Swansea University
Prifysgol Abertawe

Cronfa
Setting Research Free

Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
Journal of Financial Econometrics

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa49769

Paper:
Cai, Y. & Stander, J. (2019). The Threshold GARCH Model: Estimation and Density Forecasting for Financial Returns*. *Journal of Financial Econometrics*
http://dx.doi.org/10.1093/jjfinec/nbz014

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

http://www.swansea.ac.uk/library/researchsupport/ris-support/

Title:

The threshold GARCH model: estimation and density forecasting for financial returns

Authors:

Yuzhi Cai, Swansea University, UK

Julian Stander, University of Plymouth, UK

Address for correspondence:

Dr Yuzhi Cai

Department of Accounting and Finance

School of Management

Swansea University

United Kingdom

Email: y.cai@swansea.ac.uk

Tel: +44 1792 606865

Abstract

We consider multiple threshold value-at-risk (VaR$_t$) estimation and density forecasting for financial data following a threshold GARCH model. We develop an α-quantile quasi-maximum likelihood estimation (QMLE) method for VaR$_t$ by showing that the associated density function is an α-quantile density and belongs to the tick-exponential family. This establishes that our estimator is consistent for the parameters of VaR$_t$. We propose a density forecasting method for quantile models based on VaR$_t$ at a single non-extreme level, which overcomes some limitations of existing forecasting methods with quantile models. We find that for heavy-tailed financial data our α-quantile QMLE method for VaR$_t$ outperms the Gaussian QMLE method for volatility. We also find that density forecasts based on VaR$_t$ outperform those based on the volatility of financial data. Empirical work on market returns shows that our approach also outperforms some benchmark models for density forecasting of financial returns.

Key words: α-quantile density, density forecasting, QMLE, threshold, value-at-risk (VaR).

JEL classification numbers: C1, C5

1 Introduction

This paper is motivated by the need to produce forecasts in the form of probability density functions for financial time series. We focus on threshold GARCH (TGARCH) models due to their popularity and usefulness in financial analysis. TGARCH models have been studied by many researchers, including Glosten et al. (1993) who developed a TGARCH model, the so called GJR-GARCH model, to study the impact of negative and positive returns on conditional volatility dynamics. Zakoian (1994) also proposed a TGARCH model for similar purposes. Park et al. (2009) studied persistent TGARCH processes, 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 changes in the value-at-risk (VaR) process.

We work with a general TGARCH model with k-regimes (see, e.g., Yu et al., 2010) defined by

$$\begin{aligned} x_t &= \varepsilon_t \sqrt{h_t}, \\ h_t &= \sum_{j=1}^{k} \left(\alpha_{j0} + \sum_{p=1}^{p_j} \alpha_{jp} x_{t-p}^2 + \sum_{q=1}^{q_j} \beta_{jq} h_{t-q}\right) I\left[x_{t-d} \in \Omega_j\right], \end{aligned} \tag{1}$$

where ε_ts are independently and identically distributed (iid) random variables with mean 0 and variance 1, the so-called delay parameter d is a positive integer, $\Omega_j = [\gamma_{j-1}\ \gamma_j)$ in which the γ_js are real-valued thresholds such that $-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_{k-1} < \gamma_k = \infty$, and $p_j \geq 0$ and $q_j \geq 0$ define the order of the volatility dynamics in regime j. Moreover, $\alpha_{j0} > 0$, $\alpha_{jp} \geq 0$, $\beta_{jq} \geq 0$, and $I[\cdot]$ is the standard indicator function. Liu et al. (1997) proved that, under some regularity conditions, there exist stationary and ergodic solutions satisfying model (1).

The TGARCH model (1) is different from that studied by Zakoian (1994), which is defined by $x_t = \sigma_t \varepsilon_t$ and $\sigma_t = \alpha_0 + \sum_{i=1}^{p} \left(\alpha_i^+ x_{t-i}^+ - \alpha_i^- x_{t-i}^-\right) + \sum_{j=1}^{q} \beta_j \sigma_{t-j}$, where α_0, α_i^+, α_i^- and β_j for $i = 1, \ldots, p$ and $j = 1, \ldots, q$ are model parameters, and $x_t^+ = \max(x_t, 0)$ and $x_t^- = \min(x_t, 0)$. Here, the dynamics are defined through σ_t (the volatility) rather than h_t (the variance), and the positive and negative parts of x_{t-i} are used rather than squared values.

Model (1) is also different from the GJR-GARCH model which is defined as $x_t = \sigma_t \varepsilon_t$ and $\sigma_t^2 = \omega + \sum_{i=1}^{p} \left(\alpha_i + \xi_i I[x_{t-i} < 0]\right) x_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$, where ω, α_i, ξ_i and β_j for $i = 1, \ldots, p$ and $j = 1, \ldots, q$ are the model parameters. Here, the dynamics are defined for the variance as $h_t = \sigma_t^2$, but σ_t^2 follows different expressions according to the sign of x_{t-i} for $i = 1, \ldots, p$. These two models do not involve a delay parameter d or any thresholds.

Let VaR_t be the one step ahead τth quanitle of $x_t | \mathbf{x}_{t-1}$, where $0 < \tau < 1$, and $\mathbf{x_{t-1}} = (\mathbf{x_1}, \ldots, \mathbf{x_{t-1}})$. Yu et al. (2010) further showed that if x_t follows model (1), then VaR_t

satisfies

$$\text{VaR}_t \;= s_Q \sqrt{\textstyle\sum_{j=1}^{k} \left(a_{j0} + \sum_{p=1}^{p_j} a_{jp} x_{t-p}^2 + \sum_{q=1}^{q_j} b_{jq} \text{VaR}_{t-q}^2\right) I\left[x_{t-d} \in \Omega_j\right]}\,, \qquad (2)$$

where $a_{jp} = Q^2(\tau)\alpha_{jp}$, in which $Q(\tau)$ is the τth quantile of ε_t, $b_{jq} = \beta_{jq}$ and $s_Q = \text{sign}(Q(\tau))$, that is, $s_Q = 1$ (-1) if $Q(\tau) > (<)$ 0. We further define $s_Q = 0$ for $Q(\tau) = 0$, but we will not consider this case as this leads to a model that is not well defined. Since Yu et al. (2010) did not provide an explicit proof for this result, we provide one in Appendix I. Note that VaR_t defined by (2) depends on τ, but we have dropped τ to simplify the notation.

For model (1), the Gaussian quasi-maximum likelihood estimation (QMLE) method guarantees that the estimator is consistent for the parameters of h_t. However, Komunjer (2005a, b) showed that it is not appropriate to use the Gaussian QMLE for h_t to obtain quantile estimate for VaR_t in (2) because the corresponding conditional quantile estimator obtained in this way is generally biased.

Hence, in this paper we do not use the Gaussian QMLE method for h_t, but instead, we propose an α-quantile QMLE method for the estimation of VaR_t of financial data that follow (1). More specifically, we showed that model (1) can be expressed by $x_t = \text{VaR}_t + \sqrt{g_t}\, v_t$, where v_t and g_t are given in Proposition 1 below. Then for this equivelent model, we developed a density function for v_t and showed in Proposition 2 below that this density function is an α-quantile density and belongs to the tick-exponential family, defined by Komunjer (2005a, b). This ensures that the α-quantile QMLE method delivers consistent parameter estimates for VaR_t using results in Komunjer (2005a, b).

It is worth noting that our α-quantile density for v_t is different from the skewed-Laplace density used by the standard quantile regression method, see Koenker & Bassett (1978) and Koenker (2005), although this skewed-Laplace density also belongs to the tick-exponential family (Komunjer, 2005b). The difference between the two densities is that our α-quantile density for v_t has a unit variance, which allows us to take account of the effect of heterogeneity when estimating VaR_t.

Although both Gaussian QMLE for h_t and α-quantile QMLE for VaR_t deliver consistent estimates, we prefer the latter one. This is because existing results suggest that the Gaussian QMLE for h_t may not perform well for financial data with heavy-tails, see e.g. Hall and Yao (2003), which could lead to incorrect forecasts. In fact, we have conducted extensive simulation studies for both estimation and forecasting. The results show that for heavy-tailed financial data the α-quantile QMLE for VaR_t outperforms the Gaussian QMLE for h_t in both estimation and forecasting.

Parameters can be estimated by using frequentist methods, which usually involve a grid search for thresholds and the delay parameter. See, e.g. Yu et al. (2010). In this paper, we present a Markov chain Monte Carlo (MCMC) method to illustrate an alternative way for parameter estimation in this case.

Density forecasting has become an important area of research in the analysis of financial data, see e.g. Timmermann (2000). For model (1), density forecasts can be obtained easily if we know h_t and the distribution of ε_t. However, without knowing the distribution of ε_t, little discussion about density forecasting using only h_t can be found in the literature.

On the other hand, there is some work in the literature that allows us to obtain density forecasts for financial returns without using the distribution of ε_t but by using quantiles, which correspond to VaR_t. For example, Taylor (2005) proposed a method for constructing financial volatility forecasts from VaR estimates. Cai (2010) presented a forecasting method for quantile SETAR time series models, while Cai et al. (2012) combined density forecasts from different quantile AR models. All these forecasting methods require many quantile models to be estimated, corresponding to a sequence of quantile levels that cover the entire distributional range of the underlying process. Consequently, models at extreme quantile levels also need to be estimated. However, due to the lack of information at extreme levels, these models can perform poorly compared to those at non-extreme levels, leading to unsatisfactory forecasting results. Moreover, if no monotonic restrictions are applied, the estimated quantiles at different levels may cross, resulting in invalid estimates. Hence, the above density forecasting methods can lead to incorrect density forecasts due to

these limitations.

To overcome the limitations of the existing methods discussed above, we also propose a simple density forecasting method. The method can use the Gaussian QMLE for h_t or use the α-quantile QMLE for VaR_t to produce density forecasts for financial returns that follow a TGARCH model. We will focus on density forecasts based on VaR_t in this paper in order to provide methodology for density forecasting with quantile models. We show that our method is able to deliver density forecasts by using VaR_t at a single non-extreme level for financial returns that follow a TGARCH model. This explains how our density forecasting method overcomes the limitations of existing methods discussed above. We conduct extensive simulation studies to compare the performance of density forecasts obtained using h_t with those obtained using VaR_t. The results show that density forecasts obtained using VaR_t outperform those obtained using h_t for TGARCH models.

The rest of the paper is organized as follows. Our estimation and forecasting methods are presented in Sections 2 and 3 respectively. In Section 4 we discuss the results of simulation studies. In Section 5 we illustrate our method by conducting empirical work on Hang Seng and S&P500 daily closing returns over a five year period. We compare our results with those obtained from ARMA-GARCH and GJR-GARCH models. Some further discussion and suggestions for future research are given in Section 6.

2 Parameter estimation

2.1 Quasi-likelihood estimation of the parameters of VaR_t process

In this section we consider the estimation of the parameters a_{jp}, b_{jq}, γ_j, d and s_Q that appear in (2), conditional on the values of k, p_j and q_j for all possible values of j. We do not consider the estimation of k, p_j and q_j in detail, but provide a brief discussion about this in Section 6.

