

A quantile function approach to the distribution of financial returns following TGARCH models

Yuzhi Cai¹, and Guodong Li²

¹ School of Management, Swansea University, United Kingdom

² Department of Statistics & Actuarial Science, University of Hong Kong, China

Address for correspondence: Yuzhi Cai, School of Management, Swansea University, Swansea SA1 8EN, UK.

E-mail: y.cai@swansea.ac.uk.

Phone: (+44) 1792 606865.

Fax: (+44) 1792 295601.

Abstract: We develop a novel quantile function approach to the distribution of financial returns that follow threshold GARCH models. We propose a Bayesian method to do estimation and forecasting simultaneously, which ensures that the density forecasts can take account of the variation of model parameters. The Bayesian method also allows us to handle multiple thresholds easily. We conduct extensive simulation studies and apply our method to Nasdaq returns. The results show that our approach is robust to model specification errors and outperforms some commonly used benchmark models.

Key words: Density forecasting; financial returns; quantile function; threshold GARCH, MCMC.

1 Introduction

GARCH models (Engle, 1982 and Bollerslev, 1986) can elegantly capture the volatility clustering in asset returns which accounts for both their theoretical appeal and their empirical success. However, GARCH models are not able to explain some other features of asset returns. For example, stock returns are negatively correlated with changes in returns volatility, which was first noticed by Black (1976). This is because GARCH models assume that only the magnitude and not the positivity or negativity of unanticipated excess returns affect the volatilities of financial returns. For more detailed discussions on this aspect, see Nelson (1991).

To overcome the limitations of GARCH models, alternative models such as threshold GARCH (TGARCH) models have been developed in the literature. For example, Glosten et al. (1993) proposed a TGARCH model, the so-called GJR-GARCH model, which has been used to study the impact of negative and positive returns on conditional volatility dynamics widely. Zakoian (1994) also proposed a threshold GARCH model (denoted by T-GARCH) for similar purposes. Nelson (1991) developed the exponential GARCH (E-GARCH) model, and Schwert (1990) proposed the absolute value GARCH (AVGARCH) model. Yang and Chang (2008) considered a double-threshold GARCH model with applications to stock and currency markets, and Yu et al. (2010) extended the CAViaR idea (Engle and Manganelli, 2004) to TGARCH

and mixture-GARCH models in order to take into account possible nonlinearity and structural change in the value-at-risk (VaR) process. Chen and So (2006) proposed a threshold heteroscedastic model that integrates threshold nonlinearity and GARCH-type conditional variance for modeling mean and volatility asymmetries in financial markets. An excellent review on threshold time series models in finance was given by Chen et al. (2011). It is seen that these TGARCH models allow us to overcome some of the limitations of GARCH models, and hence in this paper we focus on the TGARCH model proposed by Yu et al. (2010) as it defines a more general TGARCH model.

However, the above TGARCH models mainly focus on the conditional volatility of financial returns, rather than the entire conditional distribution of financial returns. Moreover, the limitations of the existing estimation and forecasting methods with such models have affected the use of these models in practice. Hence it is important to conduct further investigations on the estimation and forecasting entire conditional distributions of financial returns following TGARCH models with multiple thresholds. For these reasons, we develop a novel quantile function approach to the distribution of financial data that follow a TGARCH model.

Quantile *regression* method (Koenker and Bassett, 1978, Koenker, 2005) could be used to study the conditional distribution of a response variable as it allows us to estimate a sequence of conditional quantiles of the response variable. See, e.g. Koenker and Zhao (1996), Taylor (2008) and Cai and Stander (2008). However, if no extra restriction is imposed, the quantile crossing is an unavoidable problem when we perform the quantile inference at more than one quantile level, and the situation becomes more serious when these quantile levels are close to each other; see Bondell et al. (2010).

Moreover, for TGARCH models, it is usually difficult to optimize the target function of a quantile estimation since the iterative function of conditional variances is involved (Xiao and Koenker, 2009) and the situation becomes much worse when there are more than two regimes (Yu et al. 2010).

On the other hand, a quantile *function* approach could also be used to study the conditional distributions of financial returns (Gilchrist, 2000, So and Chung, 2015, Cai, 2016). This approach allows us to estimate the entire distribution function via its quantile function, rather than a sequence of quantiles, of a response variable. It is worth noting that, under certain conditions, quantile functions can be added, multiplied or even transformed to generate another quantile function easily. Hence, from a distributional point of view, it is more convenient to work with quantile functions rather than other equivalent probability functions. Although it has received less attention in the literature, in this paper, we show the potential of this statistical modelling method for studying distributions of financial returns via TGARCH models.

Note that Cai (2016) discussed a general quantile function model but only focused on simple models where no structure change in mean or variance was involved. Estimation of models with a structure change, such as TGARCH models, can be challenging due to multiple thresholds (Yu et al. 2010 and Xiao and Koenker, 2009). Moreover, the work of Cai (2016) did not consider multiple step ahead forecasting for financial returns with quantile function models. Some work can be found in the literature on forecasting with quantile models. For example, Cai (2010) studied forecasting with quantile self-exciting threshold autoregressive time series models. Gaglianone and Lima (2012) proposed a method for constructing density forecasts from quan-

tile regression. Cai et al. (2013) considered density forecasts with quantile double AR models. A common feature of these forecasting methods is that the forecasts are obtained after a model has been estimated. One of the limitations of this two-step procedure is that the forecasts cannot take into account the random variation of the model parameters. In this paper we address the estimation and forecasting issues that have not been addressed by Cai (2016) for TGARCH models due to the importance of these models in finance. We overcome the limitations of the existing methods discussed above by doing estimation and forecasting with TGARCH models simultaneously by using a Bayesian approach.

Bayesian approach for GARCH models is one of the methods that have often been used recently in the literature. For example, Bauwens and Lubrano (2002) used a Bayesian method to study option pricing with asymmetric GARCH models; Ausin and Galeano (2007) considered a Bayesian estimation of the Gaussian mixture GARCH models; Dellaportas and Vrontos (2007) used a Bayesian method to estimate volatility asymmetries with a class of tree structured multivariate GARCH models; Vrontos et al. (2012) conducted a full Bayesian analysis of GARCH and EGARCH models that consists of parameter estimation, model selection, and volatility prediction; and more recently, Jensen and Maheu (2013) proposed a Bayesian approach to semiparametric multivariate GARCH modeling. An excellent review on Bayesian inference methods for GARCH models can be found in Virbickaite et al. (2015). In this paper, we also use a Bayesian approach because it will allow us to do estimation and forecasting with TGARCH models simultaneously and to deal with multiple thresholds easily.

In summary, this paper makes contributions to the literature about estimating and forecasting the entire distributions of financial returns. Specifically, we develop a novel

method that enables us to analyze distributions of financial returns and to obtain density forecasts for future returns that take into account the variations of the model parameters. We focus our discussions on TGARCH models for the reasons discussed above. Our analysis delivers two main results. First, our method is robust to model specification errors. Second, our method not only provides a better fitted model to the financial returns considered in this paper but also provides much improved density forecasts, compared with the benchmark models.

The paper is organized as follows. Section 2 introduces a quantile function TGARCH model and Section 3 discusses the estimation and forecasting method. Results of simulation studies and applications to Nasdaq returns are presented in Sections 4 and 5 respectively. Section 6 provides some further discussions and conclusions.

2 Quantile function TGARCH model

We focus on the general J -regime TGARCH model (Yu et al. 2010):

$$\begin{aligned} x_t &= \varepsilon_t \sqrt{h_t}, \\ h_t &= \sum_{j=1}^J \left(\alpha_0^{(j)} + \sum_{i=1}^{p_j} \alpha_i^{(j)} x_{t-i}^2 + \sum_{\ell=1}^{q_j} \beta_\ell^{(j)} h_{t-\ell} \right) I(\gamma_{j-1} \leq x_{t-d} < \gamma_j), \end{aligned} \quad (2.1)$$

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 orders of the model. Moreover, $\alpha_0^{(j)} > 0$, $\alpha_i^{(j)} \geq 0$, $\beta_\ell^{(j)} \geq 0$. The ε_t 's are usually assumed to be i.i.d. following the $N(0, 1)$ distribution.

Note that model (2.1) does not contain a location process and Yu et al. (2010) pointed out that this model can better explain nonlinear phenomena in the financial

market and it can also be regarded as an extension of the CAViaR model of Engle and Manganelli (2004), which has been used extensively in finance. Moreover, it has been accepted widely that the variance of returns can be predicted using particular time series models but the returns on financial assets may not be predictable at short horizons; see e.g. Granger (1992) and Franses and Van Dijk (1996). Therefore, we follow this trend and focus on model (2.1). However, it is worth noting that the methodology developed in this paper can be extended to a model that also contains a location process.

