_1, \ell_2 = 1}^3 (\Omega_{\ell_1} \times \Omega_{\ell_2}), i = 0, 1, j = 1, \dots, J - 1, d \in D\}$. Model (6) says that conditional distribution of y_t is a scaled standardized GLD with mean 0 and variance h_t, and the skewness and tail shape of the conditional distribution of y_t are jointly controlled by η_1 and η_2. Moreover, given model (6), the ES can be estimated by using Proposition 3 below.

Proposition 3 *Suppose y_t follows Model (6). Then the conditional expected shortfall of y_t at level τ, denoted by $ES_{t\tau}$, is given by:*

$$
ES_{t\tau} = \begin{cases}
\dfrac{\sqrt{h_t}}{\tau\sigma(\boldsymbol{\eta})}\left\{\dfrac{\tau^{\eta_1+1} - \tau}{\eta_1(\eta_1+1)} + \dfrac{(1-\tau)^{\eta_2+1} - (1-\tau)}{\eta_2(\eta_2+1)}\right\} & if\ \tau \le 0.5 \\[3mm]
\dfrac{\sqrt{h_t}}{(1-\tau)\sigma(\boldsymbol{\eta})}\left\{\dfrac{\tau - \tau^{\eta_1+1}}{\eta_1(\eta_1+1)} + \dfrac{(1-\tau) - (1-\tau)^{\eta_2+1}}{\eta_2(\eta_2+1)}\right\} & if\ \tau > 0.5.
\end{cases}
\tag{7}
$$

See Appendix C for a proof.

Clearly, to obtain the conditional ES estimates, we first need to know how to estimate the model parameters, which we discuss in the next section.

3 | ESTIMATION METHOD

3.1 | Likelihood function

Various methods have been developed in the literature in order to estimate GLD models. For example, the method of moment was proposed by Karian et al. (1996), the method of least squares was discussed by Ozturk and Dale (1985), the method of percentiles was developed by Karian and Dudewicz (1999), a starship estimation method was proposed by King and MacGillivray (1999), the MLE method was discussed by Gilchrist (2000) and a two-step procedure that combines the method of moment or percentile and the MLE to fit the GLD to data was developed by Su (2007a). Su (2007b) also pointed out that the MLE method not only is more efficient but also tends to

produce GLD that has closer first four moments to the data analysed.

For our model, to write out the likelihood function of the model parameters, we first need to note that the relationship between the density function $f(y)$ of a random variable Y and its quantile function $y = Q(\tau)$ is given by $f(y) = (d(Q(\tau))/d\tau)^{-1}$. Moreover, it follows from model (6) that, for each y_t there exists $\tau_t \sim U(0, 1)$ such that $y_t = Q(\tau_t, \boldsymbol{\eta})\sqrt{h_t}$. Then, the likelihood of the observed returns $\mathbf{y}_T$, conditional on the initial observations $\mathbf{y}_{d_0}$, is given by

$$\pi(\mathbf{y}_T|\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},d) = \prod_{t=d_0+1}^{T} \left(\frac{\sqrt{h_t} Q'(\tau_t, \boldsymbol{\eta})}{\sigma(\boldsymbol{\eta})} \right)^{-1}$$

$$= \prod_{t=d_0+1}^{T} \frac{\sigma(\boldsymbol{\eta})}{\sqrt{h_t}\left\{ \tau_t^{\eta_1-1} + (1-\tau_t)^{\eta_2-1} \right\}}$$

where $\sigma(\boldsymbol{\eta})$ is given by (4).

However, it is very difficult to use the above estimation methods for the proposed model because of the thresholds and the delay parameter involved in the volatility process. Hence, we propose a MCMC estimation method for the parameters which requires us to derive the posterior density function of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ and d.

3.2 | Posterior distribution

It follows from the Bayesian theorem that the posterior density function of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ and d is proportional to the product of the likelihood of observed returns and the prior density function of the parameters. Let $\pi(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, d| \mathbf{y}_T)$ and $\pi(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\gamma}, d)$ be the posterior and prior density functions respectively. Then, conditional on $\mathbf{y}_{d_0}$, we have

$$\pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},d|\mathbf{y}_T) \propto \pi(\mathbf{y}_T|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},d)\pi(\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},d)$$

$$= \prod_{t=d_0+1}^{T} \frac{\sigma(\boldsymbol{\eta})}{\sqrt{h_t}\left\{ \tau_t^{\eta_1-1} + (1-\tau_t)^{\eta_2-1} \right\}} \pi(\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},d). \quad (8)$$

Proposition 4 *Let* $\Omega = \left\{ \alpha_{0j}, \alpha_{1j}, \beta_{1j}, \gamma_j, \eta_v, d \mid m_0 \leq \alpha_{0j} < M, \right.$ $0 \leq \alpha_{1j}, \beta_{1j} \leq a < 1, 0 \leq \alpha_{1j} + \beta_{1j} \leq a < 1, \quad \gamma_j \in R, \boldsymbol{\eta} \in \cup_{\ell_1,\ell_2=1}^{3} (\Omega_{\ell_1} \times \Omega_{\ell_2}), i = 0, 1, j = 1, \ldots, J-1, d \in D \right\}$. *If* $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)$ *is a well-defined density function on* Ω, *i.e., the integral of* $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)$ *on* Ω *is finite, then the posterior density function (8) is also well defined on* Ω.

See Appendix D for a proof. It is seen that the difference between $\bar{\Omega}$ and Ω is in the range of values that α_{0j} can take. As m_0 is very small, this difference can be safely ignored from a practical point of view, but the proof in Appendix D shows that it guarantees that the posterior distribution is well defined on Ω, which is important for the parameter estimations using a MCMC method.

3.3 | Prior distribution

In practice, it is reasonable to assume that the thresholds $\gamma_j \in (y_{\min}, y_{\max})$ for all possible j, where $y_{\min} = \min\{y_1, \ldots, y_T\}$ and $y_{\max} = \max\{y_1, \ldots, y_T\}$. To simplify the calculations, we let the prior density function $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\eta})\pi(d)$, where γ_j is uniformly distributed on $(y_{\min}, y_{\max})$, d is uniformly distributed on D, and α_{ij}, β_{1j} and $-\eta_v$ follow a log-normal distribution truncated on the parameter space Ω. Clearly, the prior distribution is well defined on the parameter space of the posterior distribution.

The detailed formulas for the prior density function are given in Appendix E, where the σ_{ij}'s, s_{1j}'s and λ_v's are the corresponding scale parameters, which measure the strength of the prior information: large (small) values represent weak (strong) prior information used in the parameter estimation.

In this paper, we deliberately let $\sigma_{ij} = s_{1j} = \lambda_v = 2$. This means that, for example, the *standard deviation (SD)* of α_{ij} is given by $\sqrt{e^{\sigma_{ij}^2}\left(e^{\sigma_{ij}^2}-1\right)} = 54.1$, which is very large. Hence, the results obtained in this paper are robust to the choice of $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)$.

3.4 | MCMC algorithm

One of the basic MCMC methods is the Metropolis-Hastings algorithm (see for example, Gamerman and Lopes, 2002, and Geyer, 2011). In this method, we first generate a candidate parameter value from a chosen distribution and then we accept this proposed value as the next parameter value with a known probability. By repeating the two steps multiple times, we produce a sequence of parameter values that form a Markov chain in the parameter space Ω, whose equilibrium distribution is the posterior distribution of the parameters; see for example, Brooks (1998) and O'Hagan and Forster (2004).

For our method, we let $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ and d be the current parameter values, and $\boldsymbol{\beta}'$, $\boldsymbol{\gamma}'$, $\boldsymbol{\eta}'$ and d' the proposed values. Appendix F shows the detailed steps of our MCMC

algorithm. Hence, after a burn-in period, we may collect posterior samples of parameters from the Markov chain generated by our MCMC algorithm, based on which the model parameters can be estimated, and hence, conditional ES estimates at a given level can be obtained.

It is seen from Appendix F that our estimation method can handle multiple thresholds and the delay parameter easily and has no difficulties in handling the complex structure of the parameter space. However, it is worth noting that, in the MLE method, a grid search procedure for the threshold and delay parameter usually needs to be used, which becomes impractical when the number of thresholds is greater than one.

4 │ BENCHMARK MODELS AND EVALUATION OF ES ESTIMATES

4.1 │ Benchmark models and ES estimations

There are many models in the literature that have been used to estimate ES, see Nadarajah et al. (2014). For the purpose of this study, we selected our benchmark models as follows.

First, note that our model is a parametric model that is related to a GARCH model. Hence, our main benchmark models are also chosen to be parametric. Specifically, we use S-GARCH, GJR-GARCH, E-GARCH and T-GARCH models as our benchmark models, for which the models of order (1, 1) are defined by $y_t = \sigma_t \varepsilon_t$, where σ_t is given below:

S-GARCH(1,1) model: $\sigma_t^2 = \omega + \alpha_j \varepsilon_{t-1}^2 + \beta_j \sigma_{t-1}^2$.

E-GARCH(1,1) model: $\ln(\sigma_t^2) = \omega + \alpha_1 \varepsilon_{t-1} + \gamma_1(|\varepsilon_t - 1| - E|\varepsilon_{t-1}|) + \beta_1 \ln(\sigma_{t-1}^2)$.

GJR-GARCH(1,1) model: $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \gamma_1 I_{t-1} \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, where $I_{t-1} = 1$ if $\varepsilon_{t-1} \leq 0$ and $I_{t-1} = 0$ otherwise.

T-GARCH(1, 1)) model: $\sigma_t = \omega + \alpha_1(|\varepsilon_{t-1}| - \eta_1 \varepsilon_{t-1}) + \beta_1 \sigma_{t-1}$.

Moreover, the innovation term of each GARCH model is assumed to follow any one of the following standard distributions, that is, the skewed t-, NIG, skewed GED, GHYP and JSU distributions, resulting in 20 benchmark GARCH models in total. These distributions were chosen because they are able to capture much more complicated data structures of financial returns and have become more and more popular distributions in the finance literature. It is seen that, apart from the S-GARCH models, all other models are able to deal with the asymmetric impact of positive and negative returns.

As these benchmark GARCH models are parametric, a detailed formula for ES estimation can be derived for each model. However, rather than doing this for each model, we give a general formula for ES in a quantile function form in Proposition 5 for all benchmark GARCH models.

Proposition 5 *For the above benchmark GARCH models, the conditional expected shortfall at the τ_0th level is given by*

$$ES_{t\tau_0} = \begin{cases} \dfrac{\sigma_t}{\tau_0} \displaystyle\int_0^{\tau_0} Q_{\varepsilon_t}(v)dv & \text{if } \tau_0 \leq 0.5 \\[3mm] \dfrac{\sigma_t}{1-\tau_0} \displaystyle\int_{\tau_0}^1 Q_{\varepsilon_t}(v)dv & \text{if } \tau_0 > 0.5, \end{cases}$$

where $Q_{\varepsilon_t}(v)$ is the quantile function of the innovation term of the models.

See Appendix G for a proof. Note that $ES_{t\tau_0}$ given in Proposition 5 can now be estimated by using the statistical software R.

It is worth noting that our proposed model is also a quantile-based model. Recently, Taylor (2008) proposed an exponentially weighted quantile regression method (denoted by E-QR) for VaR and ES. This method is also quantile-based but it is semi-parametric because it does not assume a distribution when estimating VaR. Hence we choose the E-QR model as our another benchmark model. For the E-QR model, the conditional VaR can also be estimated by using the statistical software R and the conditional ES estimates can then be obtained by using the formulas (13) and (14) in the paper of Taylor (2008).

Finally, we compare our ES estimates with those obtained from the empirical estimation method. This is the simplest non-parametric method, according to which, the ES can be estimated as the average of the observed returns that are smaller (or larger) than the empirical τ_0th quantile of the returns if $\tau_0 \leq 0.5$ (or $\tau_0 > 0.5$).