The conventional maximum likelihood approach to parameter estimation specifies a density function for $x_t|\mathbf{x}_{t-1}$, assuming it to be the true density function of the underlying process. Unfortunately, the consequences of model mis-specification can be serious. The QMLE method is a common approach aimed at overcoming problems associated with model mis-specification. As explained by White (1982, 1994), the QMLE method may be defined by minimizing the Kullback-Leibler Information Criterion (KLIC) of the true density function of $x_t|\mathbf{x}_{t-1}$ relative to the specified density function. The KLIC reaches its minimum value 0 if and only if the specified density function is the correct one. In practice, the Gaussian distribution is commonly used as the specified density function for the QMLE. As well as White (1982, 1994), we refer to Wedderburn (1974), McCullagh and Nelder (1989), McCullagh (1991), Firth (1993), Heyde (1997), Davidson (2001), Pawitan (2001), Davison (2003) and Kuan (2004) for detailed discussions about the QMLE method.

The classical Gaussian QMLE approach focuses on the mean and the scale of $x_t|\mathbf{x}_{t-1}$ with respect to the mean; see Section 3.5 of Faraway (2016), for example. Unfortunately, it is not appropriate to use the Gaussian QMLE for h_t to obtain quantile estimate for VaR_t because, as Komunjer (2005a, b) pointed out, the corresponding conditional quantile estimator obtained in this way is generally biased. Hence, we estimate process (2) by using the quantile QMLE framework of Komunjer (2005a, b), which requires us to work with a density that is an α-quantile density or belongs to the tick-exponential family. Definition of these concepts appear in Appendix III.

As a first step in achieving this, we consider $u_t = x_t - \text{VaR}_t$, which we model as $\sqrt{g_t}\, v_t$, where v_t are iid random variables with τth quantile 0 and variance 1. This modelling assumption implies that

$$x_t = \text{VaR}_t + \sqrt{g_t}\, v_t. \tag{3}$$

That the τth quantile of $x_t|\mathbf{x}_{t-1}$ is VaR_t follows from the fact that the τth quantile of v_t is 0. It is worth reemphasizing that, in the following, we do not do Gaussian QMLE on

model (1), but instead, we do α-quantile QMLE on model (3), where g_t is given below in the Proposition 1.

Proposition 1 *If x_t satisfies both models (1) and (3), then $g_t = (h_t + VaR_t^2)/E[v_t^2]$. Furthermore,*

$$g_t = \sum_{j=1}^{k} \left(\phi_{j0} + \sum_{p=1}^{p_j} \phi_{jp} x_{t-p}^2 + \sum_{q=1}^{q_j} b_{jq} g_{t-q} \right) I\left[x_{t-d} \in \Omega_j\right], \tag{4}$$

where $\phi_{jp} = (\alpha_{jp} + a_{jp})/E[v_t^2]$ for $j = 1, \ldots, k$ and $p = 0, 1, \ldots, p_j$.

See Appendix II for a proof. It follows from the positivity assumptions made about the α_{jp}s that $\phi_{j0} > 0$ and $\phi_{jp} \geq 0$ for $j = 1, \ldots, k$ and $p = 1, \ldots, p_j$. It follows from Proposition 1 that g_t depends on τ as VaR_t is the τth quantile of $x_t|\mathbf{x}_{t-1}$, but by definition h_t does not depend on τ.

Next, we propose a suitable distribution for v_t. If a random variable W follows the skewed-Laplace distribution, it has probability density function $f(w) = \tau(1-\tau)\exp\{-w(\tau - I[w < 0])\}$. It can be shown that $\Pr(W \leq 0) = \tau$, so that the τth quantile of W is 0, and that $\text{Var}[W] = (1 - 2\tau + 2\tau^2) / \left\{(1-\tau)^2 \tau^2\right\}$. Let the random variable $V = (1-\tau)\tau W/\sqrt{1 - 2\tau + 2\tau^2}$ so that $\text{Var}[V] = 1$. The change of variable formula yields that the probability density function of V takes the form

$$f(v) = \sqrt{1 - 2\tau + 2\tau^2}\ \exp\left\{ v\sqrt{1 - 2\tau + 2\tau^2}/(\tau - I[v \geq 0]) \right\}, \tag{5}$$

while 0 remains the τth quantile of V; see Chen et al. (2009).

To justify the use of (5) in the estimation of VaR_t, we need to establish the relation between the density function defined by (5) and the α-quantile density and/or tick-exponential family defined by Komunjer (2005a, b). Doing this ensures that we can estimate the parameters of VaR_t in a consistent way.

Proposition 2 *The density function defined by (5) is an α-quantile density, as defined by Komunjer (2005a). It also belongs to the tick-exponential family of Komunjer (2005b).*

See Appendix III for a proof. Therefore, we can use the density function (5) for v_t in the α-quantile QMLE method for the estimation of VaR_t. It is worth noting that Komunjer (2005a, b) pointed out that the skewed-Laplace distribution also belongs to the tick-exponential family, hence the standard quantile regression approach of Koenker & Bassett (1978) is a special case of the quantile QMLE of Komunjer (2005a, b). The difference between our α-quantile density and the skewed-Laplace distribution used by Koenker & Bassett (1978) is that our density has a unit variance, which allows us to take into account the heterogeneity in financial data when estimating VaR_t.

Since $x_t - \text{VaR}_t = \sqrt{g_t} v_t$, it follows again from the change of variable formula that

$$f_{x_t|\mathbf{x}_{t-1}}(\text{VaR}_t, \sqrt{g_t} \mid \mathbf{x}_{t-1}, \boldsymbol{\theta}) = \frac{\sqrt{1-2\tau+2\tau^2}}{\sqrt{g_t}} \exp\left\{ \frac{(x_t - \text{VaR}_t)\sqrt{1-2\tau+2\tau^2}}{\sqrt{g_t}\,(\tau - I[x_t \geq \text{VaR}_t])} \right\}, \tag{6}$$

where $\boldsymbol{\theta}$ represents the parameter vector of the density function. Because $x_t - \text{VaR}_t = \sqrt{g_t}\, v_t$, then the scale of $x_t|\mathbf{x}_{t-1}$ with respect to VaR_t is $\sqrt{g_t}$. The different roles of g_t and h_t may be explained as follows: g_t allows us to take into account the heterogeneity in x_t when estimating VaR_t, while h_t in GARCH type models such as model (1) allows us to take into account the heterogeneity in x_t when estimating the conditional mean of x_t.

Now let $t_0 = \max_{1\leq j\leq k}(p_j, q_j, d)$, $\mathbf{V}_t = (\text{VaR}_1^2, \text{VaR}_2^2, \ldots, \text{VaR}_t^2)$ and $\mathbf{g}_t = (g_1, \ldots, g_t)$. Then conditional on the initial values $\mathbf{V}_{t_0}$, $\mathbf{x}_{t_0}$, $\mathbf{g}_{t_0}$ and the values of k, p_j and q_j for all possible j, the quasi-likelihood function of the parameters is given by

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) &= \prod_{t=t_0+1}^{T} f_{x_t|\mathbf{x}_{t-1}}(\text{VaR}_t, \sqrt{g_t} \mid \mathbf{x}_{t-1}, \boldsymbol{\theta}) \\ &= \tilde{M} \prod_{t=t_0+1}^{T} \frac{1}{\sqrt{g_t}} \exp\left\{ \frac{(x_t - \text{VaR}_t)\sqrt{1-2\tau+2\tau^2}}{\sqrt{g_t}\,(\tau - I[x_t \geq \text{VaR}_t])} \right\}, \end{aligned} \tag{7}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q)'$, $\boldsymbol{\beta}$ is a vector containing all the a's, b's and ϕ's parameters, $\boldsymbol{\gamma}$ contains thresholds, $\mathbf{p}$ and $\mathbf{q}$ contain all p's and q's respectively, and $\tilde{M} = (1 - 2\tau +$

$2\tau^2)^{(T-t_0)/2}$ is a constant.

Therefore, it follows from the theory of the quantile QMLE developed by Komunjer (2005a, b) that, if $\hat{\boldsymbol{\theta}}$ is the QMLE obtained from the above quasi-likelihood function, then $\hat{\boldsymbol{\theta}}$ is consistent for the parameters of the conditional quantile VaR_t.

The parameters can now be estimated by maximizing the quasi-likelihood function (7). In the next section, we discuss an alternative method for parameter estimations, i.e. a Bayesian approach to parameter estimation implemented using an MCMC algorithm.

2.2 Posterior distribution of the parameters $\boldsymbol{\theta}$

We propose a Bayesian approach to parameter estimation and inference implemented using an MCMC algorithm. Our Bayesian approach requires that the posterior distribution of the model parameters $\boldsymbol{\theta}$ has a finite integral over its parameter space. For this purpose, we slightly modify the parameter space by requiring that $\phi_{j0} \geq \phi_0 > 0$ for all j, where ϕ_0 is a fixed, very small positive number, and still denote the modified space by Ω. We set ϕ_0 to be 10^{-30} in order to allow the parameter space to be as wide as possible. Hence this requirement has little impact on parameter estimation.

Let $\pi_\theta(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0})$ be the prior density function of the parameters. By following the work of McCulloch and Tsay (1993, 1994) and Rosenberg and Young (1999), we assume that the prior distribution does not depend on $\mathbf{x}_{t_0}, \mathbf{V}_{t_0}$ and $\mathbf{g}_{t_0}$. We also assume that the prior distribution takes the form $\pi_\theta(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0})$ $= \pi_\beta(\boldsymbol{\beta})\pi_\gamma(\boldsymbol{\gamma})\pi_d(d)\pi_s(s_Q)$. Therefore, the posterior density function of $\boldsymbol{\theta}$ is given by

$$\begin{aligned} & \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) \\ \propto \; & L(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0})\pi_\beta(\boldsymbol{\beta})\pi_\gamma(\boldsymbol{\gamma})\pi_d(d)\pi_s(s_Q). \end{aligned} \tag{8}$$

Proposition 3 If the prior density function

$$\pi_\theta(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) = \pi_\beta(\boldsymbol{\beta})\pi_\gamma(\boldsymbol{\gamma})\pi_d(d)\pi_s(s_Q)$$

is well defined on Ω, i.e. $\int_\Omega \pi_\theta(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) d\boldsymbol{\theta} < \infty$, then the posterior distribution defined by (8) is also well defined on Ω, i.e. $\int_\Omega \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) d\boldsymbol{\theta} < \infty$, where $\boldsymbol{\theta}$ represents the vector of all model parameters.

See Appendix IV for a proof. It follows from Proposition 3 that for the posterior distribution function to be well defined, we only need to ensure that the prior density function is well defined on Ω. To achieve this, we used a log-normal distribution as the prior distribution for each of the parameters a_{jp}, b_{jq} and ϕ_{jp} as they should be non-negative, and a normal distribution for each of the thresholds γ_j since they may take any real numbers. We let the prior distribution of s_Q be uniform on its two parameters $\{-1, +1\}$, and we let the prior distribution of d be uniform on $\{1, \ldots, d_0\}$, where d_0 is the largest value of d that we would like to consider; following Yu et al. (2010), we set $d_0 = 3$. An explicit expression of the prior density function $\pi_{\boldsymbol{\theta}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0})$ is given in Appendix V. Because it is the product of proper densities, the prior density function is well defined on Ω as required.