Now we extend model (2.1) by introducing the quantile function of ε_t , denoted by $Q(\tau, \boldsymbol{\eta})$, into the model. It follows from the definition of a quantile function that we must have $\tau = P\{\varepsilon_t \leq Q(\tau, \boldsymbol{\eta})\}$ for all $\tau \in (0, 1)$. Hence, if x_t follows model (2.1), then $\tau = P\{\varepsilon_t \leq Q(\tau, \boldsymbol{\eta})\} = P\{\varepsilon_t \sqrt{h_t} \leq Q(\tau, \boldsymbol{\eta}) \sqrt{h_t} \mid \mathbf{x}_{t-1}\} = P\{x_t \leq Q(\tau, \boldsymbol{\eta}) \sqrt{h_t} \mid \mathbf{x}_{t-1}\}$, i.e. the conditional quantile function of x_t is given by

$$\begin{aligned} Q_{x_t}(\tau \mid \mathbf{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d, J, \mathbf{p}, \mathbf{q}, \mathbf{x}_{ini}, \mathbf{h}_{ini}) &= Q(\tau, \boldsymbol{\eta}) \sqrt{h_t} \\ &= Q(\tau, \boldsymbol{\eta}) \sqrt{\sum_{j=1}^J \left(\alpha_0^{(j)} + \sum_{i=1}^{p_j} \alpha_i^{(j)} x_{t-i}^2 + \sum_{\ell=1}^{q_j} \beta_\ell^{(j)} h_{t-\ell} \right) I(\gamma_{j-1} \leq x_{t-d} < \gamma_j)}, \end{aligned} \quad (2.2)$$

where

$$\begin{aligned} \boldsymbol{\gamma} &= \{\gamma_1, \dots, \gamma_{J-1}\}, \quad \boldsymbol{\beta} = \{\alpha_0^{(j)}, \dots, \alpha_{p_j}^{(j)}, \beta_1^{(j)}, \dots, \beta_{q_j}^{(j)}, j = 1, \dots, J\}, \\ \mathbf{p} &= \{p_1, \dots, p_J\}, \quad \mathbf{q} = \{q_1, \dots, q_J\}, \quad L = \max\{d_{\max}, p_j, q_j, j = 1, \dots, J\} \\ \mathbf{x}_t &= \{x_t, x_{t-1}, \dots, x_{L+1}\}, \quad \mathbf{x}_{ini} = \{x_L, \dots, x_1\}, \quad \mathbf{h}_{ini} = \{h_L, \dots, h_1\}, \end{aligned}$$

and d_{\max} is the maximum value of the delay parameter d . We call model (2.2) the quantile function threshold GARCH model, denoted by QF-TGARCH.

Model (2.2) tells us that, conditional on the financial returns in the past, the proba-

bility for today's return to be less than $Q(\tau, \boldsymbol{\eta})\sqrt{h_t}$ equals to τ . So if we let $\tau = 0.5$, then the conditional median value of today's return is given by $Q(0.5, \boldsymbol{\eta})\sqrt{h_t}$. If we let $\tau = 0.025$, then there is a 95% chance that today's return lies between $Q(0.025, \boldsymbol{\eta})\sqrt{h_t}$ and $Q(0.975, \boldsymbol{\eta})\sqrt{h_t}$, given the past returns. Generally, when τ varies between 0 and 1 we have the entire conditional distribution of today's return.

Note that if $Q(\tau, \boldsymbol{\eta})$ in model (2.2) is the quantile function of $N(0, 1)$, then both models (2.1) and (2.2) are equivalent. Hence, model (2.2) extends model (2.1). Moreover, compared with model (2.1), model (2.2) focuses on how past returns affects the *entire conditional distribution* of current returns, and compared with *quantile regression* approach, our approach does not suffer from the quantile crossing problem, because model (2.2) defines a proper conditional quantile function for x_t . Hence, as τ varies, any two estimated quantile curves will not cross.

So and Chung (2015) proposed a two-step procedure for estimating conditional quantiles by combining the empirical quantile function of ε_t and a quasi maximum likelihood estimator of model parameters. Their approach avoids specifying a specific distribution for ε_t but is again difficult to deal with multiple thresholds and the delay parameter in the estimation of TGARCH models.

Therefore, we adopt a parametric approach for which we need to specify a distribution for ε_t . It is well-known that financial returns have some stylized features including volatility clustering, tail properties and extreme fluctuations. A QF-TGARCH model with a normal distribution may not be able to explain these distributional properties of financial returns. We need to develop a more useful model from (2.2) that allows us to study conditional distributions of financial returns and that is robust to model specification errors.

The work of Freimer et al. (1988) and Fournier et al. (2007) suggest that the generalized Lambda distribution (GLD) is a good candidate for our model because it can provide a very accurate approximation to many standard distributions such as normal, log-normal, Weibull, t-, F- and skewed t-distributions as well as others (see Figure 1 for some evidence).

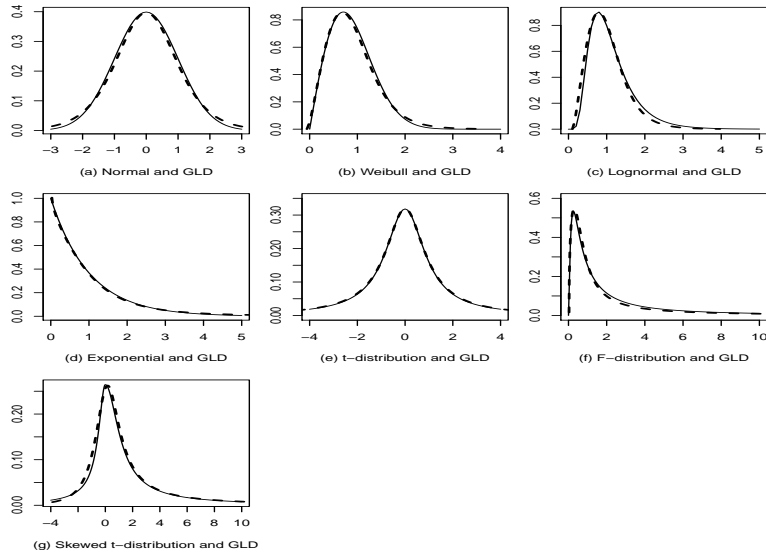


Figure 1: Density functions of seven commonly used distributions (continuous curves) and those of generalized Lambda distributions (dashed curves).

Specifically, if we let

$$Q(\tau, \boldsymbol{\eta}) = \frac{\tau^{\eta_1} - 1}{\eta_1} - \frac{(1 - \tau)^{\eta_2} - 1}{\eta_2}, \quad \eta_1 < 0, \eta_2 < 0, \quad (2.3)$$

then model (2.2) defines a GLD (Freimer et al., 1988, Fournier et al., 2007) with location 0, scale $\sqrt{h_t}$ and right and left tails controlled by η_1 and η_2 respectively. If $\eta_1 < \eta_2$ the distribution is skewed to the left and if $\eta_2 < \eta_1$ it is skewed to the right. When $\eta_1 = \eta_2$, it is symmetric and in this case it becomes the Turkey lambda

distribution. Moreover, the parameters η_1 and η_2 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 (Gilchrist, 2000). Hence, the use of the GLD will make model (2.2) robust to model specification errors, which is also confirmed by our simulation studies.

Before ending this section, it is worth mentioning that apart from the GLD, there exist many other quantile functions that could also be used for statistical modelling. For example, the five-parameter lambda distribution is defined by $Q(\tau) = \lambda + (\eta/2)\{(1 - \delta)\tau^\alpha - (1 + \delta)(1 - \tau)^\beta\}$ and the Burr XII distribution is given by $Q(\tau) = \{-1 + (1 - \tau)^{-\alpha}\}^\beta$. Detailed definitions of these and many other quantile functions can be found in Gilchrist (2000).

In the next section we discuss our estimation and forecasting method for the QF-TGARCH model (2.2) with $Q(\tau, \boldsymbol{\eta})$ defined by (2.3).

3 Estimation and forecasting method

Let the observed financial return series be $\mathbf{x}_T = \{x_T, \dots, x_1\}$, and let $\mathbf{x}_{pre} = \{x_{T+M}, \dots, x_{T+1}\}$ be the first M future returns that we want to predict. Hence, \mathbf{x}_{pre} is a latent parameter vector of the model. We assume that both \mathbf{x}_T and \mathbf{x}_{pre} follow the same model. Our task is to estimate the model parameters and to predict future returns \mathbf{x}_{pre} simultaneously.

It is worth reemphasizing that model (2.2) defined by (2.3) says that the conditional distribution of x_t is GLD. In the literature, different methods have been de-

veloped for the estimation of GLD models. For example, Karian et al. (1996) discussed the method of moment, Qzturk and Dale (1985) considered the method of least squares, Karian and Dudewicz (1999) proposed the method of percentiles, King and MacGillivray (1999) developed a starship estimation method, Gilchrist (2000) discussed the MLE method and Su (2007a) proposed a two-step procedure that combines the methods of moment or percentile and the MLE to fit the GLD to data. Su (2007b) also pointed out that the MLE method is not only more efficient but also tends to produce GLD that has closer first four moments to the data set.

In our case, apart from the parameters of our GLD model, we also need to deal with forecasting and multiple thresholds issues, which cannot be easily achieved by the above methods. Hence we consider a Bayesian approach to estimation and forecasting, which allows us to do the estimation and forecasting *simultaneously* and to deal with multiple thresholds easily. Our approach ensures that the forecasts are able to take into account the variation of the model parameters.

It is worth mentioning that our estimation and forecasting method does not estimate the values of J , \mathbf{p} and \mathbf{q} . Their values will be determined according to Bayes factors. Hence, we will remove them, as well as the initial values \mathbf{x}_{ini} and \mathbf{h}_{ini} , from our formulae below to simplify the notations.

3.1 Posterior distribution

To estimate the parameters $\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}$ and d simultaneously, we need to derive the posterior distribution function of these parameters. Note that the posterior density function of the parameters is given by $\pi(\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d \mid \mathbf{x}_T) \propto \pi(\mathbf{x}_T \mid \mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)\pi(\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)$, where the first term is simply the likelihood of the

observed returns \mathbf{x}_T and the second term is a prior density function of $\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}$ and d . However, this setting ignores the dependency between \mathbf{x}_{pre} and \mathbf{x}_T , which is not desirable.