4.2 │ Evaluation of ES estimates

We follow the suggestions of McNeil and Frey (2000) to back-test our ES estimates. First, we define the standardized residuals, denoted by r_t. Specifically, for our model, we let $r_t = (y_t - \hat{ES}_{t\tau_0})/\sqrt{\hat{h}_t}$, and for the benchmark GARCH models, we let $r_t = (y_t - \hat{ES}_{t\tau_0})/\hat{\sigma}_t$, where $\hat{ES}_{t\tau_0}$ is the estimated conditional ES of y_t at level τ_0 and $\sqrt{\hat{h}_t}$ and $\hat{\sigma}_t$ are the estimated conditional volatilities from our model and the benchmark GARCH models respectively. Then, we have the following result.

Proposition 6 *For model (6), let $R_t = (y_t - ES_{t\tau_0})/\sqrt{h_t}$ and for the benchmark GARCH models, let $R_t = (y_t - ES_{t\tau_0})/\sigma_t$. Then under these models, the corresponding R_t are i.i.d. random variables and $E(R_t | y_t \leq VaR_{t\tau_0}) = 0$ if $\tau_0 \leq 0.5$ and $E(R_t | y_t \geq VaR_{t\tau_0}) = 0$ if $\tau_0 > 0.5$.*

See Appendix H for a proof. We further define the exceedance residuals by $\{r_{1t} = r_t; y_t \leq \hat{VaR}_{t\tau_0}, 1 \leq t \leq T\}$ if $\tau_0 \leq 0.5$, and $\{r_{2t} = r_t; y_t \geq \hat{VaR}_{t\tau_0}, 1 \leq t \leq T\}$ otherwise. Then, following the suggestions of McNeil and Frey (2000), we used the Ljung-Box test for zero autocorrelation in $\{r_t\}$, and the Bootstrap method (see, for example, Efron and Tibshirani, 1993) for the zero mean of $\{r_{jt}\}$, where $j = 1, 2$. A rejection on the null of either of the tests suggests that the models used for ES estimation may not be satisfactory, and hence the ES estimates could be biased.

For the E-QR model, we do not have the estimates for the volatility process. Hence, we follow the work of Taylor (2008) and calculate the standardized residuals by $r_t = (y_t - \hat{ES}_{t\tau_0})/\hat{VaR}_{t\tau_0}$, where $\hat{VaR}_{t\tau_0}$ is the estimated τ_0th conditional quantile of y_t. Then we again use the Bootstrap test for the exceedance residuals and the Ljung-Box test for the standardized residuals in order to evaluate the dynamic properties of the ES estimates.

For the empirical method, we define r_t as we did for the E-QR model because the sample *SD* can be seriously affected by extreme returns. Similarly, the Bootstrap and Ljung-Box tests are also used for zero mean and iid of the residuals respectively.

5 | EMPIRICAL STUDY

We conducted an empirical study on six daily stock returns, denoted by x_t. They are: the French CAC40 (2/1/2004–October 19, 2011) (see Yahoo Finance, 2019a), the German DAX30 (2/1/2004–2/11/2011) (see Yahoo Finance, 2019b), British FTSE100 (2/1/2004–November 28, 2011) (see Yahoo Finance, 2019c), Japanese Nikkei225 (2/1/2004–1/3/2012) (see Yahoo Finance, 2019d), the US Nasdaq (2/1/2004–8/12/2011) (see Yahoo Finance, 2019e) and the US S&P500 (2/1/2004–8/12/2011) (see Yahoo Finance, 2019f). Each sample period consists of 2000 daily returns and covers the financial crisis period between 2008 and 2009.
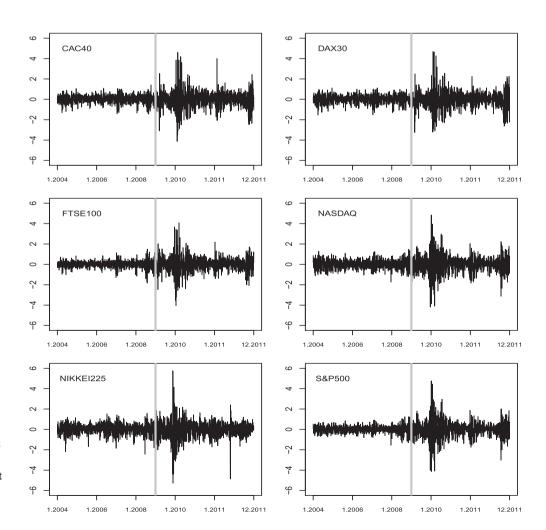

FIGURE 1 Time series plots of the data, where the first part of the data (before the lighter vertical line) in each plot corresponds to the first moving window of 1,000 days

Figure 1 shows the time series plots of the data and Table 1 provides some summary statistics about the data. It is seen that the average returns in the observation periods are slightly negative and standard deviations of the returns are similar. Moreover, all returns are only slightly skewed as all skewness statistics are between -0.5 and 0.5 or very close to them. However, the excess kurtosis for all returns is very high. Table 1 suggests that the overall shape of the distribution of returns is slightly skewed with heavy tails.

We adopted a moving window approach, suggested by Kuester et al. (2006), for model estimations. Specifically, we defined a moving window $W_t = \{x_t, ..., x_{t + 1000 - 1}\}$, where $t = 1, ..., 1000$. Hence, each moving window contains 1,000 observations. The first part of the data (before the lighter vertical line) in each plot of Figure 1 corresponds to the first moving window W_1 of 1,000 days. Note that we did not estimate the conditional mean of the returns by following the common practice (see Poon and Granger, 2003). Hence, we let $y_\ell = x_\ell - \bar{x}$, where $x_\ell \in W_t$ and $\bar{x}$ is the mean of the 1,000 in-sample returns

in W_t, and the parameter estimation methods were applied to y_ℓ.

Due to the popularity of T-GARCH models with two regimes in the finance literature, in this empirical study, we let $J = 2$ in model (6). Note that the benchmark threshold GARCH models also have two regimes. We checked BIC values of the benchmark GARCH models with various orders. We found that the minimum value of BIC always corresponds to a model of order (1, 1). See Tables 4-6 in Appendix I. Hence, all relevant models used in this empirical study have order (1, 1) for comparison purposes.

In this study, we consider the conditional ES estimates at the following six levels: 1%, 2.5%, 5%, 95%, 97.5% and 99%. Hence, for each return series and on each moving window, 20 benchmark GARCH models, 1 empirical model and 6 E-QR models (one for each level) were estimated, resulting in 162 estimated benchmark models on each moving window as we considered six return series in total. For illustration purposes, the estimated parameters of our models corresponding to W_1 and W_{350} are

TABLE 1 Summary statistics of the data

	CAC40	DAX30	FTSE100	NASDAQ	NIKKEI225	S&P500
Mean	−0.006	−0.005	−0.002	−0.007	−0.010	−0.007
SD	0.640	0.628	0.561	0.648	0.692	0.604
Skewness	0.037	0.016	−0.146	−0.204	−0.571	−0.303
Kurtosis	7.663	7.277	8.199	6.444	9.312	9.872

TABLE 2 The estimated parameters derived using the returns in W_1 and W_{350}, respectively

CAC40	α_{01}	α_{11}	α_{02}	α_{12}	β_{11}	β_{12}	η_1	η_2	γ_1	d
W_1	0.192	0.142	0.082	0.118	0.128	0.183	−0.009	−0.009	−0.166	2
W_{350}	0.207	0.239	0.021	0.103	0.383	0.729	−0.010	−0.009	−0.230	1
DAX30	α_{01}	α_{11}	α_{02}	α_{12}	β_{11}	β_{12}	η_1	η_2	γ_1	d
W_1	0.171	0.126	0.085	0.138	0.198	0.169	−0.010	−0.009	−0.036	2
W_{350}	0.250	0.251	0.021	0.100	0.219	0.735	−0.010	−0.009	−0.250	1
FTSE100	α_{01}	α_{11}	α_{02}	α_{12}	β_{11}	β_{12}	η_1	η_2	γ_1	d
W_1	0.109	0.215	0.055	0.154	0.196	0.222	−0.010	−0.008	−0.151	1
W_{350}	0.110	0.203	0.033	0.131	0.630	0.685	−0.009	−0.009	−0.181	1
NASDAQ	α_{01}	α_{11}	α_{02}	α_{12}	β_{11}	β_{12}	η_1	η_2	γ_1	d
W_1	0.188	0.079	0.118	0.084	0.178	0.126	−0.009	−0.009	−0.019	2
W_{350}	0.028	0.176	0.008	0.064	0.809	0.859	−0.010	−0.009	−0.139	1
NIKKEI225	α_{01}	α_{11}	α_{02}	α_{12}	β_{11}	β_{12}	η_1	η_2	γ_1	d
W_1	0.270	0.224	0.112	0.046	0.189	0.270	−0.010	−0.008	−0.165	1
W_{350}	0.401	0.244	0.039	0.159	0.216	0.644	−0.009	−0.009	−0.221	1
S&P500	α_{01}	α_{11}	α_{02}	α_{12}	β_{11}	β_{12}	η_1	η_2	γ_1	d
W_1	0.110	0.164	0.063	0.088	0.161	0.159	−0.010	−0.009	−0.063	2
W_{350}	0.012	0.197	0.011	0.048	0.790	0.867	−0.009	−0.008	−0.016	1

given in Table 2, where W_{350} contains observations in the 2008–2009 financial crisis period. It is seen that the volatility persistence of all six return series is much stronger during the financial crisis period. It is worth noting that the estimates of η_1 and η_2 are very similar, which suggests that the conditional distribution of the returns is only slightly skewed. Moreover, the absolute value of η_1 and η_2 are small, suggesting that the conditional density function of the y_t has heavy tails on both sides and that both tails have similar shapes.

For each moving window W_t, the conditional ES estimates at a given level can be calculated by using Proposition 3 for our model and those discussed in Section 4 for the benchmark models. Hence, when t takes values from 1 to 1,000, we obtain one-day-ahead post-sample ES estimates for the next 1,000 days at a given level.

Figure 2 shows the one-step-ahead conditional ES estimates at levels 1%, 5%, 95% and 99% for the last 1,000 days of the six return series by using our model. Clearly, the dynamics of the ES processes are similar to those of the returns.

Before checking the ES estimates, we check the coverage probability of $\hat{VaR}_{t\tau_0}$ at six quantile levels, that is, let τ_0 equal to 1%, 2.5%, 5%, 95%, 97.5% and 99% respectively. The MSE value between the true levels of τ_0 and the corresponding estimated coverage probabilities of $\hat{VaR}_{t\tau_0}$ provides a simple measure about the performance of a model in studying VaR: A good model corresponds to a smaller value of MSE. The MSE values for all models are given in Table 7 in Appendix I. Clearly, all models performed similarly well in studying VaR, which suggests that the proposed model is compatible with the benchmark models on this aspect and which also provides
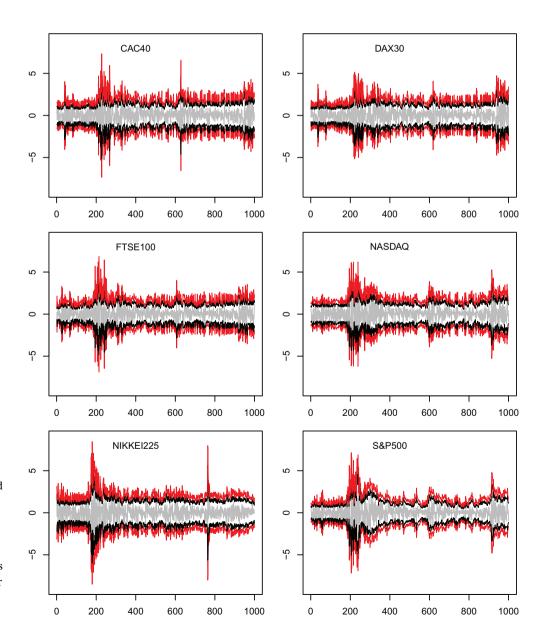