2.3 MCMC method for sampling from the posterior distribution

The basic idea of an MCMC method is to generate a sequence of model parameters taking values in the parameter space Ω such that they form a Markov chain, the equilibrium distribution of which is the posterior distribution of the parameters; see Brooks (1998) for details. Often, this is achieved by using the Metropolis-Hastings algorithm in which a candidate parameter value is simulated from a chosen distribution and this proposed value is accepted as the next in the sequence with a known probability; see Gamerman and Lopes, 2002, Ch. 6, and Geyer, 2011, for example.

Let $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, d and s_Q be the current values of the Markov chain on Ω. We will use the notation $\boldsymbol{\beta}'$, $\boldsymbol{\gamma}'$, d' and s'_Q for the proposed values. Our MCMC method consists of the following steps.

Step 1. Propose d' and s'_Q by simulating $d' \sim \{1, \ldots, d_0\}$ and $s'_Q \sim \{-1, 1\}$ uniformly.

Step 2. Propose $\boldsymbol{\beta}'$ by simulating its components from the following log-normal or truncated log-normal distributions:

$$\begin{aligned}
&\ln a'_{jp} \sim N(\ln a_{jp},\, \tilde{\sigma}^2_{jp}), \quad \ln b'_{jq} \sim N(\ln b_{jq},\, \tilde{s}^2_{jq}), \quad \text{for all } j, p, q \\
&\ln \phi'_{jp} \sim N(\ln \phi_{jp},\, \tilde{\lambda}^2_{jp}), \quad \text{for all } j, p \neq 0 \\
&\ln \phi'_{j0} \sim N(\ln \phi_{j0},\, \tilde{\lambda}^2_{j0}), \quad \text{for all } j \text{ such that } \phi'_{j0} \geq \phi_0
\end{aligned}$$

where $\tilde{\sigma}_{jp}$, $\tilde{s}_{jq}$ and $\tilde{\lambda}_{jp}$ are the scales of the respective log-normal distributions.

Step 3. Propose $\boldsymbol{\gamma}'$ by using the following truncated normal distributions:

For $j = 1, \ldots, k-1$, simulate $\gamma'_j \sim N(\gamma_j, \tilde{\xi}^2_j)$ such that $\gamma'_j \in (a_j, b)$, where $a_1 = \underline{\gamma}$ and $a_j = \gamma'_{j-1}$ for $j > 1$, $b = \bar{\gamma}$, and $\underline{\gamma}$ and $\bar{\gamma}$ are two proper values that define a range of possible threshold values. Hence we have $\underline{\gamma} < \gamma'_1 < \gamma'_2 < \cdots < \gamma'_{k-1} < \bar{\gamma}$. In this paper we take $\underline{\gamma}$ and $\bar{\gamma}$ as the 25% and 75% sample quantiles of x_t respectively. Of course other values within the range of the samples could also be used.

Step 4. Define $\Omega'_j = [\gamma'_{j-1}, \gamma'_j)$ and calculate VaR'_t and g'_t for $t = t_0 + 1, \ldots, T$:

$$\begin{aligned}
\text{VaR}'_t &= s'_Q \sqrt{\textstyle\sum_{j=1}^{k} \left(a'_{j0} + \sum_{p=1}^{p_j} a'_{jp} x^2_{t-p} + \sum_{q=1}^{q_j} b'_{jq} \text{VaR}'^2_{t-q}\right) I\left[x_{t-d'} \in \Omega'_j\right]} \\
g'_t &= \textstyle\sum_{j=1}^{k} \left(\phi'_{j0} + \sum_{p=1}^{p_j} \phi'_{jp} x^2_{t-p} + \sum_{q=1}^{q_j} b'_{jq} g'_{t-q}\right) I\left[x_{t-d'} \in \Omega'_j\right].
\end{aligned}$$

Step 5. Accept the proposed values with probability $\min\{ABC, 1\}$, where A, B and C are given in Appendix VI.

Step 6. If the proposed values are accepted, let $(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q) = (\boldsymbol{\beta}', \boldsymbol{\gamma}', d', s'_Q)$; otherwise, discard the proposed values. Go to Step 1.

The Metropolis-Hastings algorithm construction guarantees that the equilibrium distribution of the Markov chain is the posterior distribution of the model parameters. Hence,

after a burn-in period, the posterior sample of the model parameters can be used to estimate the model parameters. In this paper, the estimates for d and s_Q were taken as the mode of the respective posterior samples because these two parameters take integer values. For all the other parameters, the estimates were taken as the average value of the respective posterior samples.

3 Density forecasting

3.1 Method of density forecasting

Let $\widehat{\text{VaR}}_t$ be the value of VaR_t evaluated at the α-quantile QMLE of the parameters. Our task now is to use $\widehat{\text{VaR}}_t$ to obtain an m-step ahead density forecast for x_t that follows the TGARCH model (1), given the information available up to time T, where $m = 1, \ldots, M$.

Since $x_t = \varepsilon_t\sqrt{h_t}$ by model (1), the τth conditional quantile of x_t is given by $\text{VaR}_t = Q(\tau)\sqrt{h_t}$. This gives $e_t = x_t/\text{VaR}_t = \varepsilon_t/Q(\tau)$. So, if x_t follows the TGARCH model (1), then e_t should be iid. This result leads to the following non-parametric forecasting method.

Step 1. Calculate $\hat{e}_t = x_t/\widehat{\text{VaR}}_t$ for $t = t_0, \ldots, T$.

Step 2. Estimate the distribution of e_t, denoted by $g(e)$, by using $\hat{e}_t$ with a non-parametric method such as kernel density estimation (see, e.g. Silverman, 1986). As the α-quantile QMLE method does not depend on the distribution of ε_t, the estimation of $g(e)$ does not require any further information about ε_t.

Step 3. Given $\mathbf{x}_T$, set $m = 1$:

(a) Calculate $\widehat{\text{VaR}}_{T+m}$

(b) Simulate $e \sim g(e)$

(c) Calculate $\hat{x}_{T+m} = e\,\widehat{\text{VaR}}_{T+m}$

(d) If $m < M$, let $m = m + 1$ and go to (a).

By repeating Step 3 multiple times, a random sample of x_{T+m} values can be produced and this sample can be used to obtain density forecasts and any other predictive quantity of interest about the returns. This explains why our forecasting method differs from other quantile regression based methods in the literature, and why we need VaR_t at a single quantile level τ only. In Section 4, we will study the effects of τ on the performance of the forecasting method.

Although this paper focuses on density forecasting with VaR_t at a single quantile level, it is worth noting that the above method can be modified so that h_t can be used to obtain density forecasts as well. This can be done by replacing $\widehat{\text{VaR}}_t$ in the above method by $\sqrt{\hat{h}_t}$, where $\hat{h}_t$ may be estimated by using the Gaussian QMLE method. However, our results in Section 4 suggest that density forecasts obtained using VaR_t outperform those obtained using h_t.

3.2 Evaluation of density forecasts

We have seen that the forecasting method for x_t following the TGARCH model (1) depends on the assumption that the $e_t = x_t/\text{VaR}_t$ are iid. This implies that it is important to check the autocorrelation structure of the $\hat{e}_t$s before using the forecasting method.

In addition, as $\widehat{\text{VaR}}_t$ is the estimated τth quantile of $x_t|\mathbf{x}_{t-1}$, the τth empirical quantile of the $\hat{e}_t$s should be close to 1, which is equivalent to the τth empirical quantile of $\hat{u}_t = x_t - \widehat{\text{VaR}}_t$ being 0. This can also be checked based on either overall or local coverage probabilities.

To perform this check using local coverage probabilities, we may use the moving window method proposed by Cai et al. (2012). Specifically, first a window width W_0 is selected, and then for $i = 1, \ldots, T - W_0 + 1$, the τth empirical quantile of $\{\hat{e}_t, t = i, \ldots, i + W_0 - 1\}$ and of $\{\hat{u}_t, t = i, \ldots, i + W_0 - 1\}$ are calculated respectively. Finally

these local empirical quantiles can be checked against 0 or 1. Any value of W_0 can be used provided that it is large enough to obtain proper empirical estimates of the local quantiles.

To evaluate and compare density forecasts in practice we propose using the approach of Diebold et al. (1998) . We do this because their approach does not depend on the methods that were used to obtain the density forecasts. We will now outline the basic idea. Let $\hat{f}(x_{T+m})$ be the density forecast of x_{T+m}, and let the probability integral transform of x_{T+m} be defined by $z_m = \int_{-\infty}^{x_{T+m}} \hat{f}(u)du$, where $m = 1, \ldots, M$. Diebold et al. (1998) showed that density forecasts could be checked by testing if the probability integral transforms obtained from the density forecasts are iid $U(0, 1)$. If they are, then good density forecasts have been obtained.

Diebold et al. (1998) further suggested that the Kolmogorov-Smirnov test (KS-test) could be used to check the probability integral transforms, but also pointed out that this test is not constructive in that, if rejection occurs, the test itself provides no guidance as to why. Because of this, Diebold et al. (1998) also suggested using graphical methods such as plotting the empirical distribution of the probability integral transforms together with an associated 95% confidence interval, and then comparing it with the $U(0, 1)$ cumulative distribution function. They also suggested using the autocorrelation function (ACF) plot of $w_m = z_m - \bar{z}$ to check the dependence between these probability integral transforms, where $\bar{z}$ is the mean of z_m.

In our cases, we do not have a mathematical expression for the density forecast at time $T + m$. However, the density forecast can be estimated by using the random sample obtained from the forecasting method. Hence, the probability integral transforms can be easily obtained.

4 Simulation studies

4.1 Simulation study about estimation and forecasting: Single threshold model

Let us consider the following model with one threshold:

$$x_t = \varepsilon_t \sqrt{h_t}, \quad h_t = \begin{cases} 0.2 + 0.25x_{t-1}^2 + 0.7h_{t-1}, & x_{t-1} < 0, \\ 0.1 + 0.15x_{t-1}^2 + 0.85h_{t-1}, & x_{t-1} \geq 0, \end{cases} \tag{9}$$

where $\varepsilon_t \sim N(0,1)$. So, $d = 1, p_1 = 1, q_1 = 1, k = 2$, $\gamma_1 = 0$, $\Omega_0 = (-\infty, 0)$ and $\Omega_1 = (0, \infty)$. For model (9), the one-step ahead VaR_t at a level τ is given by

$$\text{VaR}_t = \begin{cases} s_Q \sqrt{0.2\, Q(\tau)^2 + 0.25\, Q(\tau)^2 x_{t-1}^2 + 0.7\, \text{VaR}_{t-1}^2}\,, & x_{t-1} < 0, \\ s_Q \sqrt{0.1\, Q(\tau)^2 + 0.15\, Q(\tau)^2 x_{t-1}^2 + 0.85\, \text{VaR}_{t-1}^2}\,, & x_{t-1} \geq 0. \end{cases} \tag{10}$$

where $Q(\tau)$ is the τth quantile of ε_t. So, the true parameter values of expression (10) and the values of the ϕ_{jp}s that appear in expression (4) for g_t are known once τ is fixed. Table 1 shows the true parameter values at four different levels of τ.