Hence we rewrite the posterior density function as $\pi(\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d \mid \mathbf{x}_T) \propto \pi(\mathbf{x}_{pre} \mid \mathbf{x}_T, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d) \pi(\mathbf{x}_T \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d) \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)$, where the first two terms can be easily written out explicitly by using the dependence structure between x_t 's, and the last term is a prior density function of $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}$ and d .

It follows from model (2.2) that, for each x_t there exists $\tau_t \sim U(0, 1)$ such that $x_t = Q(\tau_t, \boldsymbol{\eta})\sqrt{h_t}$. Moreover, we 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 $f(y) = (d(Q(\tau))/d\tau)^{-1}$. Therefore, the posterior density function becomes

$$\pi(\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d \mid \mathbf{x}_T) \propto \prod_{t=L+1}^{T+M} \{Q'(\tau, \boldsymbol{\eta})\sqrt{h_t}\}^{-1} \Big|_{\tau=\tau_t} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d).$$

where $Q'(\tau, \boldsymbol{\eta}) = dQ(\tau, \boldsymbol{\eta})/d\tau$. Moreover, when $Q(\tau, \boldsymbol{\eta})$ is defined by (2.3), the posterior density function of the parameters is given by

$$\begin{aligned} & \pi(\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d \mid \mathbf{x}_T) \\ & \propto \prod_{t=L+1}^{T+M} \{(\tau_t^{\eta_1-1} + (1 - \tau_t)^{\eta_2-1})\sqrt{h_t}\}^{-1} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d). \end{aligned} \tag{3.1}$$

In practice, it is reasonable to assume that the thresholds $\gamma_j \in (x_{\min}, x_{\max})$ for all possible j , where $x_{\min} = \min\{x_1, \dots, x_T\}$ and $x_{\max} = \max\{x_1, \dots, x_T\}$. Moreover, let $\alpha_0^{(j)} \in [e_0, \infty)$ for all $1 \leq j \leq J$, where $e_0 > 0$ is a very small number such that its effect can be ignored, and we set $e_0 = 10^{-30}$ in the calculations throughout this paper.

Then it is easy to see that $\tau^{\eta_1-1} + (1 - \tau)^{\eta_2-1} \geq 1$ for all $\eta_1, \eta_2 < 0$ and $\tau \in (0, 1)$, and

$$Q'(\tau, \boldsymbol{\eta})\sqrt{h_t} = \{\tau^{\eta_1-1} + (1 - \tau)^{\eta_2-1}\} \sqrt{h_t} \geq \sqrt{\sum_{j=1}^J e_0 I(\gamma_{j-1} \leq x_{t-d} < \gamma_j)}.$$

It follows from this inequality that, without further constrains on the parameters, once $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)$ is a well-defined prior density function, the posterior density function of $\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}$ and d is also well defined from equation (3.1).

3.2 Prior distribution

For the sake of easy illustration and calculation, 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 (x_{\min}, x_{\max}) , d is uniformly distributed on $[1, d_{\max}]$,

$$\pi(\alpha_i^{(j)}) = \frac{1}{\sqrt{2\pi} \alpha_i^{(j)} \sigma_{ij}} e^{-\ln^2 \alpha_i^{(j)} / 2\sigma_{ij}^2}, \quad \pi(\beta_\ell^{(j)}) = \frac{1}{\sqrt{2\pi} \beta_\ell^{(j)} s_{\ell j}} e^{-\ln^2 \beta_\ell^{(j)} / 2s_{\ell j}^2},$$

$$\pi(\eta_v) = \frac{1}{\sqrt{2\pi} (-\eta_v) \lambda_v} e^{-\ln^2(-\eta_v) / 2\lambda_v^2},$$

and the σ_{ij} 's, $s_{\ell j}$'s and λ_v 's are the corresponding scale parameters.

Clearly, the prior distribution is well defined on the parameter space of the posterior distribution. The strength of the prior information involved in the estimation and forecasting procedure is controlled by the scale parameters. For example, σ_{ij} controls the strength of the prior information on $\alpha_i^{(j)}$. This is because, given σ_{ij} , the standard deviation of $\alpha_i^{(j)}$ is given by $\sqrt{e^{\sigma_{ij}^2}(e^{\sigma_{ij}^2} - 1)}$. A small (large) standard deviation represents strong (weak) prior information on $\alpha_i^{(j)}$. As in practice we usually do not have any information on $\alpha_i^{(j)}$ and all other parameters, we deliberately let $\sigma_{ij} = s_{\ell j} = \lambda_v = 2$. This means that the standard deviation of these parameters are all 54.1, which is very large. As a result, almost no prior information is used in the estimation and forecasting procedure and we almost completely rely on the data to tell us the behavior of financial returns. This suggests that for the same data the estimation results do not depend on the choice of $\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)$.

3.3 MCMC algorithm

It is clear that the posterior distribution is not a standard distribution and it also contains latent variables, which suggest that a Markov chain Monte Carlo (MCMC) method would be appropriate for the estimation and forecasting. A Gibbs sampling method is one of the popular methods for parameter estimation. However, in our case it is difficult to obtain marginal distributions for each parameter or for blocks of parameters. Hence, to facilitate the use of the proposed model, we adopt the Metropolis-Hastings MCMC method for the parameter estimation. Denote by \mathbf{x}_{pre} , $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}$ and d the current values of parameters, and by \mathbf{x}'_{pre} , $\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\eta}'$ and d' the proposed values. Then we suggest the following MCMC algorithm to estimate the parameters and to do forecasting simultaneously.

Step 1. Obtain the proposed value of \mathbf{x}'_{pre} , $\boldsymbol{\beta}', \boldsymbol{\eta}'$: For $j = 1, \dots, J$, $i = 0, \dots, p_j$, $\ell = 1, \dots, q_j$, and $v = 1, 2$, simulate

$$\begin{aligned} \ln \alpha_i^{(j)'} &\sim N\left(\ln \alpha_i^{(j)}, \tilde{\sigma}_{ij}^2\right), & \ln \beta_\ell^{(j)'} &\sim N\left(\ln \beta_\ell^{(j)}, \tilde{s}_{\ell j}^2\right), \\ \ln(-\eta'_v) &\sim N\left(\ln(-\eta_v), \tilde{\lambda}_v^2\right). \end{aligned}$$

Step 2. Obtain the proposed d' : Simulate $d' \sim U[1, d_{\max}]$.

Step 3. Obtain the proposed thresholds $\boldsymbol{\gamma}'$: Simulate $\boldsymbol{\gamma}'$:

- Let $a_1 = x_{\min}$, $b = x_{\max}$.
- For $j = 1, \dots, J - 1$, simulate $\gamma'_j \sim N(\gamma_j, \xi_j^2)$ such that $\gamma'_j \in (a_j, b)$. That is:
 - Simulate $u \sim U(0, 1)$.

- Let $c_0 = (a_j - \gamma_j)/\xi_j$, $c_1 = (b - \gamma_j)/\xi_j$.
- 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 $N(\gamma_j, \xi_j^2)$ and $x_{\min} < \gamma'_1 < \gamma'_2 < \dots < \gamma'_{k-1} < x_{\max}$.

Step 4. Construct $[\gamma'_{j-1}, \gamma'_j)$ using the proposed γ'_j .

Step 5. For $t = L + 1, \dots, T$, calculate h'_t :

$$h'_t = \sum_{j=1}^J \left(\alpha_0^{(j)'} + \sum_{i=1}^{p_j} \alpha_i^{(j)'} x_{t-i}^2 + \sum_{\ell=1}^{q_j} \beta_\ell^{(j)'} h'_{t-\ell} \right) I(\gamma'_{j-1} \leq x'_{t-d} < \gamma'_j).$$

Step 6. For $m = 1, \dots, M$, simulate x'_{T+m} :

- Simulate $\tau'_{T+m} \sim U(0, 1)$.
- Calculate

$$h'_{T+m} = \sum_{j=1}^J \left(\alpha_0^{(j)'} + \sum_{i=1}^{p_j} \alpha_i^{(j)'} x_{T+m-i}^2 + \sum_{\ell=1}^{q_j} \beta_\ell^{(j)'} h'_{T+m-\ell} \right) I(\gamma'_{j-1} \leq x'_{T+m-d'} < \gamma'_j),$$

where $x'_{T+m-i} = x_{T+m-i}$ if $T + m - i \leq T$.

- Let $x'_{T+m} = Q(\tau'_{T+m}, \gamma') \sqrt{h'_{T+m}}$.