FIGURE 2 One-step-ahead conditional 1% (lower red), 5% (lower black), 95% (top black) and 99% (top red) ES estimates for the last 1,000 days, obtained from our model. The grey curves are the observed returns [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Number of rejections in the ES back-tests (Ljung-Box and Bootstrap) for all models at levels 1%, 2.5%, 5%, 95%, 97.5% and 99%

Our model	CAC40 (L, B) (0, 1)	DAX30 (L, B) (0, 0)	FTSE100 (L, B) (0, 0)	NASDAQ (L, B) (0, 0)	NIKKEI225 (L, B) (0, 0)	SP500 (L, B) (0, 0)	Total (L, B) (0, 1)
E-QR	CAC40 (L, B) (0, 1)	DAX30 (L, B) (0, 3)	FTSE100 (L, B) (0, 2)	NASDAQ (L, B) (0, 2)	NIKKEI225 (L, B) (1, 2)	SP500 (L, B) (0, 1)	Total (L, B) (1, 11)
Empirical	CAC40 (L, B) (1, 0)	DAX30 (L, B) (0, 0)	FTSE100 (L, B) (1, 0)	NASDAQ (L, B) (1, 0)	NIKKEI225 (L, B) (1, 0)	SP500 (L, B) (1, 0)	Total (L, B) (5, 0)
S-GARCH	CAC40 (L, B)	DAX30 (L, B)	FTSE100 (L, B)	NASDAQ (L, B)	NIKKEI225 (L, B)	SP500 (L, B)	Total (L, B)
Skewed t-	(0, 0)	(2, 1)	(0, 0)	(0, 0)	(0, 2)	(1, 6)	(3, 9)
Skewed GED	(0, 4)	(0, 5)	(0, 5)	(0, 6)	(0, 6)	(0, 6)	(0, 32)
NIG	(0, 4)	(0, 5)	(0, 5)	(0, 6)	(0, 5)	(0, 6)	(0, 31)
HGYP	(0, 4)	(0, 4)	(0, 6)	(0, 5)	(0, 5)	(0, 6)	(0, 30)
SJU	(0, 4)	(0, 5)	(0, 6)	(0, 6)	(0, 5)	(0, 6)	(0, 32)
GJR-GARCH	CAC40 (L, B)	DAX30 (L, B)	FTSE100 (L, B)	NASDAQ (L, B)	NIKKEI225 (L, B)	SP500 (L, B)	Total (L, B)
Skewed t-	(0, 0)	(0, 0)	(0, 0)	(0, 2)	(0, 1)	(2, 6)	(2, 9)
Skewed GED	(0, 5)	(0, 5)	(0, 6)	(0, 6)	(0, 5)	(0, 6)	(0, 33)
NIG	(0, 5)	(0, 5)	(0, 6)	(0, 6)	(0, 6)	(3, 6)	(3, 34)
HGYP	(0, 5)	(0, 5)	(0, 6)	(0, 5)	(0, 6)	(1, 6)	(1, 33)
SJU	(0, 6)	(0, 5)	(0, 6)	(0, 6)	(0, 6)	(2, 6)	(2, 35)
E-GARCH	CAC40 (L, B)	DAX30 (L, B)	FTSE100 (L, B)	NASDAQ (L, B)	NIKKEI225 (L, B)	SP500 (L, B)	Total (L, B)
Skewed t-	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(1, 1)	(0, 3)	(1, 4)
Skewed GED	(0, 6)	(0, 6)	(0, 6)	(0, 6)	(1, 4)	(0, 5)	(1, 33)
NIG	(0, 6)	(0, 5)	(0, 6)	(0, 6)	(1, 5)	(0, 6)	(1, 34)
HGYP	(0, 6)	(0, 5)	(0, 6)	(0, 6)	(0, 5)	(0, 5)	(0, 33)
SJU	(0, 6)	(0, 5)	(0, 5)	(0, 6)	(1, 5)	(1, 6)	(2, 33)
T-GARCH	CAC40 (L, B)	DAX30 (L, B)	FTSE100 (L, B)	NASDAQ (L, B)	NIKKEI225 (L, B)	SP500 (L, B)	Total (L, B)
Skewed t-	(0, 0)	(0, 0)	(0, 0)	(0, 0)	(0, 4)	(1, 5)	(1, 9)
Skewed GED	(0, 6)	(0, 5)	(0, 5)	(0, 6)	(0, 4)	(0, 5)	(0, 31)
NIG	(0, 6)	(0, 5)	(0, 6)	(0, 6)	(0, 4)	(0, 5)	(0, 32)
HGYP	(0, 6)	(0, 6)	(0, 6)	(0, 6)	(1, 5)	(6, 6)	(7, 35)
SJU	(0, 6)	(0, 5)	(0, 5)	(0, 6)	(1, 5)	(0, 5)	(1, 32)

some supports to the use of the benchmark models in studying VaR for financial data.

However, it is worth noting that a model that performs well in estimating VaR may not perform well in estimating ES because the estimation of VaR ignores important information regarding the tails of the underlying distributions, but that of ES does not. Hence, we now check how these models perform regarding the estimation of ES for financial returns. We back-test the ES estimates at the six levels. Tables 8 to 13 in Appendix I show the p-values of the Ljung-Box (LB) test for the standardized residuals and the Bootstrap test for the

TABLE 4 BIC values of GARCH models of order (p, q), $p, q = 1, 2$

CAC 40					
S-GARCH / order	**Skewed t-**	**Skewed GED**	**NIG**	**GHYP**	**JSU**
(1, 1)	0.86	0.86	0.85	0.86	0.85
(1, 2)	0.86	0.86	0.86	0.87	0.86
(2, 1)	0.86	0.86	0.86	0.87	0.86
(2, 2)	0.87	0.87	0.87	0.87	0.87
GJR-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.82	0.82	0.82	0.83	0.82
(1, 2)	0.83	0.83	0.83	0.83	0.83
(2, 1)	0.83	0.83	0.83	0.84	0.83
(2, 2)	0.84	0.84	0.84	0.84	0.84
E-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.80	0.80	0.80	0.81	0.80
(1, 2)	0.81	0.81	0.81	0.81	0.81
(2, 1)	0.81	0.81	0.81	0.82	0.81
(2, 2)	0.81	0.82	0.82	0.82	0.82
T-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.81	0.81	0.81	0.82	0.81
(1, 2)	0.81	0.81	0.81	0.82	0.81
(2, 1)	0.81	0.81	0.81	0.82	0.81
(2, 2)	0.81	0.81	0.81	0.82	0.81
DAX 30					
S-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.95	0.94	0.95	0.96	0.95
(1, 2)	0.96	0.95	0.96	0.96	0.96
(2, 1)	0.96	0.95	0.96	0.96	0.96
(2, 2)	0.97	0.96	0.96	0.97	0.96
GJR-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.93	0.92	0.93	0.93	0.93
(1, 2)	0.94	0.93	0.93	0.94	0.93
(2, 1)	0.94	0.94	0.94	0.95	0.94
(2, 2)	0.95	0.94	0.95	0.95	0.95
E-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.92	0.91	0.92	0.92	0.92
(1, 2)	0.93	0.92	0.92	0.93	0.92
(2, 1)	0.93	0.92	0.93	0.93	0.93
(2, 2)	0.94	0.93	0.94	0.94	0.94
T-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.92	0.92	0.92	0.93	0.92
(1, 2)	0.92	0.92	0.92	0.93	0.92
(2, 1)	0.92	0.92	0.92	0.93	0.92
(2, 2)	0.92	0.92	0.92	0.93	0.92

FTSE 100					
S-GARCH / order	**Skewed t-**	**Skewed GED**	**NIG**	**GHYP**	**JSU**
(1, 1)	0.51	0.51	0.51	0.51	0.51
(1, 2)	0.52	0.52	0.51	0.52	0.51
(2, 1)	0.51	0.52	0.51	0.52	0.51
(2, 2)	0.52	0.52	0.52	0.53	0.52
GJR-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.48	0.48	0.48	0.49	0.48
(1, 2)	0.49	0.49	0.49	0.49	0.49
(2, 1)	0.49	0.50	0.49	0.50	0.49
(2, 2)	0.50	0.50	0.50	0.51	0.50
E-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.47	0.48	0.47	0.48	0.47
(1, 2)	0.48	0.48	0.48	0.49	0.48
(2, 1)	0.48	0.48	0.48	0.49	0.48
(2, 2)	0.49	0.49	0.49	0.50	0.49
T-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.47	0.47	0.47	0.48	0.47
(1, 2)	0.47	0.47	0.47	0.48	0.47
(2, 1)	0.47	0.47	0.47	0.48	0.47
(2, 2)	0.47	0.47	0.47	0.48	0.47
NASDAQ					
S-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	1.07	1.06	1.07	1.08	1.07
(1, 2)	1.08	1.07	1.08	1.08	1.08
(2, 1)	1.07	1.07	1.07	1.08	1.07
(2, 2)	1.08	1.07	1.08	1.08	1.08
GJR-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	1.05	1.05	1.05	1.06	1.05
(1, 2)	1.06	1.06	1.06	1.07	1.06
(2, 1)	1.06	1.06	1.06	1.07	1.06
(2, 2)	1.07	1.06	1.06	1.07	1.06
E-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	1.05	1.05	1.05	1.06	1.05
(1, 2)	1.06	1.05	1.06	1.06	1.06
(2, 1)	1.05	1.05	1.05	1.05	1.05
(2, 2)	1.05	1.05	1.05	1.06	1.05
T-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	1.06	1.05	1.06	1.06	1.06
(1, 2)	1.06	1.05	1.06	1.06	1.06
(2, 1)	1.06	1.05	1.06	1.06	1.06
(2, 2)	1.06	1.05	1.06	1.06	1.06

TABLE 5 BIC values of GARCH models of order (p, q), $p, q = 1, 2$

TABLE 6 BIC values of GARCH models of order (p, q), $p, q = 1, 2$

NIKKEI 225					
S-GARCH / order	**Skewed t-**	**Skewed GED**	**NIG**	**GHYP**	**JSU**
(1, 1)	1.34	1.34	1.34	1.34	1.33
(1, 2)	1.35	1.35	1.34	1.35	1.34
(2, 1)	1.35	1.35	1.34	1.35	1.34
(2, 2)	1.35	1.35	1.35	1.35	1.35
GJR-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	1.33	1.32	1.32	1.33	1.32
(1, 2)	1.33	1.33	1.33	1.34	1.33
(2, 1)	1.34	1.33	1.33	1.34	1.33
(2, 2)	1.34	1.34	1.34	1.35	1.34
E-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	1.32	1.31	1.31	1.32	1.31
(1, 2)	1.32	1.32	1.32	1.33	1.32
(2, 1)	1.32	1.32	1.32	1.33	1.32
(2, 2)	1.33	1.32	1.33	1.33	1.33
T-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	1.31	1.31	1.31	1.32	1.31
(1, 2)	1.31	1.31	1.31	1.32	1.31
(2, 1)	1.31	1.31	1.31	1.32	1.31
(2, 2)	1.31	1.31	1.31	1.32	1.31
S&P 500					
S-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.51	0.50	0.51	0.51	0.51
(1, 2)	0.52	0.51	0.52	0.52	0.52
(2, 1)	0.51	0.51	0.51	0.52	0.51
(2, 2)	0.52	0.51	0.52	0.52	0.52
GJR-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.48	0.48	0.48	0.49	0.48
(1, 2)	0.49	0.48	0.49	0.50	0.49
(2, 1)	0.50	0.49	0.49	0.50	0.49
(2, 2)	0.50	0.49	0.49	0.50	0.49
E-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.47	0.47	0.47	0.48	0.47
(1, 2)	0.48	0.47	0.48	0.48	0.48
(2, 1)	0.47	0.47	0.47	0.48	0.47
(2, 2)	0.49	0.48	0.48	0.49	0.49
T-GARCH / order	Skewed t-	Skewed GED	NIG	GHYP	JSU
(1, 1)	0.49	0.48	0.48	0.49	0.48
(1, 2)	0.49	0.48	0.48	0.49	0.48
(2, 1)	0.49	0.48	0.48	0.49	0.48
(2, 2)	0.49	0.48	0.48	0.49	0.48