Table 1: True parameter values at four levels of τ

τ	0.05	0.25	0.75	0.95
a_{10}	0.541	0.091	0.091	0.541
a_{11}	0.676	0.114	0.114	0.676
a_{20}	0.271	0.045	0.045	0.271
a_{21}	0.406	0.068	0.068	0.406
b_{11}	0.700	0.700	0.700	0.700
b_{21}	0.850	0.850	0.850	0.850
s_Q	-1	-1	1	1
ϕ_{10}	0.391	0.208	0.208	0.391
ϕ_{11}	0.489	0.260	0.260	0.489
ϕ_{20}	0.196	0.104	0.104	0.196
ϕ_{21}	0.293	0.156	0.156	0.293

We simulated a time series from model (9) of length 500, shown in Figure 1(a), which

we treat as the observed data.

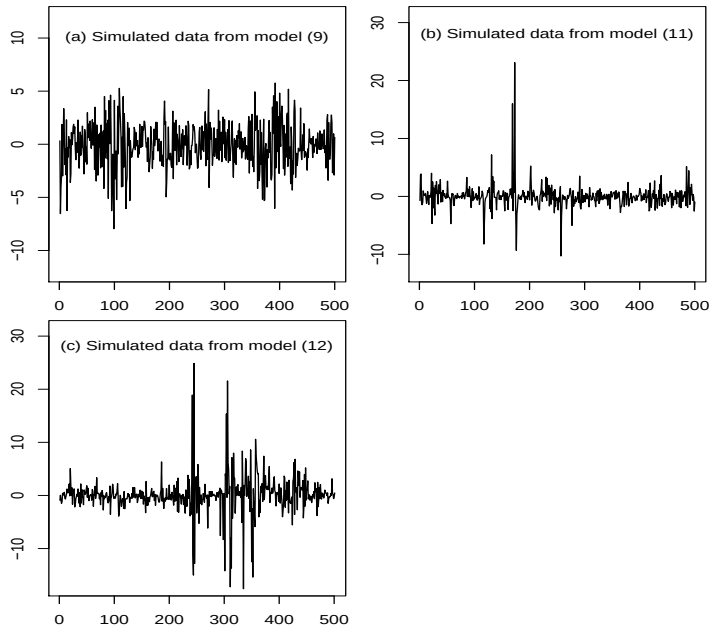

Figure 1: Plots of the simulated time series.

To check the performance of our method, we will (i) estimate the one-step ahead VaR_t at levels $\tau = 0.05, 0.25, 0.75$ and 0.95 where these values have been chosen to cover a broad range; (ii) discuss multiple step ahead density forecasts; and (iii) explore multiple step ahead VaR forecasts at different quantile levels based on these density forecasts.

4.1.1 Model estimation

Let us first estimate the one-step ahead VaR_t defined by (10) at the four τ levels. All initial parameter values required by the MCMC method were chosen by randomly simulating a number between 0 and 1, except for d that was randomly chosen from 1, 2 and 3. The Markov chain was run for 5×10^6 steps and the first 10^5 values were removed as burn-in. Time series plots of the saved parameter values show that the Markov chain has converged for all four values of τ, as illustrated in Figure 2 for $\tau = 0.25$; plots for the other τ values are similar. Table 2 shows the 95% confidence interval of the average estimated parameters

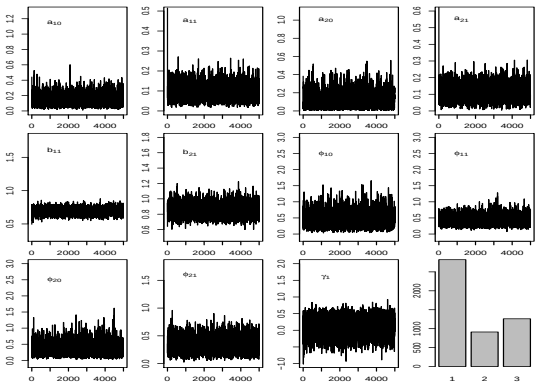

Figure 2: The posterior samples of the model parameters when $\tau = 0.25$.

Table 2: 95% confidence interval of the average estimated parameters over 100 simulations

	TRUE	Lower bound	Upper bound	MSE
a_{10}	0.091	0.074	0.300	0.013
a_{11}	0.114	0.059	0.202	0.002
a_{20}	0.046	0.000	0.259	0.010
a_{21}	0.068	0.007	0.134	0.001
b_{11}	0.700	0.621	0.775	0.002
b_{21}	0.850	0.684	0.866	0.008
γ_1	0.000	-0.181	0.170	0.008

over 100 simulations for $\tau = 0.25$, together with the MSE between each estimate and the true parameter values in model (10). Note that over the 100 simulations the estimated delay parameter d and the sign parameter s_Q were always the same as the true values, i.e. 1 and -1 respectively, and hence are not shown in Table 2. It can be seen from Table 2 that, on average, the true parameters of model (10) are well within the corresponding 95% confidence intervals, and the MSE values further confirm that the true and estimated parameters of model (10) are very close.

For comparison purposes, let $\tilde{e}_t = x_t/\text{VaR}_t$ be the value of e_t evaluated at the true

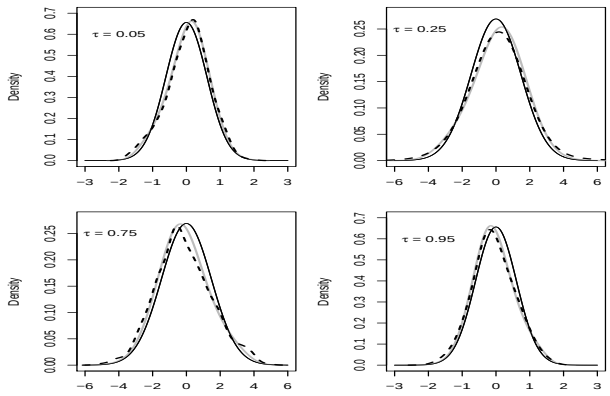

Figure 3: Probability density functions of $\varepsilon_t/Q(\tau)$ (black continuous curve), $\tilde{e}_t$ (grey) and $\hat{e}_t$ (dashed) for $\tau = 0.05, 0.25, 0.75$ and 0.95.

parameter values given in Table 1, so that the distribution of $\tilde{e}_t$ should be very similar to the theoretical distribution of $\varepsilon_t/Q(\tau)$. Figure 3 shows the density function plots of $\hat{e}_t$, $\tilde{e}_t$ and $\varepsilon_t/Q(\tau)$. It can be seen that all these density functions are very similar. Indeed, when the KS-test was used to compare these distributions, the p-values were found to be at least 0.0492, which confirmed that the distributions are not different from each other at a 1% level of significance. We also conducted the Ljung-Box test to check the independence between $\hat{e}_t$. The results show that all p-values of the test are greater than 0.77, strongly supporting that $\hat{e}_t$ are iid. These results suggest that the method performs well.

Two sets of empirical coverage probabilities were calculated by using one-step ahead VaR_t (using the true parameter values) and $\widehat{\text{VaR}}_t$. Table 3 shows that $\widehat{\text{VaR}}_t$'s do have good coverage probabilities at all the quantile levels considered. This is quantified further using RMSE, the square root of the mean squared errors between the τ's and the empirical coverage probabilities.

Table 3: Empirical coverage probabilities at different levels

Using VaR_t					
True probability τ	0.05	0.25	0.75	0.95	RMSE
Count	20	136	397	474	
Estimated probability	0.04	0.272	0.794	0.948	0.01350

Using $\widehat{\text{VaR}}_t$					
True probability τ	0.05	0.25	0.75	0.95	RMSE
Count	28	116	387	477	
Estimated probability	0.056	0.240	0.774	0.954	0.01349

4.1.2 Density forecasts

The following simulation steps were used to obtain a random sample of x_{T+m}, conditional on $\mathbf{x}_T$, as the true conditional density function of $x_{T+m}|\mathbf{x}_T$ is not analytically available: (i) Start from h'_{T+m-1}, x'_{T+m-1} and $\varepsilon'_{T+m-1} \sim N(0,1)$, where $h'_{T+m-1} = h_{T+m-1}$, $x'_{T+m-1} = x_{T+m-1}$, if $T+m-1 \leq T$, and $m = 1, \ldots, M$. We let $M = 25$ in this simulation study, but any other value of M can also be used. (ii) Let $\varepsilon'_{T+m} \sim N(0,1)$, use (9) to calculate h'_{T+m}, and let $x'_{T+m} = \epsilon'_{T+m}\sqrt{h'_{T+m}}$. By repeating these steps multiple times, a random sample of x_{T+m} can be obtained. These samples can be used to estimate the conditional density function of x_{T+m}, denoted by $f_m(x \mid \mathbf{x}_T)$. As $f_m(x \mid \mathbf{x}_T)$ should be very close to the theoretical density, we compare our density forecasts with it.

Now let $\hat{f}_{m,\tau}(x \mid \mathbf{x}_T)$ be the density forecast obtained from our forecasting method by using the estimated parameters values, where $m = 1, \ldots, 25$ and $\tau = 0.05$, 0.25, 0.75 and 0.95. Figure 4 shows the density forecasts for $m = 1, 5, 10, 15, 20$ and 25. Each panel presents nine curves. For example, for $m = 10$, the darker continuous curve corresponds to $f_{10}(x \mid \mathbf{x}_T)$, while the grey curves correspond to $f_{10,\tau}(x \mid \mathbf{x}_T)$ and $\hat{f}_{10,\tau}(x \mid \mathbf{x}_T)$ for $\tau = 0.05, 0.25, 0.75$ and 0.95. Figure 4 suggests that these density forecasts are close to the required $f_m(x \mid \mathbf{x}_T)$. The results of the KS-test show that both $f_{m,\tau}(x \mid \mathbf{x}_T)$ and $\hat{f}_{m,\tau}(x \mid \mathbf{x}_T)$ are not different from $f_m(x \mid \mathbf{x}_T)$ at the 1% level of significance for all values of m and τ except for $(m, \tau) = (1,\ 0.75)$. These results suggest that the effect of τ on the density forecasts is not important.

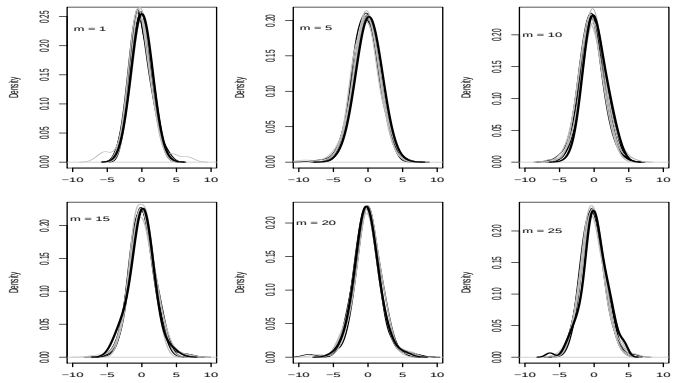

Figure 4: Plots of the predictive density functions $f_m(x \mid \mathbf{x}_T)$ (darker curves), and density forecasts $f_{m,\tau}(x \mid \mathbf{x}_T)$ and $\hat{f}_{m,\tau}(x \mid \mathbf{x}_T)$ (grey curves), where $m = 1, 5, 10, 15, 20, 25$ and $\tau = 0.05, 0.25, 0.75, 0.95$. $f_m(x \mid \mathbf{x}_T)$ represents the theoretical m-step ahead predictive density function of $x_t|\mathbf{x}_{t-1}$, while $f_{m,\tau}(x \mid \mathbf{x}_T)$ and $\hat{f}_{m,\tau}(x \mid \mathbf{x}_T)$ represent the density forecasts obtained from the forecasting method by using the true and estimated parameter values respectively.