Step 7. Accept the proposed value with probability $\min\{ABCDE, 1\}$, where

$$\begin{aligned} A &= \frac{\pi(\mathbf{x}'_{pre}, \boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\eta}', d' \mid \mathbf{x}_T, J, \mathbf{p}, \mathbf{q}, \mathbf{x}_{ini}, \mathbf{h}_{ini})}{\pi(\mathbf{x}_{pre}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d \mid \mathbf{x}_T, J, \mathbf{p}, \mathbf{q}, \mathbf{x}_{ini}, \mathbf{h}_{ini})} \\ &= \prod_{t=L+1}^{T+M} \frac{Q'(\tau_t, \boldsymbol{\eta}) \sqrt{h_t}}{Q'(\tau'_t, \boldsymbol{\eta}') \sqrt{h'_t}} \frac{\pi(\boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\eta}', d' \mid J, \mathbf{p}, \mathbf{q}, \mathbf{x}_{ini}, \mathbf{h}_{ini})}{\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d \mid J, \mathbf{p}, \mathbf{q}, \mathbf{x}_{ini}, \mathbf{h}_{ini})} \end{aligned}$$

$$\begin{aligned}
&= \prod_{t=L+1}^{T+M} \frac{(\tau_t^{\eta_1-1} + (1-\tau_t)^{\eta_2-1}) \sqrt{h_t}}{((\tau_t')^{\eta_1'-1} + (1-\tau_t')^{\eta_2'-1}) \sqrt{h_t'}} \prod_{j=1}^J \prod_{i=0}^{p_j} \frac{\alpha_i^{(j)}}{\alpha_i^{(j)'}} e^{-(\ln^2 \alpha_i^{(j)'} - \ln^2 \alpha_i^{(j)})/2\sigma_{ij}^2} \\
&\quad \times \prod_{\ell=1}^{q_j} \frac{\beta_\ell^{(j)}}{\beta_\ell^{(j)'}} e^{-(\ln^2 \beta_\ell^{(j)'} - \ln^2 \beta_\ell^{(j)})/2s_{\ell j}^2} \times \prod_{v=1}^2 \frac{\eta_v}{\eta_v'} e^{-(\ln^2(-\eta_v') - \ln^2(-\eta_v))/2\lambda_v^2}, \\
B &= \frac{q(\boldsymbol{\alpha}' \rightarrow \boldsymbol{\alpha})q(\boldsymbol{\beta}' \rightarrow \boldsymbol{\beta})q(\boldsymbol{\eta}' \rightarrow \boldsymbol{\eta})}{q(\boldsymbol{\alpha} \rightarrow \boldsymbol{\alpha}')q(\boldsymbol{\beta} \rightarrow \boldsymbol{\beta}')q(\boldsymbol{\eta} \rightarrow \boldsymbol{\eta}')} = \prod_{j=1}^J \prod_{i=0}^{p_j} \frac{\alpha_i^{(j)'}}{\alpha_i^{(j)}} \prod_{\ell=1}^{q_j} \frac{\beta_\ell^{(j)'}}{\beta_\ell^{(j)}} \prod_{v=1}^2 \frac{\eta_v'}{\eta_v},
\end{aligned}$$

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

as d is uniformly distributed on $[1, d_{\max}]$,

$$D = \frac{q(\boldsymbol{\gamma}' \rightarrow \boldsymbol{\gamma})}{q(\boldsymbol{\gamma} \rightarrow \boldsymbol{\gamma}')} = \prod_{j=1}^{J-1} \frac{\Phi((b - \gamma_j)/\xi_j) - \Phi((\gamma'_{j-1} - \gamma_j)/\xi_j)}{\Phi((b - \gamma'_j)/\xi_j) - \Phi((\gamma_{j-1} - \gamma'_j)/\xi_j)}$$

with $\gamma_0 = x_{\min}$,

$$E = \frac{q(\mathbf{x}'_{pre} \rightarrow \mathbf{x}_{pre})}{q(\mathbf{x}_{pre} \rightarrow \mathbf{x}'_{pre})} = 1$$

as τ'_{T+m} is simulated uniformly on $(0, 1)$, and $q(z \rightarrow z')$ represents the probability density function of z' conditional on z .

By repeating these steps multiple times, a Markov chain of parameters is hence generated. The theory of Markov chains (see, e.g. O'Hagan and Forster, 2004) guarantees that the equilibrium distribution of the Markov chain is the posterior distribution of parameters and the convergence does not depend on the initial values required by a MCMC method. As a result, we may collect posterior samples of parameters from the Markov chain after a burn-in period, and use them for further statistical inferences.

It is worth noting that it is important to design each step of a sampling scheme carefully if one uses the Metropolis-Hastings method. For example, in our case, we need to ensure that the proposed thresholds should satisfy the monotone condition, i.e. $x_{\min} < \gamma'_1 < \gamma'_2 < \dots < \gamma'_{k-1} < x_{\max}$. Among various ways of achieving this, we found that the method we proposed to use in this step is the most efficient one. It is seen that the first five steps of our sampling scheme is for model parameters, in which Steps 2 and 3 are for the delay parameter and thresholds respectively, and Step 6 is for forecasting. Clearly, the sampling scheme makes it very easy to deal with multiple thresholds, the delay parameter and the latent parameters \mathbf{x}_{pre} . Hence, it provides a useful approach that allows us to overcome the limitations of the MLE method.

Our simulation studies show that the average acceptance rate is 0.3735 and the average computational time is about 152.0323 seconds per 10^5 iterations on a laptop (Intel(R) Core™ i5). Our experience with the sampling scheme suggests that the algorithm converges quickly: a burn-in period of the first 5000 iterations is enough for our simulation and application studies.

For the model selection, we use Bayes factors suggested by Koop and Potter (2003).

The Bayes factor may be defined by

$$\frac{p((J, \mathbf{p}, \mathbf{q}) \mid \mathbf{x}_T)}{p((J', \mathbf{p}', \mathbf{q}') \mid \mathbf{x}_T)} = \frac{p(\mathbf{x}_T \mid (J, \mathbf{p}, \mathbf{q}))}{p(\mathbf{x}_T \mid (J', \mathbf{p}', \mathbf{q}'))},$$

where $(J, \mathbf{p}, \mathbf{q})$ and $(J', \mathbf{p}', \mathbf{q}')$ represent two models with different order and number of thresholds. The best model corresponds to the largest value of $p((J, \mathbf{p}, \mathbf{q}) \mid \mathbf{x}_T)$, hence the largest $p(\mathbf{x}_T \mid (J, \mathbf{p}, \mathbf{q}))$, which can be estimated by (see Gelfand and Dey, 1994)

$$\begin{aligned} & \{p(\mathbf{x}_T \mid (J, \mathbf{p}, \mathbf{q}))\}^{-1} \approx U^{-1} \\ & \times \sum_{u=1}^U g((\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)^{(u)}) \{p(\mathbf{x}_T \mid (J, \mathbf{p}, \mathbf{q}), (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)^{(u)}) p((\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)^{(u)} \mid (J, \mathbf{p}, \mathbf{q}))\}^{-1}, \end{aligned}$$

where $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d)^{(u)}$ is the u th posterior sample, $g(\cdot)$ is an arbitrary density function and U is the number of posterior samples collected.

We will also use several other methods to compare a couple of good models according to Bayes factors. These methods require us to check the residuals of the estimated models, the empirical coverage probabilities of the estimated quantile curves, forecasting performance and financial interpretations. By combining the Bayes factors with these methods, we are able to identify a model that is not only statistically sound but also financially meaningful.

4 Simulation studies

We conduct two simulation experiments to evaluate the performance of the proposed methodology. The first experiment is for two QF-TGARCH models with two and three regimes respectively. In this experiment the volatility persistence of the first model is stronger than the second one. Hence, different levels of volatility persistence can also be considered. The second experiment is to evaluate the robustness of the proposed approach to model specification errors. This experiment also involves two data generating models, both of them have high volatility persistence level and both have a fat-tailed t-distribution for their innovation terms, but we will fit a QF-TGARCH model to the data generated from the two models.

In the first simulation experiment, the first data generating process is

$$Q_{x_t}(\tau \mid \mathbf{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d) = \left(\frac{\tau^{-0.06} - 1}{-0.06} - \frac{(1 - \tau)^{-0.01} - 1}{-0.01} \right) \sqrt{h_t}, \quad (4.1)$$

with

$$h_t = \begin{cases} 0.02 + 0.05x_{t-1}^2 + 0.8h_{t-1}, & x_{t-1} < 0, \\ 0.06 + 0.05x_{t-1}^2 + 0.85h_{t-1}, & x_{t-1} \geq 0, \end{cases} \quad (4.2)$$

and the second data generating process is

$$Q_{x_t}(\tau | \mathbf{x}_{t-1}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, d) = \left(\frac{\tau^{-0.06} - 1}{-0.06} - \frac{(1 - \tau)^{-0.12} - 1}{-0.12} \right) \sqrt{h_t}, \quad (4.3)$$

where

$$h_t = \begin{cases} 0.1 + 0.2x_{t-1}^2 + 0.08h_{t-1}, & x_{t-2} < 0, \\ 0.25 + 0.15x_{t-1}^2 + 0.15h_{t-1}, & 0 \leq x_{t-2} < 0.2, \\ 0.9 + 0.1x_{t-1}^2 + 0.3h_{t-1}, & x_{t-2} \geq 0.2. \end{cases} \quad (4.4)$$

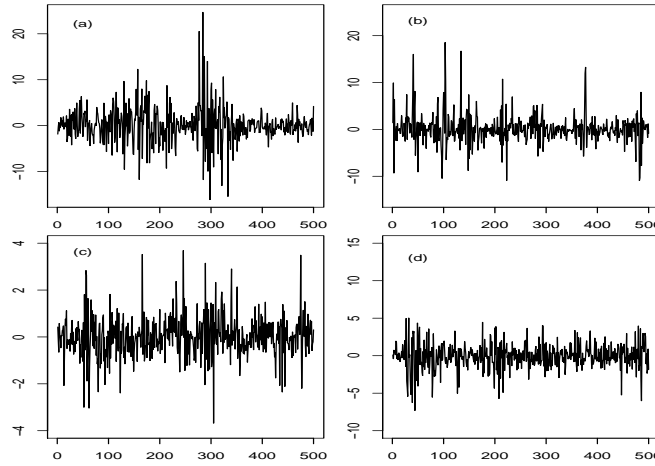


Figure 2: Time series plots of the simulated series from (a) model (4.1), (b) model (4.3), (c) model (4.5) and (d) model (4.6).