	CAC40	DAX30	FTSE100	NASDAQ	NIKKEI225	S&P500
Our model	0.001	0.000	0.000	0.001	0.000	0.001
E-QR	0.000	0.000	0.000	0.000	0.000	0.000
S-GARCH						
Skewed t	0.000	0.000	0.000	0.000	0.000	0.000
Skewed GED	0.000	0.000	0.000	0.000	0.000	0.000
NIG	0.000	0.000	0.000	0.000	0.000	0.000
HGYP	0.000	0.000	0.000	0.000	0.000	0.000
JSU	0.000	0.000	0.000	0.000	0.000	0.000
GJR-GARCH						
Skewed t	0.000	0.000	0.000	0.000	0.000	0.000
Skewed GED	0.000	0.000	0.000	0.000	0.000	0.000
NIG	0.000	0.000	0.000	0.000	0.000	0.000
HGYP	0.000	0.000	0.000	0.000	0.000	0.000
JSU	0.000	0.000	0.000	0.000	0.000	0.000
E-GARCH						
Skewed t	0.000	0.000	0.000	0.000	0.000	0.000
Skewed GED	0.000	0.000	0.000	0.000	0.000	0.000
NIG	0.000	0.000	0.000	0.000	0.000	0.000
HGYP	0.000	0.000	0.000	0.000	0.000	0.000
JSU	0.000	0.000	0.000	0.000	0.000	0.000
T-GARCH						
Skewed t	0.000	0.000	0.000	0.000	0.000	0.000
Skewed GED	0.000	0.000	0.000	0.000	0.000	0.000
NIG	0.000	0.000	0.000	0.000	0.000	0.000
HGYP	0.000	0.000	0.000	0.000	0.000	0.000
JSU	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 7 MSEs between the true level τ_0 and the estimated coverage probability of $\hat{VaR}_{t\tau_0}$ for all models and all data sets

exceedance residuals. Tables 3 summarizes the results of the tests, where L and B give the total number of rejections in the LB test and Bootstrap test respectively among the six levels. For example, for the CAC40 return series, we have (L, B) = (0, 1) for our model. Then the number 0 shows that the standardized residuals on all six levels are iid, and the number 1 indicates that except for one level, the exceedance residuals have a zero mean. Hence, a lower number of rejections in both tests corresponds to a better model for conditional ES estimation.

The final column of Table 3 gives the total number of rejections for each back-test among the six return series. So the results for our model are very good because there is only one Bootstrap rejection that occurred for CAC40 at a single level only.

Several findings can be seen from Table 3: (i) Our model outperforms all benchmark models regarding the post-sample ES estimation at all levels considered in this paper, which provides further evidence on the robustness

of the proposed model as it performed very well throughout the observation period including the financial crisis period. (ii) Among the benchmark GARCH models, the overall performance of the skewed t-distribution is better than all other standard distributions, which provides some support to the used of skewed t-distribution in practice. (iii) The empirical method did not reject the null of the Bootstrap test but the standardized residuals could be correlated. (iv) The ES estimates obtained from most benchmark GARCH models and the E-QR models rejected the null hypothesis of the Bootstrap test, which suggests that the ES estimates obtained from these models could be biased.

Some possible reasons for the satisfactory performance of our model are discussed below. One is that the ES estimates from our model can take account of more information about the tails of the conditional distribution of y_t, but such information is not considered by benchmark models. Another possible reason could be that our

TABLE 8 *p*-values of the back-test for ES—CAC40

τ	Our model		E-QR		Empirical			
	LB	Bootrap	LB	Bootrap	LB	Bootrap		
.01	.44	.01	.80	.33	.84	.31		
.025	.08	.01	.53	.05	.02	.19		
.05	.09	.00	.43	.32	.00	.03		
.95	.89	.09	.85	.04	.61	.10		
.975	1.00	.08	.95	.00	.94	.31		
.99	.98	.14	.98	.01	.95	.66		
	GJR-GARCH		S-GARCH		E-GARCH		T-GARCH	
Skewed t-	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.35	.90	.55	.56	.63	.97	.32	.81
.025	.45	.25	.92	.65	.08	.42	.29	.26
.05	.35	.04	.99	.08	.63	.28	.90	.03
.95	.90	.59	.96	.74	.06	.91	.63	.62
.975	.73	.42	.57	.38	.99	.71	.89	.48
.99	.90	.70	.56	.79	.66	.96	.70	.91
Skewed GED	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.71	.00	.95	.05	.37	.00	.21	.00
.025	.25	.00	.98	.00	.55	.00	.56	.00
.05	.29	.00	.36	.00	.60	.00	.59	.00
.95	.83	.00	.55	.00	.97	.00	.88	.00
.975	.76	.00	.58	.00	.67	.00	.09	.00
.99	.98	.01	.52	.03	.18	.00	.68	.00
NIG	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.61	.00	.67	.06	.70	.00	.47	.00
.025	.35	.00	.98	.00	.64	.00	.35	.00
.05	.56	.00	.39	.00	.28	.00	.53	.00
.95	.64	.00	.98	.00	.80	.00	.88	.00
.975	.94	.00	.64	.00	.94	.00	.20	.00
.99	.80	.01	.54	.03	.02	.00	.45	.00
GHYP	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.74	.01	.83	.06	.30	.00	.11	.00
.025	.26	.00	.98	.00	.32	.00	.26	.00
.05	.12	.00	.57	.00	.84	.00	.85	.00
.95	.74	.00	.96	.00	.48	.00	.91	.00
.975	.63	.00	.63	.00	.95	.00	.47	.00
.99	.98	.00	.21	.03	.19	.00	.18	.00
JSU	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.76	.00	.83	.06	.39	.00	.24	.00
.025	.19	.00	.99	.00	1.00	.00	.83	.00
.05	.89	.00	.41	.00	.11	.00	.20	.00
.95	.14	.00	.96	.00	.99	.00	.80	.00
.975	.52	.00	.64	.00	.87	.00	.63	.00
.99	.74	.00	.55	.02	.72	.00	.62	.00

τ	Our model		E-QR		Empirical			
	LB	Bootrap	LB	Bootrap	LB	Bootrap		
.01	.01	.06	.32	.14	.17	.49		
.025	.56	.02	.45	.00	.90	.14		
.05	.02	.01	.66	.00	.07	.06		
.95	.16	.07	.99	.03	.09	.21		
.975	.13	.03	.98	.36	.40	.83		
.99	.90	.02	.80	.00	1.00	.58		
	GJR-GARCH		S-GARCH		E-GARCH		T-GARCH	
Skewed t-	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.16	.49	.00	.00	.70	.26	.24	.78
.025	.83	.06	.79	.90	.86	.09	.58	.12
.05	1.00	.43	.99	.04	.64	.11	.66	.03
.95	.25	.02	.53	.18	.66	.86	.30	.19
.975	.69	.03	.00	.08	.15	.27	.85	.09
.99	.38	.00	.56	.19	.70	.03	.23	.11
Skewed GED	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.51	.02	.64	.07	.62	.00	.42	.01
.025	.83	.00	.98	.00	.40	.00	.74	.00
.05	.98	.00	.91	.00	.94	.00	.56	.00
.95	.15	.00	.94	.00	.91	.00	.41	.00
.975	.03	.00	.53	.00	.27	.00	.71	.00
.99	.42	.00	.38	.00	.53	.00	.73	.00
NIG	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.96	.04	.66	.07	.81	.01	.48	.01
.025	.69	.00	1.00	.00	.57	.00	.77	.00
.05	.58	.00	.95	.00	.71	.00	.69	.00
.95	.18	.00	.77	.00	.87	.00	.32	.00
.975	.91	.00	.73	.00	.31	.00	.83	.00
.99	.68	.00	.79	.00	.40	.00	.84	.00
GHYP	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.90	.02	.79	.08	.39	.01	.35	.00
.025	1.00	.00	.99	.01	.83	.00	.65	.00
.05	.31	.00	.97	.00	.58	.00	1.00	.00
.95	.07	.00	.99	.00	.54	.00	.98	.00
.975	.05	.00	.77	.00	.47	.00	.72	.00
.99	.55	.00	.46	.00	.39	.00	.73	.00
JSU	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.91	.03	.73	.08	.62	.01	.40	.01
.025	.71	.00	1.00	.00	.58	.00	.88	.00
.05	.96	.00	.84	.00	1.00	.00	.84	.00
.95	.23	.00	.99	.00	.83	.00	.82	.00
.975	.27	.00	.81	.00	.26	.00	.91	.00
.99	.59	.00	.80	.00	.22	.00	.70	.00

TABLE 9 *p*-values of the back-test for ES—DAX30

TABLE 10 *p*-values of the back-test for ES—FTSE100

τ	Our model		E-QR		Empirical			
	LB	Bootrap	LB	Bootrap	LB	Bootrap		
.01	.41	.01	.18	.62	.76	.19		
.025	.16	.02	.84	.18	.02	.28		
.05	.62	.01	.04	.11	.00	.30		
.95	.87	.12	.99	.11	.99	.12		
.975	.89	.09	.88	.00	.66	.15		
.99	.77	.02	.22	.00	.80	.09		
	GJR-GARCH		S-GARCH		E-GARCH		T-GARCH	
Skewed t-	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.88	.47	.54	.56	.83	.60	.79	.70
.025	.96	.50	.88	.21	.96	.52	.98	.74
.05	.99	.26	.81	.26	.28	.35	.95	.25
.95	.97	.78	.82	.06	.28	.66	.74	.65
.975	.85	.38	.98	.06	.93	1.00	.99	.72
.99	.97	.10	.52	.62	.99	.06	.99	.02
Skewed GED	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.89	.00	.95	.00	.99	.00	.98	.00
.025	.77	.00	.96	.00	.77	.00	.91	.00
.05	.66	.00	.15	.00	.75	.00	.71	.00
.95	.94	.00	.58	.00	.59	.00	.73	.00
.975	.09	.00	.83	.00	.78	.00	.98	.00
.99	.70	.00	.84	.01	.97	.00	.96	.01
NIG	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.93	.00	.62	.00	1.00	.00	.93	.00
.025	.90	.00	.98	.00	.99	.00	.72	.00
.05	.79	.00	.57	.00	.95	.00	.69	.00
.95	.83	.00	.98	.00	.86	.00	.60	.00
.975	.75	.00	.88	.00	.99	.00	.58	.00
.99	.95	.00	.84	.01	.98	.00	1.00	.00
GHYP	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.96	.00	.99	.00	.93	.00	.89	.00
.025	.93	.00	.97	.00	.85	.00	.54	.00
.05	.66	.00	.02	.00	.88	.00	.88	.00
.95	.76	.00	.83	.00	.27	.00	.77	.00
.975	.01	.00	.90	.00	.99	.00	.49	.00
.99	.41	.00	.83	.00	.93	.00	.90	.00
JSU	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.94	.00	.89	.00	.98	.01	.95	.01
.025	.93	.00	.82	.00	.93	.00	.74	.00
.05	.90	.00	.37	.00	.88	.00	.53	.00
.95	.54	.00	.59	.00	.68	.00	.35	.00
.975	.01	.00	.67	.00	.96	.00	.81	.00
.99	.90	.00	.84	.00	.95	.00	.98	.00