4.1.3 VaR forecasts

Any predictive quantity of interest about x_{T+m} given $\mathbf{x}_T$ can be obtained from our density forecasts. For illustration purposes, we consider m-step ahead VaR forecasts in this simulation study due to their importance in finance. We obtained two sets of VaR forecasts, denoted by $\text{VaR}_{T+m,\tau,\tau'}$ and $\widetilde{\text{VaR}}_{T+m,\tau,\tau'}$, where $\text{VaR}_{T+m,\tau,\tau'}$ is the m-step ahead τ'th VaR forecast obtained from $f_{m,\tau}(x \mid \mathbf{x}_T)$, and $\widetilde{\text{VaR}}_{T+m,\tau,\tau'}$ is the same quantity obtained from $\hat{f}_{m,\tau}(x \mid \mathbf{x}_T)$, where $m = 1, \ldots, 25$, $\tau = 0.05, 0.25, 0.75, 0.95$ and $\tau' = 0.001, 0.005, 0.01, 0.05, 0.25, 0.5, 0.75, 0.95, 0.99, 0.995, 0.999$.

We compare these two sets of VaR forecasts by calculating the value of RMSE between $\text{VaR}_{T+m,\tau,\tau'}$ and $\widetilde{\text{VaR}}_{T+m,\tau,\tau'}$. The results are shown in Table 4. We can see that the VaR forecasts $\widetilde{\text{VaR}}_{T+m,\tau,\tau'}$ are similar for different τ values in this simulation study, which further confirms that a VaR_t process at a single quantile level is sufficient for us to obtain density forecasts for market returns that follow TGARCH model (1). However, to avoid potential problems associated with the estimation of VaR_t at extreme levels due to lack of information, we suggest using VaR_t at a non-extreme level for density forecasting and for predicting other quantities of interest about the process under study.

Table 4: RMSE values between m-step ahead VaR forecasts $\text{VaR}_{T+m,\tau,\tau'}$ and $\widetilde{\text{VaR}}_{T+m,\tau,\tau'}$ for $m = 1, \ldots, 25$

$\tau \setminus \tau'$	0.001	0.005	0.01	0.05	0.25	0.50	0.75	0.95	0.99	0.995	0.999
0.05	0.52	0.38	0.35	0.29	0.12	0.03	0.10	0.37	0.75	0.71	0.73
0.25	1.72	1.31	0.99	0.39	0.19	0.04	0.11	0.30	0.94	1.10	1.40
0.75	0.58	0.54	0.67	0.20	0.06	0.02	0.04	0.22	0.36	0.54	0.79
0.95	0.87	0.66	0.59	0.25	0.15	0.04	0.11	0.32	0.73	0.80	1.12

Note that we have also repeated the above simulation study 100 times. The results obtained were consistent with those shown above, and hence are not discussed further.

4.2 Simulation study comparing estimation and forecasting: Double threshold model

This simulation study highlights the advantage of the α-quantile QMLE approach for model estimation and density forecasting for heavy-tailed financial returns that follow TGARCH models. As discussed in Section 1, the variance process h_t of a TGARCH model may be estimated by using the Gaussian QMLE method. Hall and Yao (2003) observed that, for the Gaussian QMLE method with GARCH models, standard asymptotic theory breaks down for financial data with heavy-tails. Although Mikosch & Straumann (2006) provided a new limit theory for parameter estimation in this case, it is still not clear how the Gaussian QMLE method performs for volatility estimation and density forecasting with heavy-tailed financial data compared with the α-quantile QMLE for VaR_t estimation and density forecasting.

Models with more than two regimes have been proven to be very useful in finance. For example, Li and Ling (2012) analyzed real GNP data over the period 1947-2009 and found that a model with three regimes (i.e. two thresholds) explains the data well. Chen et al. (2014) also used a three-regime threshold model to study the process of pair return spread, where the upper and lower regimes in the model are used for trading entry and exit signals. Medeiros and Veiga (2009) studied ten financial indices and showed evidence of three regimes for four out of the ten indices under study. They found that the first, second and third regimes are associated with very negative shocks, tranquil periods and large positive shocks respectively. Due to the usefulness of models with three regimes in finance, we consider TGARCH models with two thresholds in this section.

Specifically, we consider two models, both have two thresholds but with different

volatility persistence. The two models are given below.

$$x_t = \varepsilon_t \sqrt{h_t}, \quad h_t = \begin{cases} 0.1 + 0.2x_{t-1}^2 + 0.09h_{t-1}, & x_{t-2} < -0.2, \\ 0.25 + 0.15x_{t-1}^2 + 0.14h_{t-1}, & -0.2 \leq x_{t-2} < 0.3, \\ 0.9 + 0.15x_{t-1}^2 + 0.4h_{t-1}, & x_{t-2} \geq 0.3, \end{cases} \tag{11}$$

and

$$x_t = \varepsilon_t \sqrt{h_t}, \quad h_t = \begin{cases} 0.1 + 0.2x_{t-1}^2 + 0.55h_{t-1}, & x_{t-2} < -0.2, \\ 0.15 + 0.05x_{t-1}^2 + 0.65h_{t-1}, & -0.2 \leq x_{t-2} < 0.3, \\ 0.3 + 0.15x_{t-1}^2 + 0.62h_{t-1}, & x_{t-2} \geq 0.3, \end{cases} \tag{12}$$

where ε_t follows a t_ν-distribution, where ν represents the degrees of freedom. We let $\nu = 3, 5, 7, 10, 20, 50$ and 100. Moreover, we set the delay parameter $d = 2$ and consider $\tau = 0.75$. Hence $s_Q = \text{sign}(Q(0.75)) = 1$.

Therefore, this simulation study allows us to check the impact of ν on the performance of estimation and forecasting based on the two QMLE methods respectively. It also enables us to check the impact of volatility persistence since model (12) has stronger volatility persistence than that of model (11). For each ν, we simulated 100 time series of size 500 from the two models. Figure 1(b) and (c) show the first of these simulated time series from models (11) and (12) when $\nu = 3$ respectively. For each ν we applied the Gaussian QMLE method to estimate h_t and the α-quantile QMLE method to estimate VaR_t to each of the simulated time series, obtaining 100 estimated values for each of the parameters in h_t and VaR_t respectively.

Let $\hat{h}_{t,\ell}$ and $\widehat{\text{VaR}}_{t,\ell}^{0.75}$ be the estimates of h_t and the 75% VaR_t evaluated at the Gaussian and α-quantile QMLEs using the ℓth simulated time series, respectively. Let $\hat{w}_{t,\ell} = x_t/\sqrt{\hat{h}_{t,\ell}}$ and $\hat{v}_{t,\ell} = x_t/\widehat{\text{VaR}}_{t,\ell}^{0.75}$ be the residuals. Note that we do not include a model indicator and ν in this notation for simplicity.

For each model, let $x_{T+m,i}^{(\ell)}$ be the ith sample of $x_{T+m}|\mathbf{x}_T$ obtained by using the ℓth sim-

ulated time series and the true parameter values, where $i, \ell = 1, \ldots, 100$. Then $\left\{ x^{(\ell)}_{T+m,i} : i, \ell = 1, \ldots, 100 \right\}$ is a pooled sample of $x_{T+m}|\mathbf{x}_T$, which can be used to obtain the true density function of $x_{T+m}|\mathbf{x}_T$. Here, we let $m = 1, \ldots, M$, where $M = 25$. We have experimented with other values of M and obtained comparable results.

Similarly, let $\hat{x}^{(\ell)}_{T+m,i,h}$ and $\hat{x}^{(\ell)}_{T+m,i,V}$ be the samples obtained from our forecasting method based on $\hat{h}_{t,\ell}$ and $\widehat{\text{VaR}}^{0.75}_{t,\ell}$ respectively. Then both $\left\{ \hat{x}^{(\ell)}_{T+m,i,h},\ i = 1, \ldots, 100 \right\}$ and $\left\{ \hat{x}^{(\ell)}_{T+m,i,V},\ i = 1, \ldots, 100 \right\}$ can be used to obtain the ℓth density forecast for $x_{T+m}|\mathbf{x}_T$, where $\ell = 1, \ldots, 100$ and $m = 1, \ldots, 25$. These density forecasts can then be compared with the true density functions of $x_{T+m}|\mathbf{x}_T$.

To quantify the performance of different methods, we first examined model fitting by checking whether $\hat{w}_{t,\ell}$ and $\hat{v}_{t,\ell}\, Q(0.75)$ follow a t_ν distribution by using KS-tests for each simulated series. The first two columns of Table 5 show the number of KS-test null hypothesis rejections out of 100 tests in this situation. For the performance of forecasting, we conducted two sets of 2500 KS-tests for each model to compare the true density function of $x_{T+m}|\mathbf{x}_T$ with those predicted by using $\hat{h}_{t,\ell}$ and $\widehat{\text{VaR}}^{0.75}_{t,\ell}$ respectively. The number of times when the null hypothesis of the KS-test was rejected for each method is given in the last two columns of Table 5.

First let us consider the results on model estimations. Although both QMLE methods deliver consistent estimates, Table 5 shows that for heavy-tailed distributions, i.e. when $\nu < 10$ in our case, the α-quantile QMLE method outperforms the Gaussian QMLE method as the number of KS-test null hypothesis rejections is much smaller. When ν becomes larger, it can be seen that the Gaussian QMLE method provides better estimation results, which is what we should have expected because when ν is large the t-distribution approaches to the standard normal distribution. In this limiting case, the model is correctly specified. On the other hand, the performance of model (11) is better than that of model (12) for almost all values of ν considered here, implying that volatility persistence does have some impact on model estimations with both Gaussian and quantile QMLE methods.

Table 5: Number of null hypothesis rejections out of 100 KS-tests for model fitting and out of 2500 KS-tests for density forecasting.

	$\hat{w}_{t,\ell} \sim t_{df}$ Gaussian QMLE	$\hat{v}_{t,\ell} Q(0.75) \sim t_{df}$ Quantile QMLE	Forecasting Gaussian QMLE	Forecasting Quantile QMLE
$\nu = 3$				
Model (11)	100	14	73	52
Model (12)	100	27	217	120
$\nu = 5$				
Model (11)	66	8	55	45
Model (12)	81	24	74	71
$\nu = 7$				
Model (11)	25	10	67	54
Model (12)	49	29	50	44
$\nu = 10$				
Model (11)	9	9	79	75
Model (12)	20	22	45	43
$\nu = 20$				
Model (11)	2	6	94	74
Model (12)	5	23	44	34
$\nu = 50$				
Model (11)	1	10	98	94
Model (12)	3	23	46	44
$\nu = 100$				
Model (11)	3	8	96	89
Model (12)	2	21	39	35

Now let us consider the results about density forecasting. Table 5 shows that, for all values of ν considered in this study, density forecasts obtained based on VaR_t outperform those obtained based on h_t. On the other hand, compared with model (11), density forecasts with model (12) are worse when $\nu < 7$ but better when $\nu \geq 7$, which is true for both forecasting approaches. This suggests that the impact of volatility persistence on density forecasting depends on the tail behaviour of the distribution of financial returns.