We independently simulated 200 sequences, each of size 500, from (4.1) and (4.3) respectively. We use 500 as our sample size because knowing a method that works well in small samples is of practical importance. Figure 2 (a) and (b) present the time series plots of the first simulated series from these two models, both of which

show some common features of financial time series, such as extremes and volatility clustering.

We applied the MCMC algorithm to each of the 200 simulated series obtained from each model. The required initial values for the MCMC algorithm were obtained by simulating β , γ , η and d uniformly on $(0, 1)$, (q_1, q_2) , $(-1, 0)$ and $\{1, 2, 3\}$ respectively, where q_1 and q_2 are the lower and upper 2.5% percentiles of the data respectively. By checking the time series plots (not shown to save space) of the parameter sequences we found that a burn-in period of the first 5000 iterations is appropriate, after which the posterior samples were collected. For each run, corresponding to each simulated sequence, we recorded the Bayesian estimate and an associated 95% credible interval for each parameter, resulting in 200 estimates for each of them. Figure 3

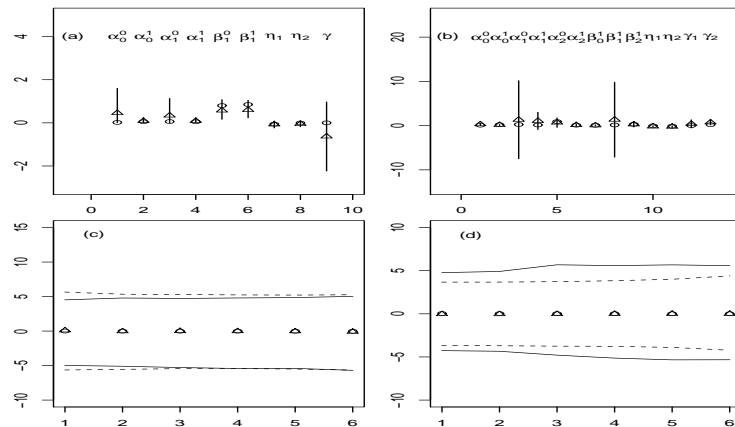


Figure 3: Overall performance in the first experiment. Upper row: true (\circ) and estimated (\triangle) parameters and the corresponding 95% credible intervals (vertical lines); Lower row: true (\circ , continuous curves) and estimated (\triangle , dashed curves) median forecasts and the central 95% probability ranges determined by the conditional distribution of x_{T+m} .

(a) and (b) show the average of these parameters (denoted by Δ) and credible intervals (vertical lines) for models (4.1) and (4.3) respectively. We see that all the true values (denoted by \circ) are well within the associated credible intervals. In fact the mean squared errors (MSE) between the true and estimated parameter values are 0.0876 and 0.3059 for models (4.1) and (4.3) respectively.

For each run we also recorded the median forecasts and the central 95% probability ranges determined by the estimated conditional distribution of x_{T+m} . Figure 3 (c) and (d) show the average of these median forecasts (Δ) and the central probability ranges (dashed) for models (4.1) and (4.3) respectively. It is seen that the median forecasts are also very close to those of the true medians (\circ). In fact the MSEs between them are 0.0025 and 0.0021 for models (4.1) and (4.3) respectively. It is also seen that, compared with the true ones, the estimated central probability ranges of the conditional distribution of x_{T+m} are also satisfactory.

We now check the residuals of the estimated models. Let $\hat{r}_{it} = x_t / \sqrt{\hat{h}_{it}}$ be the standardized residuals calculated by using the Bayesian estimates for the i th simulated series, where $i = 1, \dots, 200$. For both models we know the theoretical distribution of the error term, i.e. $Q(\tau, \boldsymbol{\eta})$. Hence we conducted the Kolmogorov-Smirnov (K-S) test to check the difference between the distributions of $\hat{r}_{it} = x_t / \sqrt{\hat{h}_{it}}$ and $Q(\tau, \boldsymbol{\eta})$ for each i . We found that, for both models, none of the tests rejected the null hypothesis of the K-S test. This further confirms that the performance of our method is satisfactory in this experiment.

In the second experiment, we consider the following two TGARCH models:

$$x_t = \varepsilon_t \sqrt{h_t}, \quad \varepsilon_t \sim t(5), \quad h_t \text{ is defined as in (4.2),} \quad (4.5)$$

and

$$x_t = \varepsilon_t \sqrt{h_t}, \quad \varepsilon_t \sim t(10), \quad (4.6)$$

where

$$h_t = \begin{cases} 0.1 + 0.2x_{t-1}^2 + 0.8h_{t-1}, & x_{t-2} < 0, \\ 0.02 + 0.05x_{t-1}^2 + 0.85h_{t-1}, & 0 \leq x_{t-2} < 0.2, \\ 0.03 + 0.01x_{t-1}^2 + 0.9h_{t-1}, & x_{t-2} \geq 0.2. \end{cases}$$

Hence, both models generate data that have high volatility persistence and are fat-tailed. We generated 200 time series, each of size 500, from models (4.5) and (4.6) respectively. Figure 2 (c) and (d) show the first simulated series from these models. In this experiment, we fit a QF-TGARCH model with $Q(\tau, \boldsymbol{\eta})$ defined by (2.3), rather than the correct TGARCH model, to each series. Hence the errors of the models are misspecified and we are not able to compare the estimated parameter values with the true ones. However, we are able to check the robustness of our model with respect to forecasting performance in this case. To achieve this, we saved the median forecasts and the central 95% probability range of the density forecasts for each series and compared them with those obtained from the true models (4.5) and (4.6) respectively.

Our results show that the MSE values between median forecasts and true ones are 0.0004 and 0.0028 for models (4.5) and (4.6) respectively, while the MSE values between the true and predicted lower bounds of the probability ranges are 0.4716 and 0.4410 for models (4.5) and (4.6) respectively, and those for the upper bounds are 0.4060 and 0.4588 for models (4.5) and (4.6) respectively. In summary, the simulation results suggest that our method works well for both parameter estimation and forecasting.

5 Application

5.1 The data

We consider the Nasdaq daily returns from 3 January 2007 to 8 June 2012, denoted by $\{y_t\}$, which contain 1370 observations. We use the first $T = 1364$ returns for model estimation and forecasting and leave the last six observations for evaluating our multiple-step ahead forecasts for the returns in the next 6 days (about a week ahead). As a result, the observation period for estimation spans from 3 January 2007 to 31 May 2012.

Figure 4 (a)-(c) give the time series plots of Nasdaq daily closing prices and returns as well as the sample autocorrelation function (ACF) of the returns. It is seen that the sample ACF stands out at lag one significantly, and hence an ARMA model was first considered,

$$y_t = \hat{a} + \hat{b} y_{t-1} + x_t = 0.013 - 0.0995 y_{t-1} + x_t, \quad (5.1)$$

where the standard errors of the estimated parameters are given by $\text{SE}(\hat{a}) = 0.041$ and $\text{SE}(\hat{b}) = 0.027$. The sample ACF of the residuals is presented in Figure 4 (d), and we may consider $\{x_t\}$ for modelling.

It is worth noting that the proposed QF-TGARCH model allows us to study the asymmetric relation between stock returns and volatility changes. As the GJR-GARCH model of Glosten et al. (1993), T-GARCH model of Zakoian (1994), E-GARCH model of Nelson (1991) and AVGARCH model of Schwert (1990) also allow us to do so, we will use these models as our benchmark models for comparison purposes. Moreover, we let the innovation term of the models follows a skewed t-distribution

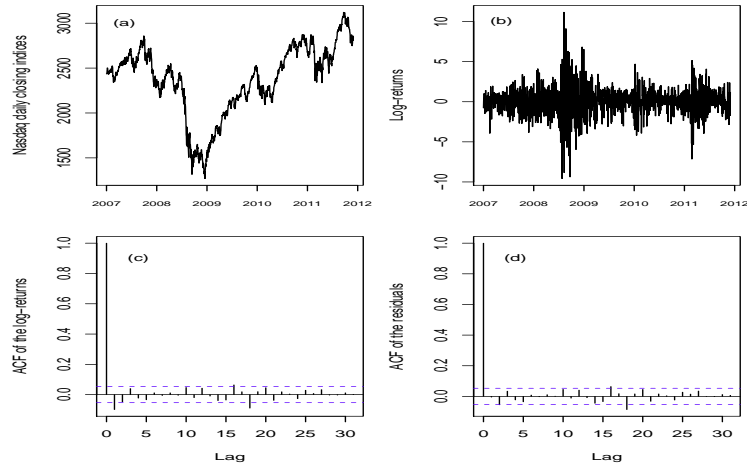


Figure 4: First row: time series plots of Nasdaq daily closing prices and their returns from 3 January 2007 to 8 June 2012. Second row: the sample ACF of $\{y_t\}$ and $\{x_t\}$.

due to its popularity in the literature. It will become clear later in the paper that, the proposed QF-TGARCH model defined by a GLD distribution outperforms the above benchmark models.