τ	Our model		E-QR		Empirical			
	LB	Bootrap	LB	Bootrap	LB	Bootrap		
.01	1.00	.16	.03	.72	.79	.32		
.025	1.00	.04	.39	.27	.25	.04		
.05	.97	.01	.79	.59	.00	.05		
.95	.36	.18	1.00	.02	.02	.01		
.975	1.00	.51	.92	.00	.82	.09		
.99	.99	.36	1.00	.00	.86	.33		
	GJR-GARCH		S-GARCH		E-GARCH		T-GARCH	
Skewed t-	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.45	.04	.38	.06	.66	.58	.16	.69
.025	.96	.01	.43	.30	.59	.54	.90	.31
.05	.37	.01	.20	.24	.85	.03	.82	.03
.95	.81	.06	.60	.01	.84	.03	.75	.01
.975	.42	.00	.81	.03	.97	.02	.11	.05
.99	.31	.00	.04	.33	.03	.01	.06	.00
Skewed GED	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.72	.00	.28	.00	.51	.00	.82	.00
.025	.79	.00	.27	.00	.71	.00	.28	.00
.05	.92	.00	.58	.00	.56	.00	1.00	.00
.95	.59	.00	.99	.00	.63	.00	.91	.00
.975	.25	.00	.84	.00	.66	.00	.83	.00
.99	.95	.00	.53	.00	.98	.00	.59	.00
NIG	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.29	.00	.24	.00	.64	.00	.35	.00
.025	.57	.00	.20	.00	.52	.00	.55	.00
.05	.54	.00	.08	.00	.96	.00	.80	.00
.95	.22	.00	.77	.00	.71	.00	.98	.00
.975	.16	.00	.46	.00	.13	.00	.14	.00
.99	.75	.00	.37	.00	.41	.00	.91	.00
GHYP	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.91	.01	.05	.01	.18	.00	.61	.00
.025	.01	.00	.19	.00	.80	.00	.91	.00
.05	.88	.00	.58	.00	.92	.00	.69	.00
.95	.88	.00	.52	.00	.24	.00	.96	.00
.975	.77	.00	.68	.00	.98	.00	.30	.00
.99	.85	.00	.52	.00	.87	.00	.78	.00
JSU	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.44	.00	.60	.00	.81	.00	.82	.00
.025	.14	.00	.13	.00	.46	.00	.77	.00
.05	.77	.00	.78	.00	.95	.00	.77	.00
.95	.95	.00	.84	.00	.99	.00	.98	.00
.975	.30	.00	.77	.00	.70	.00	.77	.00
.99	.89	.00	.28	.00	.28	.00	.29	.00

TABLE 11 p-values of the back-test for ES—NASDAQ

TABLE 12 *p*-values of the back-test for ES—NIKKEI225

	Our model		E-QR		Empirical			
τ	LB	Bootrap	LB	Bootrap	LB	Bootrap		
.01	.73	.06	.22	.30	.21	.33		
.025	.03	.02	.22	.20	.02	.03		
.05	.02	.01	.00	.62	.00	.10		
.95	.40	.11	1.00	.04	.82	.04		
.975	.52	.20	.73	.00	.98	.22		
.99	.53	.82	.74	.00	.98	.68		
	GJR-GARCH		S-GARCH		E-GARCH		T-GARCH	
Skewed t-	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.65	.88	.75	.11	.38	.50	.10	.80
.025	.45	.96	.68	.28	.67	.37	.69	.47
.05	.20	.62	.12	.57	.33	.92	.59	.70
.95	.20	.34	.07	.05	.00	.36	.11	.53
.975	.02	.00	.43	.00	.08	.00	.15	.00
.99	.24	.03	.34	.00	.70	.82	.72	.46
Skewed GED	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.77	.00	.64	.00	.49	.01	.97	.01
.025	.07	.00	.74	.00	.99	.00	.94	.00
.05	.88	.00	.47	.00	.04	.00	.24	.00
.95	.11	.00	.74	.00	.00	.00	.12	.00
.975	.47	.00	.25	.00	.19	.00	.28	.00
.99	.06	.01	.76	.00	.17	.01	.06	.01
NIG	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.90	.00	.56	.00	.31	.02	.68	.01
.025	.03	.00	.12	.00	.34	.00	.61	.00
.05	.99	.00	.31	.00	.28	.00	.27	.00
.95	.32	.00	.60	.00	.00	.00	.20	.00
.975	.31	.00	.35	.00	.28	.00	.76	.00
.99	.03	.00	.89	.01	.18	.00	.32	.00
GHYP	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.87	.00	.57	.00	.47	.01	.39	.02
.025	.73	.00	.35	.00	.66	.00	.88	.00
.05	.83	.00	.27	.00	.27	.00	.23	.00
.95	.57	.00	.89	.00	.05	.00	.00	.00
.975	.80	.00	.39	.00	.26	.00	.38	.00
.99	.05	.00	.89	.05	.17	.00	.32	.00
JSU	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.88	.00	.86	.00	.65	.02	.55	.01
.025	.62	.00	.47	.00	.79	.00	.91	.00
.05	.90	.00	.24	.00	.31	.00	.13	.00
.95	.13	.00	.28	.00	.00	.00	.00	.00
.975	.13	.00	.92	.00	.67	.00	.03	.00
.99	.03	.00	.71	.00	.18	.00	.24	.00

τ	Our model		E-QR		Empirical			
	LB	Bootrap	LB	Bootrap	LB	Bootrap		
.01	1.00	.18	.05	.13	.95	.06		
.025	1.00	.05	.43	.68	.09	.05		
.05	1.00	.01	.69	.63	.00	.01		
.95	.93	.66	1.00	.01	.04	.02		
.975	.97	.18	1.00	.00	.96	.09		
.99	.52	.06	1.00	.26	.92	.12		
	GJR-GARCH		S-GARCH		E-GARCH		T-GARCH	
Skewed t-	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.00	.00	.00	.00	.73	.00	.15	.00
.025	.11	.00	.76	.00	.61	.12	.19	.05
.05	.47	.00	.47	.00	.21	.01	.83	.00
.95	.00	.00	.09	.00	.31	.00	.00	.00
.975	.60	.00	.91	.00	.55	.00	.77	.00
.99	.66	.00	.10	.00	.26	.02	.06	.00
Skewed GED	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.14	.00	.80	.00	.41	.00	.86	.00
.025	.19	.00	.08	.00	.40	.00	.60	.00
.05	.08	.00	.67	.00	.48	.00	.17	.00
.95	.02	.00	.50	.00	.39	.00	.38	.00
.975	.80	.00	.51	.00	.71	.00	.68	.00
.99	.38	.00	.70	.00	.10	.02	.34	.08
NIG	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.00	.00	.63	.00	.45	.00	.59	.00
.025	.50	.00	.06	.00	.79	.00	.99	.00
.05	.00	.00	.27	.00	.98	.00	1.00	.00
.95	.00	.00	.10	.00	.31	.00	.16	.00
.975	.10	.00	.07	.00	.95	.00	.96	.00
.99	.69	.00	.46	.00	.14	.00	.62	.02
GHYP	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.33	.00	.47	.00	.61	.00	.00	.00
.025	.21	.00	.62	.00	.68	.00	.00	.00
.05	.29	.00	.37	.00	.64	.00	.00	.00
.95	.00	.00	.06	.00	.18	.00	.00	.00
.975	.60	.00	.20	.00	.34	.00	.00	.00
.99	.15	.00	.08	.00	.70	.02	.00	.00
JSU	LB	Bootrap	LB	Bootrap	LB	Bootrap	LB	Bootrap
.01	.00	.00	.51	.00	.50	.00	.48	.00
.025	.71	.00	.08	.00	.57	.00	.63	.00
.05	.03	.00	.09	.00	.65	.00	.84	.00
.95	.00	.00	.01	.00	.00	.00	.07	.00
.975	.70	.00	.06	.00	.87	.00	.95	.00
.99	.57	.00	.53	.00	.48	.00	.39	.01

TABLE 13 *p*-values of the back-test for ES—S&P500

model estimates the threshold, while the benchmark T-GARCH models fix the threshold at value 0, representing a zero return. However, Yu et al. (2010) pointed out that the threshold value could be slightly greater or less than 0, which suggests that a small amount of variation around 0 in financial returns may not be significant enough to change their volatility dynamics. Hence, estimating rather than fixing the threshold value would be more reasonable. Our results provide some further support to this argument.

6 | CONCLUDING DISCUSSIONS

This paper proposed a novel quantile function T-GARCH model for estimating ES of financial returns. Different from the GARCH models defined by standard distributions, the proposed model defines the conditional distribution of financial returns by using the GLD. We found that the ES estimates obtained from our model are significantly better than those obtained from the benchmark models considered in this paper. Our results show that, apart from the overall shape of the distribution of returns, it is also important to take account of specific features about the tails of the distribution when estimating the ES of returns.

It is also worth reemphasizing that the GLD is defined via a quantile function as its other equivalent functions do not have an explicit mathematical expression. There are many other distributions that are also defined by quantile functions. It is certainly worth further investigations on the use of quantile functions to build up useful models for financial analysis.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available from the following Yahoo Finance websites, which have also been cited in Section 5 (see page 16) and included in the Reference section of this paper. Yahoo Finance (2019a).

ORCID
Yuzhi Cai https://orcid.org/0000-0003-3509-9787

REFERENCES
Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, *26*, 1487–1503.
Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1997). Thinking coherently. *Risk*, *10*, 68–71.
Artzner, P. E., Delbain, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, *9*, 203–228.
Ausín, M. C., & Galeano, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, *51*, 2636–2652.
Bauwens, L., & Lubrano, M. (2002). Bayesian option pricing using asymmetric GARCH models. *Journal of Empirical Finance*, *9*, 321–342.
Bollerslev, T. (1986). Generalized autoregressive conditional Heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.
Brooks, S. P. (1998). Markov chain Monte Carlo method and its application. *The Statistician*, *47*, 69–100.
Chen, C. W. S., So, M. K. P., & Liu, F. C. (2011). A review of threshold time series models in finance. *Statistics and its Interface*, *4*, 167–181.
Corlu, C. G., & Corlu, A. (2015). Modelling exchange rate returns: Which flexible distribution to use? *Quantitative Finance*, *15*, 1851–1864.
Corrado, C. J. (2001). Option pricing based on the generalized lambda distribution. *Journal of Futures Markets*, *21*, 213–236.
Dellaportas, P., & Vrontos, I. D. (2007). Modelling volatility asymmetries: A Bayesian analysis of a class of tree structured multivariate GARCH models. *The Econometrics Journal*, *10*, 503–520.
Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall.
Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.
Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*, 367–381.
Fournier, B., Rupin, N., Bigerelle, M., Najjar, D., Iost, A., & Wilcox, R. (2007). Estimating the parameters of a generalized lambda distribution. *Computational Statistics & Data Analysis*, *51*, 2813–2835.
Freimer, M., Mudholkar, G., Kollia, G., & Lin, C. (1988). A study of the generalised Tukey lambda family. *Communications in Statistics - Theory Methods*, *17*, 3547–3567.
Gamerman, D., & Lopes, H. F. (2002). *Markov chain Monte Carlo: Stochastic simulation for Bayesian inference* (2nd ed.). FL: Chapman and Hall/CRC.
Geyer, C. J. (2011). Introduction to Markov chain Monte Carlo. In S. Brooks, A. Gelman, G. K. Jones, & X. L. Meng (Eds.), *Handbook of Markov chain Monte Carlo* (pp. 3–48). Florida: CRC Press.
Gilchrist, W. G. (2000). *Statistical Modelling with Quantile functions*, London: Chapman & Hall/CRC.
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. *Journal of Finance*, *48*, 1779–1801.
Jensen, M. J., & Maheu, J. M. (2013). Bayesian Semiparametric multivariate GARCH modeling. *Journal of Econometrics*, *176*, 3–17.
Karian, Z. A., Dudewicz, E. J., & McDonald, P. (1996). The extended generalized lambda distribution system for fitting distributions to data: History, completion of theory, tables, applications, the final word on moment fits. *Communications in Statistics - Simulation and Computation*, *25*, 611–642.
Karian, Z. A., & Dudewicz, E. J. (1999). Fitting the generalised lambda distribution to data: A method based on percentiles. *Communications in Statistics - Simulation and Computation*, *28*, 793–819.

King, R., & MacGillivray, H. (1999). A starship estimation method for the generalised lambda distributions. *Australia New Zealand Journal of Statistics*, *41*, 353–374.

Kuester, K., Mittnik, S., & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, *4*, 53–89.

Lee, P.P. (2003). The generalized lambda distribution applied to spot exchange rates. (PhD Dissertation). Carnegie Mellon University.

Liu, J., Li, W. K., & Li, C. W. (1997). On a threshold autoregression with conditional heteroscedastic variance. *Journal of Statistical Planning and Inferences*, *62*, 279–300.