In summary, our results show that as the number of degrees of freedom ν decreases, both inference and forecasting using the α-quantile QMLE method with model (3) are better than inference and forecasting using the Gaussian QMLE method with model (1).

5 Forecasting Hang Seng and S&P500 returns

We now illustrate our method by conducting empirical work on Hang Seng and S&P500 daily closing indices from 3 January 2007 to 13 July 2012. The data are available from Yahoo, with days on which the market was closed having been removed. We are interested in the returns of these daily closing indices. Figures 5(a)-(b) show time series plots of the two indices, while Figures 5(c)-(d) show the associated returns. The autocorrelation structure of the two return series can be seen from Figure 5(e)-(f).

The last 30 values of the returns were not included in our parameter estimation procedure, but were used to evaluate our out-of-sample multiple step ahead forecasts. For each return series, we will (i) estimate the one-step ahead VaR_t at the 75% level, (ii) obtain out-of-sample density forecasts, (iii) compare point forecasts with the actually observed returns, and (iv) make a comparison of our density forecasts with those obtained from other models.

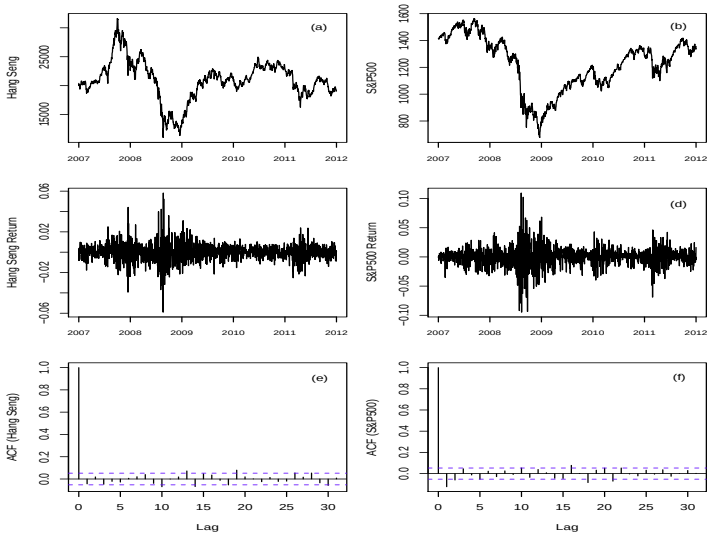

Figure 5: (a)(b) Time series plots, (c)(d) the returns and (e)(f) the autocorrelation function of the returns for the Hang Seng and S&P500 indices from 3 January 2007 to 13 July 2012.

5.1 Results from our method

5.1.1 Estimated models

For the Hang Seng daily closing indices, let x_t be the return value at time t. As Figure 5(e) shows little autocorrelation structure in the return series, our estimation method was employed on the return series directly, resulting in the following estimated model at the 75% level:

$$\text{VaR}_t = \begin{cases} \sqrt{0.0003 + 0.2311\, x_{t-1}^2 + 0.1891\, \text{VaR}_{t-1}^2}\,, & \text{if } x_{t-2} < -0.0090, \\ \sqrt{0.00008 + 0.0997\, x_{t-1}^2 + 0.1627\, \text{VaR}_{t-1}^2}\,, & \text{if } x_{t-2} \geq -0.0090. \end{cases} \tag{13}$$

For the S&P500 daily closing indices, let y_t be the return value at time t. Since Figure 5(f) shows a strong autocorrelation at lags 1 and 2, we first estimated a sequence of ARMA models for the returns and concluded that an AR(2) process satisfactorily modelled the data. Let the residuals of the fitted AR(2) model be $x_t = y_t - \hat{\ell}_1 y_{t-1} - \hat{\ell}_2 y_{t-2}$, where $\hat{\ell}_1 = -0.1344\ (0.0267)$ and $\hat{\ell}_2 = -0.0777\ (0.0267)$ are the estimated parameter values with the standard errors in brackets. By applying our estimation method to x_t, we obtained the following estimated model, again at the 75% level:

$$\text{VaR}_t^x = \begin{cases} \sqrt{0.00018 + 0.2328\, x_{t-1}^2 + 0.2412\, (\text{VaR}_{t-1}^x)^2}\,, & \text{if } x_{t-2} < -0.0059, \\ \sqrt{0.00004 + 0.0881\, x_{t-1}^2 + 0.1858\, (\text{VaR}_{t-1}^x)^2}\,, & \text{if } x_{t-2} \geq -0.0059. \end{cases} \tag{14}$$

The estimated one-step ahead VaR process for the S&P500 returns is then given by

$$\text{VaR}_t^y = \text{VaR}_t^x + \hat{\ell}_1 y_{t-1} + \hat{\ell}_2 y_{t-2}\,. \tag{15}$$

Hence, forecasts of y_t can be obtained from those of x_t by using $y_t = x_t + \hat{\ell}_1 y_{t-1} + \hat{\ell}_2 y_{t-2}$.

Now we use the one-step ahead VaR_t processes (13) and (15) to obtain density forecasts for the Hang Seng and S&P500 return series respectively.

5.1.2 Forecasting results

Using our forecasting method, we obtained density forecasts up to 30 steps ahead for both return series. To make the plots clearer, we only show the density forecasts in Figure 6(a)-(d) for $m = 1, 10, 20$ and 30, where the grey and black curves are the density forecasts for the Hang Seng and S&P500 return series respectively. It can be seen that multiple step ahead conditional density forecasts can be skewed either to the right or left.

Figure 6(e)-(f) show the point forecasts for Hang Seng and S&P500 return series respectively. In each of these plots, the black curve corresponds to the observed returns, the middle grey curve shows the m-step ahead mean forecasts, and the upper and lower grey curves provide the m-step ahead VaR forecasts at levels $\tau = 0.95$ and 0.05 respectively, where $m = 1, \ldots, 30$. The RMSE and the mean absolute difference (MAD) between the predicted and the observed returns are all less than 0.0981 for both return series, which is a small value compared with the range of the values that RMSE and MAD can take. The range between the two m-step ahead VaR forecasts also provides a measure of the volatility of $x_{T+m}|\mathbf{x}_T$. In fact, this is a measure of the variation of the distribution with respect to its median, and so we suggest using it in practice when the distribution of the returns is skewed.

5.2 Comparison of forecasts

5.2.1 Comparison with ARMA-GARCH models

For comparison purposes, we fitted an ARMA-GARCH type model to each of the two return series. For Hang Seng returns, the best fitting model is a GARCH(1,1) given by $y_t = \epsilon_t\sqrt{h_t}$, where $h_t = 3.305 \times 10^{-6} + 0.087y_{t-1}^2 + 0.905h_{t-1}$ and ϵ_t follows a skewed t-distribution with skewness 0.926 and degrees of freedom 10.

For S&P500 returns, the best fitting model was found to be an AR(2)-GARCH(1,1) and is given by $y_t = -0.095y_{t-1} - 0.057y_{t-2} + \varepsilon_t\sqrt{h_t}$, where $h_t = 2.02 \times 10^{-6} + 0.112(y_{t-1} +$

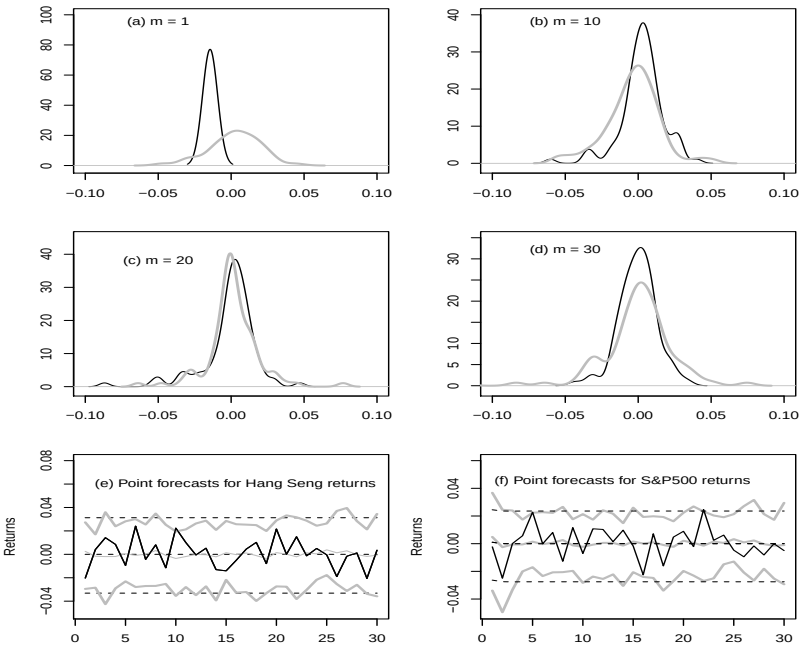

Figure 6: (a)-(d) Density forecasts for Hang Seng (grey curves) and S&P500 (black curves) returns respectively. (e)(f) Point forecasts for Hang Seng and S&P500 return series, where the black curve corresponds to the observed returns, the middle grey curve shows the m-step ahead mean forecasts, where $m = 1, \ldots, 30$, and the upper and lower grey curves provide the m-step ahead VaR forecasts at levels $\tau = 0.95$ and 0.05 respectively. The middle dashed curve shows the m-step ahead mean forecasts from the ARMA-GARCH models and the upper and lower dashed curves give the corresponding 90% prediction intervals.

$0.095y_{t-2} + 0.057y_{t-3})^2 + 0.887h_{t-1}$ and ε_t follows a skewed t-distribution with skewness 0.830 and degrees of freedom just over 6.

For both models we chose to use a skewed t-distribution for ε_t because this distribution can deal with more complex data structures. The mean forecasts for the next 30 days were obtained and are shown as the middle dashed curves in Figure 6(e)-(f), where the upper and lower dashed curves form a 90% prediction interval. It is seen that the mean forecasts from the ARMA-GARCH models decrease to 0 very quickly as expected. Note that the range of the prediction intervals do not provide a good measure for the volatility of the underlying processes.

To obtain density forecasts from the fitted ARMA-GARCH models, a conventional simulation method was used, which provides us with a random sample from the m-step ahead predictive distribution function, where $m = 1, \ldots, 30$. Hence, for each financial return series, we obtained two sets of density forecasts: one from our method and the other from the ARMA-GARCH model. We now test the differences between these density forecasts by using the method of Diebold et al. (1998) discussed in Section 3.2.

Figure 7 shows the ACF of $w_m = z_m - \bar{z}$, where $\bar{z}$ is the mean of the probability integral transforms z_m of x_{T+m}, $m = 1, \ldots, 30$. It can be seen that the correlation between the probability integral transforms is not statistically significant in all cases. So from this point of view, both models perform similarly.