5.2 Estimated QF-TGARCH models

We consider the QF-TGARCH models with two and three regimes ($J = 1$ and 2), respectively, where $Q(\tau, \boldsymbol{\eta})$ is defined by (2.3) and the orders of the models are given by $p_j = 1, 2$ and $q_j = 1, 2$ for $j = 1, 2$. Hence there are eight models in total. The initial values of $\boldsymbol{\beta}$, γ , $\boldsymbol{\eta}$ and d required by the estimation and forecasting method were simulated uniformly on $(0, 1)$, $(Q_{x,0.25}, Q_{x,0.75})$, $(-1, 0)$ and $\{1, 2, 3\}$, respectively, where $Q_{x,0.25}$ (or $Q_{x,0.75}$) is the lower (or upper) 25% percentile of $\{x_t\}$. The best two

models selected by Bayes factors are:

$$\begin{aligned}
 Q_{x_t}(\tau \mid \mathbf{x}_{t-1}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, \hat{d}) &= Q(\tau, \hat{\boldsymbol{\eta}}) \sqrt{\hat{h}_t}, \\
 h_t &= \begin{cases} 0.056 + 0.020x_{t-1}^2 + 0.023x_{t-2}^2 + 0.965h_{t-1}, & x_{t-1} < -0.246, \\ 0.016 + 0.003x_{t-1}^2 + 0.016x_{t-2}^2 + 0.764h_{t-1}, & x_{t-1} \geq -0.246, \end{cases} \quad (5.2) \\
 Q(\tau, \hat{\boldsymbol{\eta}}) &= \frac{\tau^{-0.066} - 1}{-0.066} - \frac{(1 - \tau)^{-0.009} - 1}{-0.009},
 \end{aligned}$$

and

$$\begin{aligned}
 Q_{x_t}(\tau \mid \mathbf{x}_{t-1}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, \hat{d}) &= Q(\tau, \hat{\boldsymbol{\eta}}) \sqrt{\hat{h}_t}, \\
 h_t &= \begin{cases} 0.009 + 0.028x_{t-1}^2 + 0.027x_{t-2}^2 + 0.906h_{t-1}, & x_{t-1} < 0.608, \\ 0.064 + 0.056x_{t-1}^2 + 0.045x_{t-2}^2 + 0.364h_{t-1}, & 0.608 \leq x_{t-1} < 0.678, \\ 0.008 + 0.004x_{t-1}^2 + 0.038x_{t-2}^2 + 0.645h_{t-1}, & x_{t-1} \geq 0.678, \end{cases} \\
 Q(\tau, \hat{\boldsymbol{\eta}}) &= \frac{\tau^{-0.046} - 1}{-0.046} - \frac{(1 - \tau)^{-0.004} - 1}{-0.004}, \quad (5.3)
 \end{aligned}$$

with $p(\mathbf{x}_T \mid (J, \mathbf{p}, \mathbf{q})) = 1/20.75$ and $1/19.51$ for models (5.2) and (5.3) respectively, which are very similar. So it is worth comparing them further in order to identify a model that is not only statistically sound but also financially meaningful.

For illustration purposes, we first show the time series plots for the posterior samples for model (5.2) in Figures 5 and 6, which suggest that the Markov chains generated from our method converged.

For each model, we also compare the distribution of the standardized residuals $\hat{r}_t = x_t / \sqrt{\hat{h}_t}$ with $Q(\tau, \hat{\boldsymbol{\eta}})$ and we hope that they are not significantly different. The results of the K-S test confirmed this as the test did not reject the null hypothesis at the 1% level of significance for both models (both p-values are greater than 0.02). Hence,

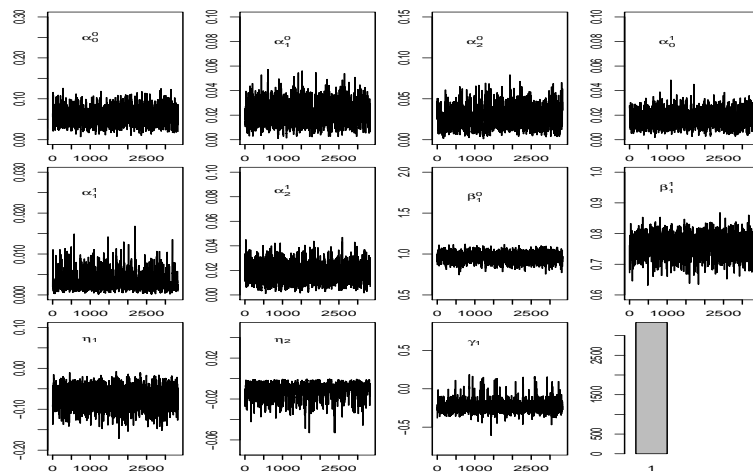


Figure 5: Time series plot for the posterior samples of the parameters in model (5.2).

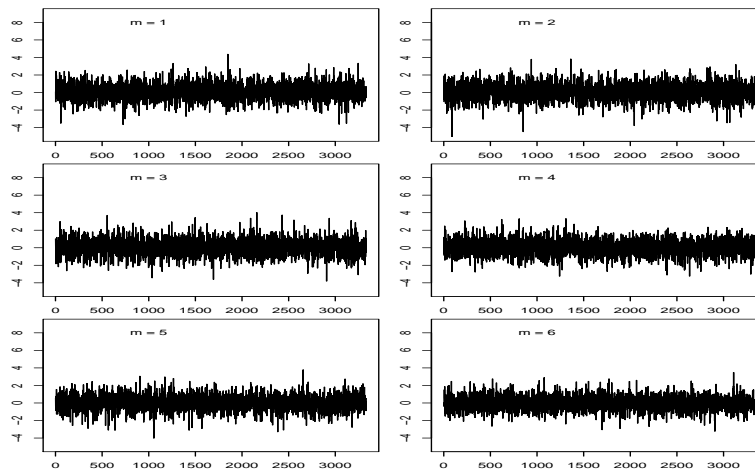


Figure 6: Time series plot for the posterior samples of the forecasts for model (5.2).

both models behave well from this point of view.

As we know the entire distribution of x_t , where $t = 1, \dots, T$, we are able to calculate the conditional quantile curves at levels $\tau = 0.01, 0.05, 0.25, 0.5, 0.75, 0.95$ and 0.99 . For a good fitted model we would expect that the proportion of the returns that are below the τ th quantile curve is close to the value of τ . Table 1 provides a summary of the results. It is seen that model (5.2) is slightly better.

Table 1: Empirical coverage rates of the estimated conditional quantile curves

τ	0.01	0.05	0.25	0.5	0.75	0.95	0.99	MSE
Model (5.2)	0.008	0.056	0.235	0.450	0.727	0.950	0.996	4.75×10^{-4}
Model (5.3)	0.013	0.062	0.246	0.451	0.718	0.946	0.998	5.20×10^{-4}

It follows from (5.1) that the conditional quantile function of the return y_t is given by

$$Q_{y_t}(\tau \mid \mathbf{y}_{t-1}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, \hat{d}) = 0.013 - 0.0995y_{t-1} + Q_{x_t}(\tau \mid \mathbf{x}_{t-1}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, \hat{d}), \quad (5.4)$$

where $Q_{x_t}(\tau \mid \mathbf{x}_{t-1}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}})$ is given by models (5.2) and (5.3) respectively. As a result, based on the posterior samples $\{x_{T+m}^{(1)}, \dots, x_{T+m}^{(U)}\}$, we can construct a random sample of y_{T+m} by letting $y_{T+m}^{(u)} = 0.013 - 0.0995y_{T+m-1}^{(u)} + x_{T+m}^{(u)}$, where $m = 1, \dots, 6$, U is the number of posterior samples and $y_T^{(u)} = y_T = -2.8652$ for all u . Hence, density forecasts of y_{T+m} can be obtained by using models (5.2) and (5.3) respectively. K-S test results show that the two density forecasts are significantly different at the 1% level for all values of m considered in this study. One of the limitations of the K-S test is that this test is not constructive in that if a rejection occurs, the test itself provides no guidance as to why. Hence, we further check the differences between the density forecasts by comparing the forecasts of mean, median, 1% quantile (LQ), 99%

quantile (UQ) and expected shortfalls (ES) at a 1% level, all of which can be derived from the density forecasts. The results are given in Table 2.

To evaluate the accuracy of the forecasts, we compare the differences between the observed data and their forecasts using the square root MSE (RMSE) and the mean absolute deviation (MAD). We also use the tracking signal (TS) method proposed by Brown (1959, 1962), which is defined as $TS_t = \sum_{t=1}^m (f_t - a_t)/C$, for $m = 1, \dots, M$, where f_t is the forecast, a_t is the actual observed value, and $C = \sum_{t=1}^m |f_t - a_t|/m$ (i.e. the MAD). An alarm will be raised if the TS value is outside $(-3.75, 3.75)$. This method has been widely used in monitoring forecasting systems to ensure that the underlying systems remain in control (Li et al., 2012). These RMSE, MAD and TS values are also given in Table 2.

It is seen that, for both models, the mean and median forecasts are very similar, although the median forecasts are slightly better than the mean forecasts for this data set. The TS values also suggest that model (5.2) is slightly better since the TS values for this model are closer to the center of the interval $(-3.75, 3.75)$, suggesting that this forecasting system is more stable. However, the two models produced very different lower and upper 1% quantile forecasts and expected shortfall forecasts, which suggest that the main difference between the two density forecasts could be in the region of tails. This explains why the K-S test rejected the null hypothesis. In summary, the results of statistical evaluations suggest that both models perform similarly from a model fitting point of view, but from a density forecasting point of view, the two models behaves very differently, especially in the tails of the density forecasts. So, which model is better for the returns?

To answer this question we now evaluate the models from a financial point of view.

Table 2: Multiple-step ahead forecasts of mean, median, 1% lower and upper quantiles (LQ and UQ) and the expected shortfalls (ES) at a 1% level. Tracking signals for the mean (TS-mean) and for the median (TS-median) forecasts.