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, *7*, 271–300.

Nadarajah, S., Zhang, B., & Chan, S. (2014). Estimation methods for expected shortfall. *Quantitative Finance*, *14*, 271–291.

Necula, C. (2009). Modelling heavy-tailed stock index returns using the generalized hyperbolic distribution. *Romanian Journal of Economic Forecasting*, *2*, 118–131.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*, 347–370.

O'Hagan, A., & Forster, J. J. (2004). *Bayesian inference*. London: Arnold.

Önalan, Ö. (2010). Financial risk management with normal inverse Gaussian distributions. *International Research Journal of Finance and Economics*, *38*, 104–115.

Ozturk, A., & Dale, R. F. (1985). Least squares estimation of the parameters of the generalized lambda distribution. *Technometrics*, *27*, 81–84.

Poon, S., & Granger, C. J. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, *41*, 478–639.

Socgnia, V. K., & Wilcox, D. (2014). A comparison of generalized hyperbolic distribution models for equity returns. *Journal of Applied Mathematics*, *2014*, 263465. https://doi.org/10.1155/2014/263465

Su, S. (2007a). Numerical maximum log likelihood estimation for generalized lambda distributions. *Computational Statistics & Data Analysis*, *51*, 3983–3998.

Su, S. (2007b). Fitting single and mixture of generalized lambda distributions to data via discretized and maximum likelihood methods: GLDEX in R. *Journal of Statistical Software*, *21*(9), 1–17.

Tarsitano, A. (2004). Fitting the generalized lambda distribution to income data. In *COMPSTAT 2004-Proceedings in Computational Statistics* (p. 2004). Prague, Czech Republic: Springer.

Taylor, J. W. (2008). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, *6*, 382–406.

Theodossiou, P. (2015). Skewed generalized error distribution of financial assets and option pricing. *Multinational Finance Journal*, *19*, 223–266.

Virbickaite, A., Ausín, C. M., & Galeano, P. (2015). Bayesian inference methods for univariate and multivariate GARCH models: A survey. *Journal of Economic Surveys*, *29*, 76–96.

Yahoo Finance (2019a). https://finance.yahoo.com/quote/%5EFCHI/history/

Yahoo Finance (2019b). https://finance.yahoo.com/quote/%5EGDAXI/history/

Yahoo Finance (2019c). https://finance.yahoo.com/quote/%5EFTSE%3FP%3DFTSE/history/

Yahoo Finance (2019d). https://finance.yahoo.com/quote/%5EN225/history/

Yahoo Finance (2019e). https://finance.yahoo.com/quote/%5EIXIC/history/

Yahoo Finance (2019f). https://finance.yahoo.com/quote/%5EGSPC/history/

Yamai, Y., & Yoshiba, T. (2002). On the validity of value-at-risk: Comparative analyses with expected shortfall. *Monetary and Economic Studies*, *20*, 57–85.

Yang, Y. L., & Chang, C. L. (2008). A double-threshold GARCH model of stock market and currency shocks on stock returns. *Mathematics and Computers in Simulation*, *79*, 458–474.

Yu, P. L. H., Li, W. K., & Jin, S. (2010). On some models for Value at Risk. *Economic Review*, *29*, 622–641.

Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, *18*, 931–955.

APPENDIX A.

Proof of Proposition 1.

For $\mu(\boldsymbol{\eta})$, we have

$$
\begin{aligned}
\mu(\boldsymbol{\eta}) = E(e_t) &= \int_0^1 Q(\tau,\boldsymbol{\eta})d\tau \\
&= \frac{1}{\eta_1}\int_0^1 (\tau^{\eta_1}-1)d\tau - \frac{1}{\eta_2}\int_0^1 \{(1-\tau)^{\eta_2}-1\}d\tau \\
&= \frac{1}{\eta_1}\left\{\frac{1}{\eta_1+1}-1\right\} - \frac{1}{\eta_2}\left\{\frac{1}{\eta_2+1}-1\right\} = \frac{1}{\eta_2+1} - \frac{1}{\eta_1+1}.
\end{aligned}
$$

For $\sigma(\boldsymbol{\eta})$, first note that

$$
\begin{aligned}
E\left(e_t^2\right) &= \int_0^1 Q^2(\tau,\boldsymbol{\eta})d\tau = \int_0^1 \left(\frac{\tau^{\eta_1}-1}{\eta_1} - \frac{(1-\tau)^{\eta_2}-1}{\eta_1}\right)^2 d\tau \\
&= \int_0^1 \left(\frac{(\tau^{\eta_1}-1)^2}{\eta_1^2} + \frac{\{(1-\tau)^{\eta_2}-1\}^2}{\eta_2^2} - 2\frac{(\tau^{\eta_1}-1)\{(1-\tau)^{\eta_2}-1\}}{\eta_1\eta_2}\right) d\tau \\
&= \frac{1}{\eta_1^2}\left(\frac{1}{2\eta_1+1} - \frac{2}{\eta_1+1} + 1\right) + \frac{1}{\eta_2^2}\left(\frac{1}{2\eta_2+1} - \frac{2}{\eta_2+1} + 1\right) \\
&\quad - \frac{2}{\eta_1\eta_2}\left(\int_0^1 \tau^{\eta_1}(1-\tau)^{\eta_2}d\tau - \frac{1}{\eta_1+1} - \frac{1}{\eta_2+1} + 1\right).
\end{aligned}
$$

Hence

$$
\begin{aligned}
\sigma(\boldsymbol{\eta})^2 &= E\left(e_t^2\right) - \mu(\boldsymbol{\eta})^2 \\
&= \frac{1}{\eta_1^2}\left(\frac{1}{2\eta_1+1} - \frac{1}{(\eta_1+1)^2}\right) + \frac{1}{\eta_2^2}\left(\frac{1}{2\eta_2+1} - \frac{1}{(\eta_2+1)^2}\right) \\
&\quad - \frac{2}{\eta_1\eta_2}\left\{B(\eta_1,\eta_2) - \frac{1}{(\eta_1+1)(\eta_2+1)}\right\},
\end{aligned}
$$

where $B(\eta_1,\eta_2) = \int_0^1 \tau^{\eta_1}(1-\tau)^{\eta_2}d\tau$. It is seen that both $\mu(\boldsymbol{\eta})$ and $\sigma(\boldsymbol{\eta})$ are continuous functions of $\boldsymbol{\eta}$ on the closed set $\cup_{i,j=1}^3 (\Omega_i \times \Omega_j)$. Hence, they are finite on this set as required. This completes the proof.

APPENDIX B.

Proof of Proposition 2.

First we show that the quantile function of ε_t is given by $Q_{\varepsilon_t}(\tau,\boldsymbol{\eta}) = \{Q(\tau,\boldsymbol{\eta})-\mu(\boldsymbol{\eta})\}/\sigma(\boldsymbol{\eta})$. It follows from the definition of a quantile function that, for any $\tau \in (0, 1)$ we have

$$
\begin{aligned}
\tau &= P[e_t \leq Q(\tau,\boldsymbol{\eta})] \\
&= P[\{e_t-\mu(\boldsymbol{\eta})\}/\sigma(\boldsymbol{\eta}) \leq \{Q(\tau,\boldsymbol{\eta})-\mu(\boldsymbol{\eta})\}/\sigma(\boldsymbol{\eta})] \\
&= P[\varepsilon_t \leq \{Q(\tau,\boldsymbol{\eta})-\mu(\boldsymbol{\eta})\}/\sigma(\boldsymbol{\eta})].
\end{aligned}
$$

Hence, by definition, the quantile function of ε_t is given by $Q_{\varepsilon_t}(\tau,\boldsymbol{\eta}) = \{Q(\tau,\boldsymbol{\eta})-\mu(\boldsymbol{\eta})\}/\sigma(\boldsymbol{\eta})$ as required.
Moreover, we have

$$\tau = P[\varepsilon_t \leq \{Q(\tau,\boldsymbol{\eta}) - \mu(\boldsymbol{\eta})\}/\sigma(\boldsymbol{\eta})]$$
$$= P\left[\sqrt{h_t}\varepsilon_t \leq \sqrt{h_t}\{Q(\tau,\boldsymbol{\eta}) - \mu(\boldsymbol{\eta})\}/\sigma(\boldsymbol{\eta})|\mathbf{y}_{t-1}\right]$$
$$= P\left[y_t \leq \sqrt{h_t}Q_{\varepsilon_t}(\tau,\boldsymbol{\eta})|\mathbf{y}_{t-1}\right].$$

Therefore, by definition, the conditional quantile function of y_t is given by.
$Q_{y_t}(\tau|\mathbf{y}_{t-1},\boldsymbol{\beta},\boldsymbol{\eta},\gamma_1,d) = \sqrt{h_t}Q_{\varepsilon_t}(\tau,\boldsymbol{\eta})$ as required.

APPENDIX C.

Proof of Proposition 3.

When $\tau \leq 0.5$, we have

$$ES_{t\tau} = E(y_t|y_t \leq \text{VaR}_\tau) = \frac{1}{\tau}\int_0^\tau Q_{y_t}\{\tau|\mathbf{y}_{t-1},\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},d\}d\tau$$

$$= \frac{\sqrt{h_t}}{\tau\sigma(\boldsymbol{\eta})}\int_0^\tau (Q(\tau,\boldsymbol{\eta}) - \mu(\boldsymbol{\eta}))d\tau$$

$$= \frac{\sqrt{h_t}}{\tau\sigma(\boldsymbol{\eta})}\left\{\frac{1}{\eta_1}\left(\frac{\tau^{\eta_1+1}}{\eta_1+1} - \tau\right) - \frac{1}{\eta_2}\left(-\frac{(1-\tau)^{\eta_2+1}}{\eta_2+1} + \frac{1}{\eta_2+1} - \tau\right) - \mu(\boldsymbol{\eta})\tau\right\}.$$

It follows from Proposition 1 that $\mu(\boldsymbol{\eta}) = 1/(\eta_2 + 1) - 1/(\eta_1 + 1)$ and hence

$$ES_{t\tau} = \frac{\sqrt{h_t}}{\tau\sigma(\boldsymbol{\eta})}\left\{\frac{\tau^{\eta_1+1} - \tau}{\eta_1(\eta_1+1)} + \frac{(1-\tau)^{\eta_2+1} - (1-\tau)}{\eta_2(\eta_2+1)}\right\}.$$

Similarly, when $\tau > 0.5$, we have

$$ES_{t\tau} = E(y_t|y_t \geq \text{VaR}_\tau) = \frac{1}{1-\tau}\int_\tau^1 Q_{y_t}\{\tau|\mathbf{y}_{t-1},\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},d\}d\tau$$

$$= \frac{\sqrt{h_t}}{(1-\tau)\sigma(\boldsymbol{\eta})}\int_\tau^1 (Q(\tau,\boldsymbol{\eta}) - \mu)d\tau$$

$$= \frac{\sqrt{h_t}}{(1-\tau)\sigma(\boldsymbol{\eta})}\left\{\frac{1}{\eta_1}\left(\frac{1}{\eta_1+1} - \frac{\tau^{\eta_1+1}}{\eta_1+1} - (1-\tau)\right) - \frac{1}{\eta_2}\left(\frac{(1-\tau)^{\eta_2+1}}{\eta_2+1} - (1-\tau)\right) - \mu(\boldsymbol{\eta})(1-\tau)\right\}$$

$$= \frac{\sqrt{h_t}}{(1-\tau)\sigma(\boldsymbol{\eta})}\left\{\frac{\tau - \tau^{\eta_1+1}}{\eta_1(\eta_1+1)} + \frac{(1-\tau) - (1-\tau)^{\eta_2+1}}{\eta_2(\eta_2+1)}\right\}.$$

This completes the proof.

APPENDIX D.

Proof of Proposition 4.

Let $\boldsymbol{\xi}$ represents the vector of parameters. Then we need to show that $\int_\Omega \pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},d|\mathbf{y}_T)d\boldsymbol{\xi} < \infty$.