Figure 8 shows the empirical distributions of z_ms (step curves) for the two return series, where the two dashed curves in each panel form a corresponding 95% confidence interval for the distribution of the probability integral transforms, and the grey straight line is the $U(0, 1)$ cumulative distribution function. Figure 8 suggests that for Hang Seng return series, the two sets of density forecasts are satisfactory, which was further confirmed by the KS-test as in both cases we cannot reject the null hypothesis of the test at the 5% level of significance. However, for the S&P500 return series, our method outperforms that of the ARMA-GARCH model, which was also confirmed by the KS-test: for our results the

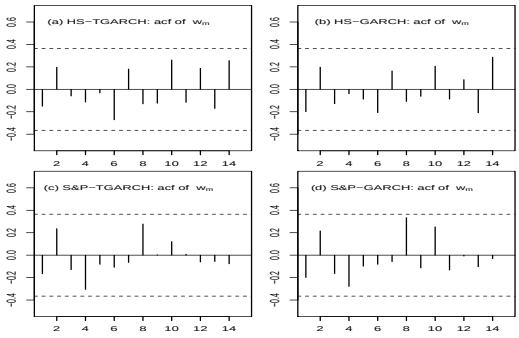

Figure 7: ACF plots of the probability integral transforms z_ms. (a)(c) are for our models, while (b)(d) are for ARMA-GARCH models.

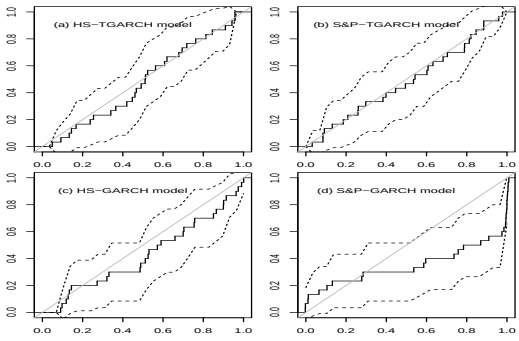

Figure 8: Plots of the empirical distributions of the probability integral transforms z_ms (step curves) together with the associated 95% confidence intervals (dashed curves). The grey line is the $U(0,1)$ cumulative distribution function. (a)(b) are for our models, while (c)(d) are for ARMA-GARCH models.

p-value is 0.96, while for the ARMA-GARCH model the p-value is less than 0.00005.

5.2.2 Comparison with GJR-GARCH models

The GJR-GARCH model of order 1 (Glosten et al., 1993) is defined by $x_t = \sqrt{h_t}\,\varepsilon_t$, where $h_t = \omega + (a + b\,[x_{t-1} < 0])x_{t-1}^2 + \delta h_{t-1}$, and ε_t is the error term of the model. In this comparison, we also let ε_t follow a skewed t-distribution.

For the Hang Seng return series, the estimated GJR-GARCH model is given by $x_t = \sqrt{h_t}\,\varepsilon_t$, where $h_t = 0.4 \times 10^{-5} + (0.035 + 0.104\,I\,[x_{t-1} < 0])x_{t-1}^2 + 0.900 h_{t-1}$ and ε_t follows a skewed t-distribution with skewness 0.931 and degrees of freedom around 18.

For the S&P500 return series, the estimated AR(2)-GJR-GARCH model is given by $y_t = -0.091 y_{t-1} - 0.055 y_{t-2} + \sqrt{h_t}\,\varepsilon_t$, where

$$h_t = 0.3\times10^{-5} + (0.1\times10^{-7} + 0.187\,I\,[x_{t-1} < 0])(y_t + 0.091 y_{t-1} + 0.055 y_{t-2})^2 + 0.897 h_{t-1}^2$$

and ε_t follows a skewed t-distribution with skewness 0.812 and degrees of freedom 6.7.

Figure 9 checks the density forecasts obtained from the above GJR-GARCH models. Clearly, the distribution of the probability integral transforms are away from $U(0,1)$ and there exists strong autocorrelation between these transforms. These results suggest that the density forecasts obtained from the GJR-GARCH models are not satisfactory. One of the possible reasons for this unsatisfactory GJR-GARCH performance could be that both the threshold and the delay parameter values in this model are fixed, rather than being estimated from the data.

6 Further comments and conclusions

This paper develops an α-quantile QMLE method for VaR_t by showing that the associated density function is an α-quantile density and it belongs to the tick-exponential family.

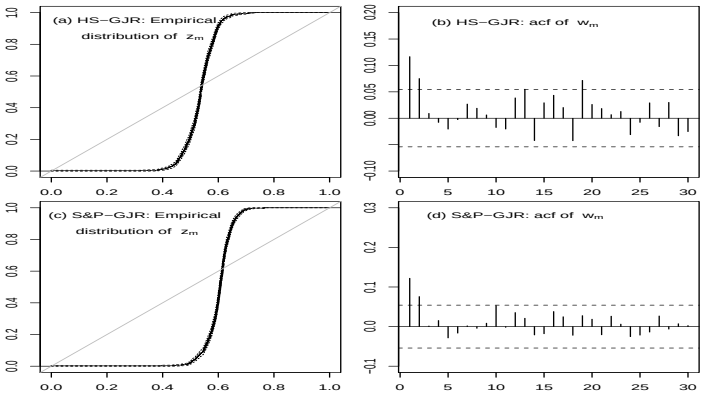

Figure 9: Density forecasting evaluation for GJR-GARCH models: (a)(c) Empirical distribution function plots of z_ms (step curves) together with an associated 95% confidence intervals (dashed curves). (b)(d) ACF plots of the z_ms.

Hence the method guarantees that our estimator is consistent for the parameters of VaR_t. Performing Bayesian inference about VaR_t provides an alternative method for parameter estimation. We also propose simple density forecasting methodology based on the VaR_t at a single non-extreme level. This density forecasting method overcomes the limitations of some existing forecasting methods with quantile models.

For model estimation, our simulation results show that the proposed α-quantile QMLE method for VaR_t outperforms the Gaussian QMLE method for h_t for heavy-tailed financial data that follow TGARCH models. For density forecasting, our results show that density forecasts based on the proposed α-quantile QMLE for VaR_t outperform those based on the Gaussian QMLE for h_t for both heavy-tailed and non-heavy-tailed financial data. Moreover, our results also show that the impact of volatility persistence on model estimation and forecasting depends on the tail behaviour of financial data. The results obtained from the empirical work on market returns show that our forecasting method also outperforms some

benchmark models for density forecasting of financial returns.

In this paper, the number of thresholds and the order of the model are assumed known. However, it is possible to develop a reversible jump MCMC method to estimate these parameters simultaneously with all the other model parameters. We leave this for future work.

Acknowledgement

We are grateful to the editor for very helpful and insightful comments about our research. We also thank an associate editor and referees for constructive suggestions that have led to substantial improvements in this paper.

Appendix I

We want to show that if x_t follows model (1), then VaR_t follows expression (2).

Let $Q(\tau)$ be the τth quantile of ε_t. It follows from (1) that the τth quantile of x_t, i.e. one-step ahead VaR_t , is given by $\text{VaR}_t = Q(\tau)\sqrt{h_t}$. Hence,

$$\begin{aligned}
\text{VaR}_t^2 &= Q^2(\tau)h_t \\
&= Q^2(\tau)\left\{\sum_{j=1}^{k}\left(\alpha_{j0}+\sum_{p=1}^{p_j}\alpha_{jp}x_{t-p}^2+\sum_{q=1}^{q_j}\beta_{jq}h_{t-q}\right)I\left[x_{t-d}\in\Omega_j\right]\right\} \\
&= \sum_{j=1}^{k}\left(Q^2(\tau)\alpha_{j0}+\sum_{p=1}^{p_j}Q^2(\tau)\alpha_{jp}x_{t-p}^2+\sum_{q=1}^{q_j}Q^2(\tau)\beta_{jq}h_{t-q}\right)I\left[x_{t-d}\in\Omega_j\right].
\end{aligned}$$

So it follows from $\text{VaR}_{t-q}^2 = Q^2(\tau)h_{t-q}$ that

$$\text{VaR}_t^2 = \sum_{j=1}^{k}\left(a_{j0}+\sum_{p=1}^{p_j}a_{jp}x_{t-p}^2+\sum_{q=1}^{q_j}b_{jq}\text{VaR}_{t-q}^2\right)I\left[x_{t-d}\in\Omega_j\right],$$

where $a_{jp} = Q^2(\tau)\alpha_{jp}$ and $b_{jq} = \beta_{jq}$. Taking the square root on both sides gives

$$\text{VaR}_t = s_Q\sqrt{\sum_{j=1}^{k}\left(a_{j0}+\sum_{p=1}^{p_j}a_{jp}x_{t-p}^2+\sum_{q=1}^{q_j}b_{jq}\text{VaR}_{t-q}^2\right)I\left[x_{t-d}\in\Omega_j\right]},$$

in which $s_Q = \text{sign}(Q(\tau))$.

Appendix II

We now prove Proposition 1. First recall that v_t are iid with τth quantile 0 and variance 1. It follows from (1) that

$$E[x_t \mid \mathbf{x}_{t-1}] = 0 \text{ and that } E[x_t^2 \mid \mathbf{x}_{t-1}] = \text{Var}[x_t \mid \mathbf{x}_{t-1}] = h_t. \tag{16}$$

Hence, it follows from (3) and (16) that

$$0 = E[x_t \mid \mathbf{x}_{t-1}] = \text{VaR}_t + \sqrt{g_t} E[v_t], \text{ so that } \sqrt{g_t}\, E[v_t] = -\text{VaR}_t,$$

and that

$$\begin{aligned} h_t &= E[x_t^2 \mid \mathbf{x}_{t-1}] = E[(\text{VaR}_t + \sqrt{g_t} v_t)^2 \mid \mathbf{x}_{t-1}] \\ &= \text{VaR}_t^2 + 2\,\text{VaR}_t \sqrt{g_t}\, E[v_t] + g_t E[v_t^2]. \end{aligned}$$

Hence,

$$h_t = \text{VaR}_t^2 - 2\,\text{VaR}_t^2 + g_t\, E[v_t^2], \text{ so that } g_t = \frac{h_t + \text{VaR}_t^2}{E[v_t^2]}$$

as required.

Moreover, it follows from this result, from (2) and from the facts that $\beta_{jq} = b_{jq}$ and v_t are iid that

$$\begin{aligned} g_t &= \frac{h_t + \text{VaR}_t^2}{E[v_t^2]} \\ &= \frac{1}{E[v_t^2]} \textstyle\sum_{j=1}^{k} \left\{ (\alpha_{j0} + a_{j0}) + \sum_{p=1}^{p_j} (\alpha_{jp} + a_{jp}) x_{t-p}^2 + b_{jq}(h_{t-q} + \text{VaR}_{t-q}^2) \right\} I[x_{t-d} \in \Omega_j] \\ &= \textstyle\sum_{j=1}^{k} \left\{ \frac{\alpha_{j0} + a_{j0}}{E[v_t^2]} + \sum_{p=1}^{p_j} \frac{\alpha_{jp} + a_{jp}}{E[v_t^2]}\, x_{t-p}^2 + b_{jq}\, \frac{h_{t-q} + \text{VaR}_{t-q}^2}{E[v_t^2]} \right\} I[x_{t-d} \in \Omega_j] \\ &= \textstyle\sum_{j=1}^{k} \left\{ \phi_{j0} + \sum_{p=1}^{p_j} \phi_{jp} x_{t-p}^2 + b_{jq} g_{t-q} \right\} I[x_{t-d} \in \Omega_j], \end{aligned}$$

where $\phi_{jp} = (\alpha_{jp} + a_{jp})/E[v_t^2]$ for $j = 1, \ldots, k$ and $p = 0, \ldots, p_j$. This completes the

proof.