QF-TGARCH model (5.2)								
m	Observed	Mean	Median	LQ	UQ	ES	TS-mean	TS-median
1	0.654	0.101	0.104	-2.059	2.143	-2.460	-1.000	-1.000
2	2.369	0.130	0.143	-2.092	2.289	-2.547	-2.000	-2.000
3	-0.483	0.094	0.102	-2.063	2.288	-2.477	-1.973	-1.956
4	0.963	0.063	0.073	-2.003	2.141	-2.473	-2.919	-2.899
5	-1.718	0.044	0.046	-2.043	2.029	-2.438	-1.122	-1.095
6	1.180	0.023	0.015	-1.943	2.034	-2.325	-2.096	-2.074
	RMSE	1.348	1.345					
	MAD	1.198	1.197					

QF-TGARCH model (5.3)								
m	Observed	Mean	Median	LQ	UQ	ES	TS-mean	TS-median
1	0.654	-0.033	-0.006	-3.429	3.197	-4.145	-1.000	-1.000
2	2.369	0.0548	0.077	-3.606	4.045	-4.466	-2.000	-2.000
3	-0.483	0.002	0.040	-3.768	3.576	-4.534	-2.165	-2.097
4	0.963	0.003	0.022	-3.562	3.762	-4.462	-3.127	-3.053
5	-1.718	0.032	0.047	-3.645	3.585	-4.923	-1.393	-1.298
6	1.180	0.006	0.007	-3.782	3.584	-4.606	-2.361	-2.266
	RMSE	1.380	1.375					
	MAD	1.228	1.226					

If the last term in model (5.4) is defined by (5.2), then it says that the distribution of tomorrow's return depends on today's return and tomorrow's volatility that depends on whether today's return is less than or greater than -0.246% . On the other hand, if it is defined by (5.3), then it says that the distribution of tomorrow's return depends on today's return and tomorrow's volatility that depends on whether today's return is less than 0.608% , or lies between 0.608% and 0.678% or greater than 0.678% . Obviously the two models provide very different financial explanations on the dynamics of Nasdaq returns from a distributional point of view.

It is worth noting that a perfect statistical model may not be able to verify the features of financial returns suggested by a financial theory or intuition from a financial theory in practice (Brooks, 2012, p.10). In finance it is well-known that the impact of negative and positive returns on conditional volatility dynamics is different, see, e.g. Glosten et al. (1993), Zakoian (1994), Nelson (1991) and Schwert (1990). This means that the threshold of these models is set at number zero, representing a zero return. Yu et al.'s (2010) work shows 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. It is seen that model (5.2) is in a good agreement with the work published in literature, but model (5.3) is not because it suggests two very close positive thresholds for the returns, which is difficult to explain from a financial point of view. Therefore, we prefer model (5.2) for this return series and will use it in the rest of the paper.

5.3 Estimated benchmark models

We estimated each of the following benchmark models to $\{x_t\}$:

$$\begin{aligned}
 x_t &= \sigma_t v_t, & v_t &\sim \text{skewed } t\text{-distribution} \\
 \text{GJR-GARCH: } \sigma_t^2 &= \omega + \sum_{i=1}^2 (a_i + b_i I_{t-i}) x_{t-i}^2 + c \sigma_{t-1}^2, \\
 \text{T-GARCH: } \sigma_t &= \omega + \sum_{i=1}^2 a_i (|x_{t-i}| - b_i x_{t-i}) + c \sigma_{t-1}, \\
 \text{E-GARCH: } \ln(\sigma_t^2) &= \omega + \sum_{i=1}^2 (a_i x_{t-i} + b_i (|x_{t-i}| - E|x_{t-i}|)) \\
 &\quad + c \ln(\sigma_{t-1}^2), \\
 \text{AVGARCH: } \sigma_t &= \omega + \sum_{i=1}^2 a_i |x_{t-i}| + c \sigma_{t-1},
 \end{aligned} \tag{5.5}$$

where $I_{t-i} = 0$ if $x_{t-i} \geq 0$, otherwise $I_{t-i} = 1$. Note that the order of these models is the same as that of model (5.2) for comparison purposes. The estimated parameters are given in Table 3.

For each model, we calculated the standardized residuals and compared its distribution with the corresponding skewed t -distribution using the K-S test. The test results show that the two distributions are significantly different at any conventional levels (all p-values are much less than 0.01). Figure 7 shows the quantile-quantile plots for these models, which explains why the K-S test rejected the null hypothesis, suggesting that these models fail to interpret the data well. It is worth mentioning that model (5.2) is much better than the benchmark models because the K-S test did not reject the null hypothesis for model (5.2) and the quantile-quantile plot for model (5.2) is also satisfactory (see the first plot in Figure 7).

We now compare the benchmark models (5.5) with the QF-TGARCH model (5.2) from a density forecasting point of view. Suppose that u_t is generated from $f_t(u_t | \mathbf{u}_{t-1})$ and $p_t(u_t)$ is the forecast of $f_t(u_t | \mathbf{u}_{t-1})$. Diebold et al. (1998) showed that, if

Table 3: Estimated parameters of the GARCH-type models

Estimated parameters of the GJR-GARCH model								
Parameter	ω	a_1	a_2	c	b_1	b_2	skew	df
Estimate	0.049	0.000	0.015	0.871	0.055	0.131	0.822	11.507
Std.Error	0.013	0.193	0.169	0.02	0.176	0.145	0.027	1.892
Estimated parameters of the T-GARCH model								
Parameter	ω	a_1	a_2	c	b_1	b_2	skew	df
Estimate	0.046	0.066	0.049	0.879	1.000	-0.338	0.810	12.025
Std.Error	0.006	0.012	0.011	0.011	0.170	0.460	0.030	3.964
Estimated parameters of the E-GARCH model								
Parameter	ω	a_1	a_2	c	b_1	b_2	skew	df
Estimate	0.020	-0.291	0.148	0.970	-0.224	0.366	0.803	14.042
Std.Error	0.005	0.042	0.042	0.006	0.060	0.061	0.031	5.433
Estimated parameters of the AVGARCH model								
Parameter	ω	a_1	a_2	c	b_1	b_2	skew	df
Estimate	0.029	0.000	0.133	0.879	-	-	0.834	9.666
Std.Error	0.010	0.031	0.037	0.018	-	-	0.029	2.609

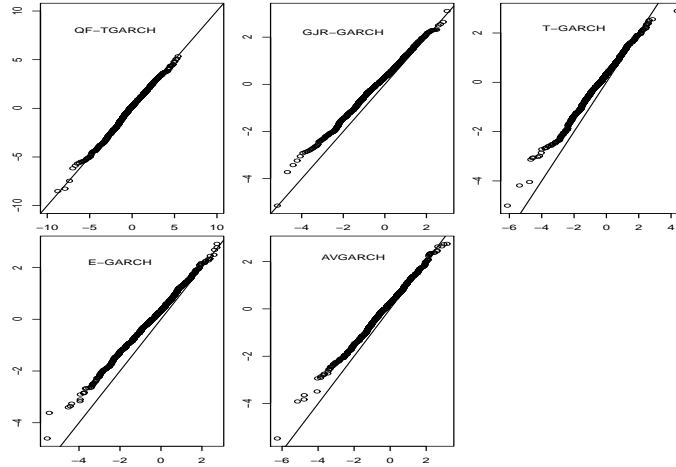


Figure 7: QQ-plots for the estimated QF-TGARCH and benchmark models.

$p_t(u_t)$ coincides with $f_t(u_t | \mathbf{u}_{t-1})$, then the sequence of probability integral transforms of u_t , denoted by $\{z_t\}$ are *i.i.d.* with $U(0, 1)$, where $z_t = \int_{-\infty}^{u_t} p_t(u) du$. We will use the approach of Diebold et al. (1998) to compare the estimated models because it does not depend on the method that is used for obtaining the forecasts.

Figure 8 shows that the density forecasts obtained from the QF-TGARCH model are significantly better because the distribution of $\{z_t\}$ for the benchmark models deviates from that of $U(0, 1)$ much more seriously. Moreover, we found that the strength of the autocorrelation structures for $\{z_t - \bar{z}\}$ and $\{(z_t - \bar{z})^2\}$ is much weaker for the QF-TGARCH model than that for the benchmark models, because for our model the sample ACFs stand out at lag one only, while those for the benchmark models show much stronger autocorrelation structures. For illustration purposes, Figure 9 shows the ACF plots for our model and the GJR-GARCH model.

Some possible reasons for the better performance of the QF-TGARCH model are

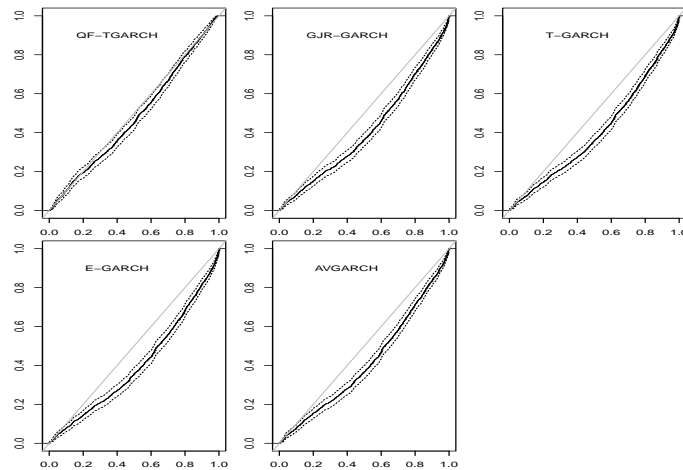


Figure 8: Distribution function of $U(0,1)$ (grey) and that of $\{z_t\}$ (dark) together with a 95% confidence interval (dotted curve) for the estimated models.