First note that $\tau^{\eta_1-1} + (1-\tau)^{\eta_2-1} \geq 1$ for all $\eta_1, \eta_2 \in \Omega$ and $\tau \in (0, 1)$. Moreover, for any $\boldsymbol{\xi} \in \Omega$, we have $\sqrt{h_t} \geq \sqrt{\sum_{j=1}^J m_0 I\left(\gamma_{j-1} \leq y_{t-d} < \gamma_j\right)}$. Hence,

$$\left\{\tau^{\eta_1-1}+(1-\tau)^{\eta_2-1}\right\}\sqrt{h_t}\geq\sqrt{\sum_{j=1}^{J}m_0 I\left(\gamma_{j-1}\leq y_{t-d}<\gamma_j\right)}\geq m_0.$$

On the other hand, it follows from Proposition 1 that $\sigma(\boldsymbol{\eta})^2<\infty$ for any η_1 and η_2 in Ω. Therefore, as $\int_\Omega \pi(\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},d)$ $d\boldsymbol{\xi}<\infty$, we see that

$$\int_\Omega \pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta},d|\mathbf{y}_T)d\boldsymbol{\xi}=M^*\int_\Omega \prod_{t=d_0+1}^{T}\frac{\sigma(\boldsymbol{\eta})}{\sqrt{h_t}\left\{\tau_t^{\eta_1-1}+(1-\tau_t)^{\eta_2-1}\right\}}\pi(\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},d)d\boldsymbol{\xi}$$

$$\leq M^*\left\{\frac{\sigma(\boldsymbol{\eta})}{m_0}\right\}^{T-d_0}\int_\Omega \pi(\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\gamma},d)d\boldsymbol{\xi}<\infty,$$

where M^* is the normalizing constant of the posterior density function. Therefore, the posterior density function is well defined on Ω as required. This completes the proof.

APPENDIX E.

Prior density functions:

$$\pi(\alpha_{0j})=\left\{\sqrt{2\pi}\sigma_{0j}\alpha_{0j}\right\}^{-1}e^{-\ln^2\alpha_{0j}/2\sigma_{0j}^2}\left\{\int_{m_0}^{M}\left\{\sqrt{2\pi}\sigma_{0j}\alpha_{0j}\right\}^{-1}e^{-\ln^2\alpha_{0j}/2\sigma_{0j}^2}d\alpha_{0j}\right\}^{-1}$$

$$=\left[\left\{\Phi\left(\frac{\ln M}{\sigma_{0j}}\right)-\Phi\left(\frac{\ln m_0}{\sigma_{0j}}\right)\right\}\sqrt{2\pi}\sigma_{0j}\alpha_{0j}\right]^{-1}e^{-\ln^2\alpha_{0j}/2\sigma_{0j}^2}$$

$$\pi(\alpha_{1j})=\left\{\sqrt{2\pi}\sigma_{1j}\alpha_{1j}\right\}^{-1}e^{-\ln^2\alpha_{1j}/2\sigma_{1j}^2}\left\{\int_0^a\left\{\sqrt{2\pi}\sigma_{1j}\alpha_{1j}\right\}^{-1}e^{-\ln^2\alpha_{1j}/2\sigma_{1j}^2}d\alpha_{1j}\right\}^{-1}$$

$$=\left[\Phi\left(\frac{\ln a}{\sigma_{1j}}\right)\sqrt{2\pi}\sigma_{1j}\alpha_{1j}\right]^{-1}e^{-\ln^2\alpha_{1j}/2\sigma_{1j}^2}$$

$$\pi(\beta_{1j})=\left\{\sqrt{2\pi}s_{1j}\beta_{1j}\right\}^{-1}e^{-\ln^2\beta_{1j}/2s_{1j}^2}\left\{\int_0^{a-\alpha_{1j}}\left\{\sqrt{2\pi}s_{1j}\beta_{1j}\right\}^{-1}e^{-\ln^2\beta_{1j}/2s_{1j}^2}d\beta_{1j}\right\}^{-1}$$

$$=\left[\Phi\left(\frac{\ln(a-\alpha_{1j})}{s_{1j}}\right)\sqrt{2\pi}s_{1j}\beta_{1j}\right]^{-1}e^{-\ln^2\beta_{1j}/2s_{1j}^2}$$

$$\pi((-\eta_v))$$

$$=\left\{\sqrt{2\pi}\lambda_v(-\eta_v)\right\}^{-1}e^{-\ln^2(-\eta_v)/2\lambda_v^2}\left\{\sum_{\ell=1}^{3}\int_{\Omega_\ell}\left\{\sqrt{2\pi}\lambda_v(-\eta_v)\right\}^{-1}e^{-\ln^2(-\eta_v)/2\lambda_v^2}d(-\eta_v)\right\}^{-1}$$

where $j=1,\dots,J$ and $v=1,2$.

APPENDIX F.

MCMC method.

Let $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$ and d be the current parameter values, and $\boldsymbol{\beta}'$, $\boldsymbol{\gamma}'$, $\boldsymbol{\eta}'$ and d' the proposed values. For simplicity, let $\sigma' = \sigma(\boldsymbol{\eta}')$ and $\sigma = \sigma(\boldsymbol{\eta})$. Then our MCMC algorithm consists of the following several steps.

Step 1. Obtain the proposed value $\ln(-\eta_v)$ by simulating $\ln\left(-\eta_v'\right) \sim N\left(\ln(-\eta_v), \tilde{\lambda}_v\right)$ such that $\eta_v \in \cup_{\ell=1}^3 \Omega_\ell$, where $v = 1, 2$.

Calculate $\sigma' = \sigma(\boldsymbol{\eta}')$ using Proposition 1.

Step 2. Obtain the proposed value $\ln\alpha_{0j}' \sim N\left(\ln\alpha_{0j}, \tilde{\sigma}_{0j}^2\right)$ such that $\ln m_0 \le \ln\alpha_{0j}' < \ln a$;

Obtain the proposed value $\ln\alpha_{1j}' \sim N\left(\ln\alpha_{1j}, \tilde{\sigma}_{1j}^2\right)$ such that $\ln\alpha_{1j}' < \ln a$; where $j = 1, ..., J$.

Step 3. Obtain the proposed value $\ln\beta_{1j}' \sim N\left(\ln\beta_{1j}, \tilde{s}_{1j}^2\right)$ such that $\ln\beta_{1j}' < \ln\left(a - \alpha_{1j}'\right)$, where $j = 1, ..., J$.

Step 4. Obtain the proposed value $d' \sim U\{1, ..., d_0\}$.

Step 5. Obtain the proposed thresholds $\boldsymbol{\gamma}'$ as follows:
- Let $a_1 = y_{\min}$, $b = y_{\max}$.
- For $j = 1, ..., J-1$, simulate $\gamma_j' \sim N\left(\gamma_j, \xi_j^2\right)$ such that $\gamma_j' \in (a_j, b)$. That is:
- Simulate $u \sim U(0, 1)$.
- Let $c_0 = (a_j - \gamma_j)/\xi_j, \quad c_1 = (b - \gamma_j)/\xi_j$.
- Let $w = u(\Phi(c_1) - \Phi(c_0)) + \Phi(c_0)$.
- Let $\gamma_j' = \xi_j \Phi^{-1}(w) + \gamma_j$.
- Let $a_{j+1} = \gamma_j'$.

Then γ_j' are random samples from $N\left(\gamma_j, \xi_j^2\right)$ and $y_{\min} < \gamma_1' < \gamma_2' < \cdots < \gamma_{k-1}' < y_{\max}$.

Step 6. Calculate h_t' for $t = d_0 + 1, ..., T$:

$$h_t' = \sum_{j=1}^J \left(\alpha_{0j}' + \alpha_{1j}' y_{t-1}^2 + \beta_{1j}' h_{t-1}'\right) I\left(\gamma_{j-1}' \le y_{t-d}' < \gamma_j'\right).$$

Step 7. Accept the proposed value with probability $\min\{AB, 1\}$, where A and B are given below.

$$A = \frac{\pi\left(\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\eta}', d' | \mathbf{y}_T\right)}{\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d | \mathbf{y}_T)} = \prod_{t=d_0+1}^T \frac{\sigma' Q'(\tau_t, \boldsymbol{\eta}) \sqrt{h_t}}{\sigma Q'(\tau_t', \boldsymbol{\eta}') \sqrt{h_t'}} \frac{\pi\left(\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\eta}', d'\right)}{\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)}$$

$$= \prod_{t=d_0+1}^T \frac{\sigma'\left(\tau_t^{\eta_1-1} + (1-\tau_t)^{\eta_2-1}\right)\sqrt{h_t}}{\sigma\left((\tau_t')^{\eta_1'-1} + (1-\tau_t')^{\eta_2'-1}\right)\sqrt{h_t'}} \prod_{j=1}^J \prod_{i=0}^1 \frac{\alpha_{ij}}{\alpha_{ij}'} e^{-\left(\ln^2\alpha_{ij}' - \ln^2\alpha_{ij}\right)/2\sigma_{ij}^2}$$

$$\times \frac{\beta_{1j}}{\beta_{1j}'} e^{-\left(\ln^2\beta_{1j}' - \ln^2\beta_{1j}\right)/2s_{1j}^2} \frac{\Phi\left(\dfrac{\ln\left(a - \alpha_{1j}\right)}{s_{1j}}\right)}{\Phi\left(\dfrac{\ln\left(a - \alpha_{1j}'\right)}{s_{1j}}\right)} \prod_{v=1}^2 \frac{\eta_v}{\eta_v'} e^{-\left(\ln^2\left(-\eta_v'\right) - \ln^2(-\eta_v)\right)/2\lambda_v^2},$$

$$B = \frac{q(\boldsymbol{\alpha}' \to \boldsymbol{\alpha}) q\left(\boldsymbol{\beta}' \to \boldsymbol{\beta}\right) q(\boldsymbol{\eta}' \to \boldsymbol{\eta})}{q(\boldsymbol{\alpha} \to \boldsymbol{\alpha}') q(\boldsymbol{\beta} \to \boldsymbol{\beta}') q(\boldsymbol{\eta} \to \boldsymbol{\eta}')} \frac{q\left(d' \to d\right)}{q\left(d \to d'\right)} \frac{q(\boldsymbol{\gamma}' \to \boldsymbol{\gamma})}{q(\boldsymbol{\gamma} \to \boldsymbol{\gamma}')},$$

where

$$\frac{q(\boldsymbol{\alpha}' \to \boldsymbol{\alpha})}{q(\boldsymbol{\alpha} \to \boldsymbol{\alpha}')}$$

$$= \prod_{j=1}^{J} \frac{\left(\sqrt{2\pi}\tilde{\sigma}_{0j}\alpha_{0j}\right)^{-1} e^{-\left(\ln\alpha_{0j} - \ln\alpha'_{0j}\right)^2 / 2\tilde{\sigma}_{0j}^2} \left\{ \int_{m_0}^{M} \left(\sqrt{2\pi}\tilde{\sigma}_{0j}\alpha_{0j}\right)^{-1} e^{-\left(\ln\alpha_{0j} - \ln\alpha'_{0j}\right)^2 / 2\tilde{\sigma}_{0j}^2} d\alpha_{0j} \right\}^{-1}}{\left(\sqrt{2\pi}\tilde{\sigma}_{0j}\alpha'_{0j}\right)^{-1} e^{-\left(\ln\alpha'_{0j} - \ln\alpha_{0j}\right)^2 / 2\tilde{\sigma}_{0j}^2} \left\{ \int_{m_0}^{M} \left(\sqrt{2\pi}\tilde{\sigma}_{0j}\alpha'_{0j}\right)^{-1} e^{-\left(\ln\alpha'_{0j} - \ln\alpha_{0j}\right)^2 / 2\tilde{\sigma}_{0j}^2} d\alpha'_{0j} \right\}^{-1}}$$

$$\times \frac{\left(\sqrt{2\pi}\tilde{\sigma}_{1j}\alpha_{1j}\right)^{-1} e^{-\left(\ln\alpha_{1j} - \ln\alpha'_{1j}\right)^2 / 2\tilde{\sigma}_{1j}^2} \left\{ \int_{0}^{a} \left(\sqrt{2\pi}\tilde{\sigma}_{1j}\alpha_{1j}\right)^{-1} e^{-\left(\ln\alpha_{1j} - \ln\alpha'_{1j}\right)^2 / 2\tilde{\sigma}_{1j}^2} d\alpha_{1j} \right\}^{-1}}{\left(\sqrt{2\pi}\tilde{\sigma}_{1j}\alpha'_{1j}\right)^{-1} e^{-\left(\ln\alpha'_{1j} - \ln\alpha_{1j}\right)^2 / 2\tilde{\sigma}_{1j}^2} \left\{ \int_{0}^{a} \left(\sqrt{2\pi}\tilde{\sigma}_{1j}\alpha'_{1j}\right)^{-1} e^{-\left(\ln\alpha'_{1j} - \ln\alpha_{1j}\right)^2 / 2\tilde{\sigma}_{1j}^2} d\alpha'_{1j} \right\}^{-1}}$$

$$= \prod_{j=1}^{J} \frac{\alpha'_{0j}}{\alpha_{0j}} \frac{\alpha'_{1j}}{\alpha_{1j}} \frac{\Phi\left(\frac{\ln M - \ln\alpha_{0j}}{\tilde{\sigma}_{0j}}\right) - \Phi\left(\frac{\ln m_0 - \ln\alpha_{0j}}{\tilde{\sigma}_{0j}}\right)}{\Phi\left(\frac{\ln M - \ln\alpha'_{0j}}{\tilde{\sigma}_{0j}}\right) - \Phi\left(\frac{\ln m_0 - \ln\alpha'_{0j}}{\tilde{\sigma}_{0j}}\right)} \frac{\Phi\left(\frac{\ln a - \ln\alpha_{1j}}{\tilde{\sigma}_{1j}}\right)}{\Phi\left(\frac{\ln a - \ln\alpha'_{1j}}{\tilde{\sigma}_{1j}}\right)}$$

$$\frac{q(\boldsymbol{\beta}' \to \boldsymbol{\beta})}{q(\boldsymbol{\beta} \to \boldsymbol{\beta}')}$$

$$= \prod_{j=1}^{J} \frac{\left(\sqrt{2\pi}\tilde{s}_{1j}\beta_{1j}\right)^{-1} e^{-\left(\ln\beta_{1j} - \ln\beta'_{1j}\right)^2 / 2\tilde{s}_{1j}^2} \left\{ \int_{0}^{a - \alpha_{1j}} \left(\sqrt{2\pi}\tilde{s}_{1j}\beta_{1j}\right)^{-1} e^{-\left(\ln\beta_{1j} - \ln\beta'_{1j}\right)^2 / 2\tilde{s}_{1j}^2} d\beta_{1j} \right\}^{-1}}{\left(\sqrt{2\pi}\tilde{s}_{1j}\beta'_{1j}\right)^{-1} e^{-\left(\ln\beta'_{1j} - \ln\beta_{1j}\right)^2 / 2\tilde{s}_{1j}^2} \left\{ \int_{0}^{a - \alpha'_{1j}} \left(\sqrt{2\pi}\tilde{s}_{1j}\beta'_{1j}\right)^{-1} e^{-\left(\ln\beta'_{1j} - \ln\beta_{1j}\right)^2 / 2\tilde{s}_{1j}^2} d\beta'_{1j} \right\}^{-1}}$$

$$= \prod_{j=1}^{J} \frac{\beta'_{1j}}{\beta_{1j}} \frac{\Phi\left(\frac{\ln\left(a - \alpha'_{1j}\right) - \ln\beta_{1j}}{\tilde{s}_{1j}}\right)}{\Phi\left(\frac{\ln\left(a - \alpha_{1j}\right) - \ln\beta'_{1j}}{\tilde{s}_{1j}}\right)}$$

$$\frac{q(\boldsymbol{\eta}' \to \boldsymbol{\eta})}{q(\boldsymbol{\eta} \to \boldsymbol{\eta}')} = \prod_{v=1}^{2} \frac{\left(\sqrt{2\pi}\tilde{\lambda}_v(-\eta_v)\right)^{-1} e^{-\left(\ln(-\eta_v) - \ln\left(-\eta'_v\right)\right)^2 / 2\tilde{\lambda}_v^2}}{\left(\sqrt{2\pi}\tilde{\lambda}_v\left(-\eta'_v\right)\right)^{-1} e^{-\left(\ln\left(-\eta'_v\right) - \ln(-\eta_v)\right)^2 / 2\tilde{\lambda}_v^2}}$$

$$\times \frac{\left\{ \sum_{\ell=1}^{3} \int_{\Omega_\ell} \left(\sqrt{2\pi}\tilde{\lambda}_v(-\eta_v)\right)^{-1} e^{-\left(\ln(-\eta_v) - \ln\left(-\eta'_v\right)\right)^2 / 2\tilde{\lambda}_v^2} d(-\eta_v) \right\}^{-1}}{\left\{ \sum_{\ell=1}^{3} \int_{\Omega_\ell} \left(\sqrt{2\pi}\tilde{\lambda}_v\left(-\eta'_v\right)\right)^{-1} e^{-\left(\ln\left(-\eta'_v\right) - \ln(-\eta_v)\right)^2 / 2\tilde{\lambda}_v^2} d\left(-\eta'_v\right) \right\}^{-1}}$$

$$= \prod_{v=1}^{2} \frac{\eta'_v}{\eta_v} \frac{C_1}{C_2},$$

where

$$C_1 = \left\{ \Phi\left(\frac{\ln(1/2 - m_0) - \ln(-\eta_v)}{\tilde{\lambda}_v} \right) - \Phi\left(\frac{\ln m_0 - \ln(-\eta_v)}{\tilde{\lambda}_v} \right) \right\}$$
$$+ \left\{ \Phi\left(\frac{\ln(1 - m_0) - \ln(-\eta_v)}{\tilde{\lambda}_v} \right) - \Phi\left(\frac{\ln(1/2 + m_0) - \ln(-\eta_v)}{\tilde{\lambda}_v} \right) \right\}$$
$$+ \left\{ \Phi\left(\frac{\ln(M) - \ln(-\eta_v)}{\tilde{\lambda}_v} \right) - \Phi\left(\frac{\ln(1 + m_0) - \ln(-\eta_v)}{\tilde{\lambda}_v} \right) \right\}$$

$$C_2 = \left\{ \Phi\left(\frac{\ln(1/2 - m_0) - \ln(-\eta_v')}{\tilde{\lambda}_v} \right) - \Phi\left(\frac{\ln m_0 - \ln(-\eta_v')}{\tilde{\lambda}_v} \right) \right\}$$
$$+ \left\{ \Phi\left(\frac{\ln(1 - m_0) - \ln(-\eta_v')}{\tilde{\lambda}_v} \right) - \Phi\left(\frac{\ln(1/2 + m_0) - \ln(-\eta_v')}{\tilde{\lambda}_v} \right) \right\}$$
$$+ \left\{ \Phi\left(\frac{\ln(M) - \ln(-\eta_v')}{\tilde{\lambda}_v} \right) - \Phi\left(\frac{\ln(1 + m_0) - \ln(-\eta_v')}{\tilde{\lambda}_v} \right) \right\}$$

$$\frac{q(d' \to d)}{q(d \to d')} = 1,$$

$$\frac{q(\gamma' \to \gamma)}{q(\gamma \to \gamma')} = \prod_{j=1}^{J-1} \frac{\Phi\left((b - \gamma_j)/\xi_j \right) - \Phi\left((\gamma'_{j-1} - \gamma_j)/\xi_j \right)}{\Phi\left((b - \gamma'_j)/\xi_j \right) - \Phi\left((\gamma_{j-1} - \gamma'_j)/\xi_j \right)},$$

as d is uniformly distributed on D, where $\gamma_0 = \gamma'_0 = y_{\min}$ and $q(z \to z')$ represents the probability density function of z' conditional on z.

APPENDIX G.

Proof of Proposition 5.

If $\tau_0 \leq 0.5$,

$$ES_{t\tau_0} = E(y_t | y_t \leq VaR_{t\tau_0}) = E\left(y_t | y_t \leq \sigma_t Q_{\varepsilon_t}(\tau_0) \right)$$
$$= \int y_t f\left[y_t | y_t \leq \sigma_t Q_{\varepsilon_t}(\tau_0) \right] dy_t$$
$$= \frac{\sigma_t}{\tau_0} \int_0^{\tau_0} Q_{\varepsilon_t}(\tau) d\tau$$

as required. Similarly we have $ES_{t\tau_0} = \frac{\sigma_t}{1 - \tau_0} \int_{\tau_0}^1 Q_{\varepsilon_t}(\tau) d\tau$ if $\tau_0 > 0.5$ as required.

APPENDIX H.

Proof of Proposition 6.

For the proposed model, it follows from Proposition 3 that $ES_{t\tau_0} = \sqrt{h_t} f(\tau_0, \boldsymbol{\eta})$ if $\tau_0 \leq 0.5$ and $ES_{t\tau_0} = \sqrt{h_t} g(\tau_0, \boldsymbol{\eta})$ otherwise, where

$$f(\tau_0, \boldsymbol{\eta}) = \frac{1}{\tau_0 \sigma(\boldsymbol{\eta})} \left\{ \frac{\tau_0^{\eta_1 + 1} - \tau_0}{\eta_1(\eta_1 + 1)} + \frac{(1 - \tau_0)^{\eta_2 + 1} - (1 - \tau_0)}{\eta_2(\eta_2 + 1)} \right\},$$

and

$$g(\tau_0, \boldsymbol{\eta}) = \frac{1}{(1-\tau_0)\sigma(\boldsymbol{\eta})} \left\{ \frac{\tau_0 - \tau_0^{\eta_1+1}}{\eta_1(\eta_1+1)} + \frac{(1-\tau_0) - (1-\tau_0)^{\eta_2+1}}{\eta_2(\eta_2+1)} \right\}.$$

Hence, $R_t = (y_t - ES_{t\tau_0})/\sqrt{h_t} = \varepsilon_t - f(\tau_0, \boldsymbol{\eta})$ if $\tau_0 \leq 0.5$ and $R_t = (y_t - ES_{t\tau_0})/\sqrt{h_t} = \varepsilon_t - g(\tau_0, \boldsymbol{\eta})$ otherwise. As ε_t are iid random variables, we see that R_t are also iid.

Furthermore, it follows from the proof of Proposition 3 that $f(\tau_0, \boldsymbol{\eta})$ is the ES of ε_t when $\tau_0 \leq 0.5$ and $g(\tau_0, \boldsymbol{\eta})$ is the ES of ε_t when $\tau_0 > 0.5$. Let $ES_{t\tau_0}^{\varepsilon_t}$ be the ES of ε_t at level τ_0. Then when $\tau_0 \leq 0.5$, we have $E(R_t | y_t \leq \text{VaR}_{t\tau_0}) = E\left\{ \varepsilon_t - ES_{t\tau_0}^{\varepsilon_t} | \varepsilon_t \leq Q_{\varepsilon_t}(\tau_0, \boldsymbol{\eta}) \right\} = 0$ by the definition of ES. Similarly, when $\tau_0 > 0.5$, we have $E(R_t | y_t \geq \text{VaR}_{t\tau_0}) = E\left\{ \varepsilon_t - ES_{t\tau_0}^{\varepsilon_t} | \varepsilon_t \geq Q_{\varepsilon_t}(\tau_0, \boldsymbol{\eta}) \right\} = 0$ as required.

Similarly, for the benchmark GARCH models, it can also be shown that R_t are iid with zero mean. This completes the proof.

APPENDIX I.

Tables

Tables 4-6 give the BIC values for all benchmark GARCH models. These BIC values show that the minimum value of BIC always corresponds to a model of order (1, 1) in each case.

Table 7 gives the MSE values between the true level τ_0 and the estimated coverage probability of $\hat{\text{VaR}}_{t\tau_0}$ for all models and all data sets.