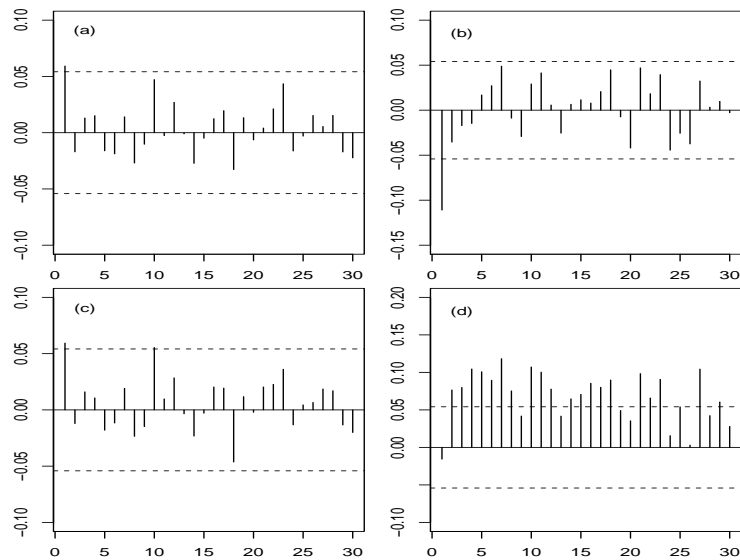


Figure 9: Sample ACFs of $\{(z_t - \bar{z})\}$ and $\{(z_t - \bar{z})^2\}$ for the QF-TGARCH model (first row) and the GJR-GARCH model (second row) respectively.

discussed below: (i) We used the GLD for our QF-TGARCH model, which allowed us to capture some data structures that cannot be captured by the skewed t-distribution. (ii) The forecasts from the QF-TGARCH model can take into account the variation of the model parameters, while those from the benchmark models cannot. (iii) The threshold in our QF-TGARCH model needs to be estimated from data, but that in the benchmark models is fixed.

In summary, this application shows that the proposed QF-TGARCH model can be used to study the entire conditional distribution of financial returns and the asymmetric relation between stock returns and volatility changes. It also outperforms the commonly used TGARCH type models with respect to model estimation and forecasting. Hence, the proposed model has the potential to be very useful in financial data analysis.

6 Conclusions and discussions

We develop a QF-TGARCH model for analyzing the distribution of financial returns that follow a TGARCH model. We suggest a Bayesian method to do parameter estimation and forecasting simultaneously. This method allows us to handle multiple thresholds, the delay parameter and forecasts easily and ensures that the forecasts can take into account the variation of the model parameters. We have shown that the proposed model outperforms some commonly used TGARCH models with respect to model estimation and forecasting. We have also shown that the QF-TGARCH model with $Q(\tau, \boldsymbol{\eta})$ defined by (2.3) is more robust to model specification errors.

It is worth noting that this paper illustrates a novel approach to studying financial

data. The proposed model can be extended in several different ways. For example, rather than using (2.3) for $Q(\tau, \boldsymbol{\eta})$ in the model, we could use many other distributions that can be defined by a quantile function only. An extension of the proposed model in this direction will lead to many new models.

In this paper we focus on the TGARCH model of Yu et al. (2010) and hence, our model does not contain a conditional location process. It is interesting to extend our approach by including a location process into the proposed model so that the conditional location, scale and forecasts can be estimated simultaneously. It is worth noting that if an autocorrelation term is included into the model as its location term, then several issues need to be considered. For example, what restrictions on the parameters of the location process should be, whether the location process depends on thresholds, whether the thresholds for the location process are the same as those for the volatility process etc. Note that an extension in this direction will require us to develop a new sampling scheme and to derive new formulae for the acceptance probabilities. We will leave this for future work.

Another way to extend our approach is to treat the number of regimes J and the values of p_j and q_j for $j = 1, \dots, J$ as parameters of the model and to be estimated simultaneously with other parameters and forecasts. However, any change in their values will result in a change in the dimension of the parameter space of the posterior distribution function. Therefore, a new estimation method, e.g. a reversible MCMC algorithm, needs to be developed in order to deal with this case.

Supplementary materials

Supplementary materials for this paper are available from

<http://www.statmod.org/smij/archive.html>.

Acknowledgements

We would like to thank three referees for very helpful and insightful comments and constructive suggestions on our research, that have led to substantial improvements of this paper.

References

- Ausn, M.C. and Galeano, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, 51, 2636-2652
- Bauwens, L. and Lubrano, M. (2002). Bayesian option pricing using asymmetric GARCH models. *Journal of Empirical Finance*, 9, 321-342.
- Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.
- Bondell, H.D., Reich, B.J. and Wang, H. (2010). Non-crossing quantile regression curve estimation. *Biometrika*, 97, 825-838.

- Brooks, C. (2012). *Introductory Econometrics for Finance*. Cambridge: Cambridge University Press.
- Brown, R.G. (1959). *Statistical forecasting for inventory control*. New York: McGraw-Hill.
- Brown, R.G. (1962). *Smoothing, forecasting and prediction of discrete time series*. Englewood Cliffs, N.J.: Prentice-Hall.
- Cai, Y. (2016). A General Quantile Function Model for Economic and Financial Time Series. *Econometric Reviews*, 35, 1173-1193.
- Cai, Y., Montes-Rojas, G. and Olmo J.(2013). A quantile double AR model: estimation and forecasting. *Journal of Forecasting*, 32, 551-560.
- Cai, Y. (2010). Forecasting for quantile self exciting threshold autoregressive time series models. *Biometrika*, 97, 199-208.
- Cai, Y. and Stander, J. (2008). Quantile self-exciting threshold autoregressive time series models. *Journal of Time Series Analysis*, 29, 186-202.
- Chen, C.W.S. and So, M.K.P. (2006). On a threshold heteroscedastic model. *International Journal of Forecasting*, 22, 73-89.
- Chen, C.W.S., So, M.K.P. and Liu, F.C. (2011). A review of threshold time series models in finance. *Statistics and its Interface*, 4, 167-181.
- Dellaportas, P. and 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.

- Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998). Evaluating density forecasts. *International Economic Review*, 39, 863-883.
- Engle, R.F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22, 367-381.
- Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1007.
- Fournier, B., Rupin, N., Bigerelle, M., Najjar, D., Iost, A. and Wilcox, R. (2007). Estimating the parameters of a generalized lambda distribution. *Computational Statistics & Data Analysis*, 51, 2813-2835.
- Freimer, M., Mudholkar, G., Kollia, G. and Lin, C. (1988). A study of the generalised Tukey lambda family. *Communications in Statistics - Theory Methods*, 17, 3547-3567.
- Franses, P.H. and Van Dijk, D. (1996). Forecasting stock market volatility using (non-linear) Garch models. *Journal of Forecasting*, 15, 229-235.
- Gaglianone, W.P. and Lime L.R. (2012). Constructing density forecasts from quantile regressions. *Journal of Money, Credit and Banking*, 44, 1589-1607.
- Gelfand, A.E. and Dey, D. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society Series B*, 56, 501-514.
- Gilchrist, W.G. (2000). *Statistical Modelling with Quantile Functions*. London: Chapman & Hall/CRC.

- Glosten, L.R., Jagannathan, R. and 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.
- Granger, C.W.J. (1992). Forecasting stock market prices: Lessons for forecasters. *International Journal of Forecasting*, 8, 3-13.
- Jensen, M.J. and Maheu, J.M. (2013). Bayesian Semiparametric Multivariate GARCH Modeling. *Journal of Econometrics*, 176, 3-17.
- Karian, Z.A., Dudewicz, E.J. and 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. and 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. and MacGillivray, H. (1999). A starship estimation method for the generalised lambda distributions. *Australia New Zealand Journal of Statistics*, 41, 353-374.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33-50.
- Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.
- Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, 12, 793-813.

- Koop, G. and Potter, S.M. (2003). Bayesian analysis of endogenous delay threshold models. *Journal of Business & Economic Statistics*, 21, 93-103.
- Li, Y.K., Lange, M. and Stocks, C. (2012). Monitoring Forecasting Systems - Revisit Trigg's Tracking Signal. The 32nd Annual International Symposium on Forecasting 24 - 27 June 2012, Boston, USA.
- Nelson D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.
- O'Hagan, A. and Forster, J.J. (2004). *Bayesian Inference*. Arnold, London.
- Ozturk, A. and Dale, R.F. (1985). Least squares estimation of the parameters of the generalized lambda distribution. *Technometrics*, 27, 81-84.
- Schwert, G.W. (1990). Stock volatility and the crash of '87. *Review of Financial Studies*, 3, 77-102.
- So M.K.P and Chung, R.S.W (2015). Statistical inference for conditional quantiles in nonlinear time series. *Journal of Econometrics*, 189, 457-472.
- 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*, Vol 21, Issue 9. <http://www.jstatsoft.org/>.
- Taylor, J.W. (2008). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics*, 6, 382-406.

- Virbickaite, A., Ausin, C.M. and Galeano, P. (2015). Bayesian inference methods for univariate and multivariate GARCH models: a Survey. *Journal of Economic Surveys*, 29, 76-96.
- Vrontos, I.D., Dellaportas, P. and Politis, D.N. (2012). Full Bayesian inference for GARCH and EGARCH models. *Journal Journal of Business & Economic Statistics*, 18, 187-198.
- Xiao, Z. and Koenker, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association*, 104, 1696-1712.
- Yang, Y.L. and 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. and Jin, S. (2010). On some models for Value at Risk. *Economic Reviews*, 29, 622-641.
- Zakoian, J.M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18, 931-955.