Appendix III

First note that τ in the density defined by (5) corresponds to α in the α-quantile density and the tick-exponential family defined by Komunjer (2005a, b).

For the first part of the Proposition 2, we need to show that the density function (5) can be written in the form

$$f_\alpha(u;\theta,\phi) = \begin{cases} \frac{2\alpha(1-\alpha)}{\phi}\exp\left[-2(1-\alpha)\frac{|u-\theta|}{\phi}\right], & u \le \theta, \\ \frac{2\alpha(1-\alpha)}{\phi}\exp\left[-2\alpha\frac{|u-\theta|}{\phi}\right], & u > \theta, \end{cases}$$

where $0 < \alpha < 1$, $\phi > 0$ and $\theta \in R$. This is the definition given by Komunjer (2005a).

We proceed as follows.

$$f(v) = \sqrt{1-2\tau+2\tau^2}\,\exp\left\{\frac{v\sqrt{1-2\tau+2\tau^2}}{\tau - I[v\ge 0]}\right\}$$

$$= \sqrt{1-2\tau+2\tau^2}\,\exp\left\{-\frac{v\sqrt{1-2\tau+2\tau^2}}{\tau(1-\tau)}(\tau - I[v \le 0])\right\}$$

$$= \frac{2\tau(1-\tau)}{2\tau(1-\tau)/\sqrt{1-2\tau+2\tau^2}}\,\exp\left\{-\frac{2v(\tau - I[(v\le 0)])}{2\tau(1-\tau)/\sqrt{1-2\tau+2\tau^2}}\right\}$$

$$= \begin{cases} \frac{2\tau(1-\tau)}{\phi}\,\exp\left\{-\frac{2(1-\tau)|v|}{\phi}\right\} & \text{if } v \le 0, \\ \frac{2\tau(1-\tau)}{\phi}\,\exp\left\{-\frac{2\tau|v|}{\phi}\right\} & \text{if } v > 0, \end{cases}$$

where $\phi = 2\tau(1-\tau)/\sqrt{1-2\tau+2\tau^2}$. Comparing this with the above definition, we see that the density function (5) is an α-quantile density with $\phi = 2\tau(1-\tau)/\sqrt{1-2\tau+2\tau^2}$ and $\theta = 0$. This completes the first part of the proof.

For the second part, we need to show that the density function defined by (5) can be

written as

$$f_\alpha(y,\mu) = \exp\left\{-(1-\tau)[a_t(\mu) - b_t(y)]I[y \le \mu] + \tau[a_t(\mu) - c_t(y)]I[y > \mu]\right\}$$

for some functions $a_t(\mu), b_t(y)$ and $c_t(y)$. This corresponds to the definition given by Komunjer (2005b).

First note that

$$f(v) = \sqrt{1-2\tau+2\tau^2}\,\exp\left\{\tfrac{v\sqrt{1-2\tau+2\tau^2}}{\tau - I[v\le 0]}\right\}$$

$$= \sqrt{1-2\tau+2\tau^2}\,\exp\left\{-\tfrac{v\sqrt{1-2\tau+2\tau^2}}{\tau(1-\tau)}(\tau - I[v \le 0])\right\}.$$

Then by letting $y = v\sqrt{1-2\tau+2\tau^2}$, we have

$$f(y) = \exp\left\{-\tfrac{y}{\tau(1-\tau)}(\tau - I[y \le 0])\right\}$$

$$= \exp\left\{-(1-\tau)\tfrac{-y}{\tau(1-\tau)}I[y \le 0] + \tau\tfrac{-y}{\tau(1-\tau)}I[y > 0]\right\}$$

$$= \exp\left\{-(1-\tau)\left[\tfrac{0}{\tau(1-\tau)} - \tfrac{y}{\tau(1-\tau)}\right]I[y \le 0] + \tau\left[\tfrac{0}{\tau(1-\tau)} - \tfrac{y}{\tau(1-\tau)}\right]I[y > 0]\right\}$$

$$= \exp\left\{-(1-\tau)\left[a_t(\mu) - b_t(y)\right]I[y \le 0] + \tau\left[a_t(\mu) - c_t(y)\right]I[y > 0]\right\},$$

where $a_t(\mu) = 0/\{\tau(1-\tau)\} = 0$ and $b_t(y) = c_t(y) = y/\{\tau(1-\tau)\}$. Hence, the density defined by (5) also belongs to the tick-exponential family. This completes the proof.

Appendix IV

To prove Proposition 3, we need to show that

$$\int_{\Omega} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) d\boldsymbol{\theta} < \infty.$$

Note that, for any $\boldsymbol{\theta} \in \Omega$, we have $g_t \geq \sum_{j=1}^{k} \phi_{j0} I\left[x_{t-d} \in \Omega_j\right] \geq \phi_0 > 0$, and that

$$\exp\left[\sqrt{1-2\tau+2\tau^2}\ \left(x_t - \text{VaR}_t\right) / \left\{\sqrt{g_t}\ \left(\tau - I\left[x_t \geq \text{VaR}_t\right]\right)\right\}\right] \leq 1.$$

Hence, using (7), we have

$$\begin{aligned}
& \int_{\Omega} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) d\boldsymbol{\theta} \\
\leq\ & \tilde{M} \prod_{t=t_0+1}^{T} \tfrac{1}{\sqrt{\phi_0}} \int_{\Omega} \pi_{\beta}(\boldsymbol{\beta}) \pi_{\gamma}(\boldsymbol{\gamma}) \pi_d(d) \pi_s(s_Q) d\boldsymbol{\theta} \\
=\ & \frac{\tilde{M}}{\phi_0^{(T-t_0)/2}} \int_{\Omega} \pi_{\beta}(\boldsymbol{\beta}) \pi_{\gamma}(\boldsymbol{\gamma}) \pi_d(d) \pi_s(s_Q) d\boldsymbol{\theta}.
\end{aligned}$$

Therefore, if the prior density function is well defined on Ω, i.e. if $\int_{\Omega} \pi_{\beta}(\boldsymbol{\beta}) \pi_{\gamma}(\boldsymbol{\gamma}) \pi_d(d) \pi_s(s_Q) d\boldsymbol{\theta} < \infty$, we must have $\int_{\Omega} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) d\boldsymbol{\theta} < \infty$. That it, the posterior distribution is well defined on Ω.

Appendix V

Since d and s_Q are uniformly distributed on $\{1, \ldots, d_0\}$ and $\{-1, 1\}$ respectively, $\pi_d(d)$ and $\pi_s(s_Q)$ are constant. The prior density functions of a_{jp}, ϕ_{jp}, b_{jq}, and γ_j are given by

$$\pi_{a_{jp}} = \frac{1}{a_{jp}\sigma_{jp}\sqrt{2\pi}} e^{-\ln^2(a_{jp})/2\sigma_{jp}^2}, \quad \pi_{\phi_{jp}} = \frac{1}{\phi_{jp}\lambda_{jp}\sqrt{2\pi}} e^{-\ln^2(\phi_{jp})/2\lambda_{jp}^2},$$

$$\pi_{b_{jq}} = \frac{1}{b_{jq}s_{jq}\sqrt{2\pi}} e^{-\ln^2(b_{jq})/2s_{jq}^2}, \quad \pi_{\gamma_j} = \frac{1}{\xi_j\sqrt{2\pi}} e^{-\gamma_j^2/2\xi_j^2},$$

where σ_{jp}, s_{jq}, λ_{jp} and ξ_j are the scale parameters of these log-normal or normal distributions respectively.

Hence, the prior density function $\pi_{\boldsymbol{\theta}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0})$ is given by

$$\pi_{\boldsymbol{\theta}}(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) = \pi_\beta(\boldsymbol{\beta})\pi_\gamma(\boldsymbol{\gamma})\pi_d(d)\pi_s(s_Q)$$

$$\propto \prod_{j=1}^{k} \prod_{p=0}^{p_j} \{e^{-\ln^2(a_{jp})/2\sigma_{jp}^2}/a_{jp}\sigma_{jp}\}\{e^{-\ln^2(\phi_{jp})/2\lambda_{jp}^2}/\phi_{jp}\lambda_{jp}\}$$

$$\times \prod_{q=1}^{q_j} \{e^{-\ln^2(b_{jq})/2s_{jq}^2}/b_{jq}s_{jq}\}\{e^{-\gamma_j^2/2\xi_j^2}/\xi_j\}$$

Appendix VI

The general formula for calculating the Metropolis-Hastings algorithm acceptance probability can be found in Brooks (1998). This formula involves the ratios of the likelihood functions and the prior probability density functions at the proposed and at the current values, multiplied by the ratio of appropriately defined transition densities. In the following we derive the formulae required for calculating the acceptance probability for our MCMC method.

$$A = L(\boldsymbol{\beta}', \boldsymbol{\gamma}', d', s_Q' \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0})/L(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid \mathbf{x}_T, k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0})$$

$$\propto \prod_{t=t_0+1}^{T} \left\{ \exp\left(a'/b'\right) / \sqrt{g_t'} \right\} / \{ \exp\left(a/b\right) / \sqrt{g_t} \}$$

$$= \prod_{t=t_0+1}^{T} \left(\sqrt{g_t} / \sqrt{g_t'} \right) \exp\left(a'/b' - a/b\right)$$

where

$$a' = \sqrt{1 - 2\tau + 2\tau^2}\, u_t', \quad a = \sqrt{1 - 2\tau + 2\tau^2}\, u_t,$$
$$b' = (\tau - I\left[u_t' \geq 0\right])\sqrt{g_t'}, \quad b = (\tau - I\left[u_t \geq 0\right])\sqrt{g_t}.$$

$$
\begin{aligned}
B &= \pi_\theta(\boldsymbol{\beta}', \boldsymbol{\gamma}', d', s'_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) / \pi_\theta(\boldsymbol{\beta}, \boldsymbol{\gamma}, d, s_Q \mid k, \mathbf{p}, \mathbf{q}, \mathbf{x}_{t_0}, \mathbf{V}_{t_0}, \mathbf{g}_{t_0}) \\
&= \textstyle\prod_{j=1}^{k} \prod_{p=0}^{p_j} \left(a_{jp}/a'_{jp}\right) \exp\{-(\ln^2 a'_{jp} - \ln^2 a_{jp})/2\sigma_{jp}^2\} \\
&\times \left(\phi_{jp}/\phi'_{jp}\right) \exp\{-(\ln^2 \phi'_{jp} - \ln^2 \phi_{jp})/2\lambda_{jp}^2\} \\
&\times \textstyle\prod_{q=1}^{q_j} \left(b_{jq}/b'_{jq}\right) \exp\{-(\ln^2 b'_{jq} - \ln^2 b_{jq})/2s_{jq}^2\} \\
&\times \textstyle\prod_{j=1}^{k} \exp\{-(\gamma_j'^2 - \gamma_j^2)/2\xi_j^2\}
\end{aligned}
$$

and $C = C_1/C_2$, in which

$$
\begin{aligned}
C_1 &= q(\boldsymbol{\beta}' \to \boldsymbol{\beta})\, q(\boldsymbol{\gamma}' \to \boldsymbol{\gamma})\, q(d' \to d)\, q(s'_Q \to s_Q) \\
C_2 &= q(\boldsymbol{\beta} \to \boldsymbol{\beta}')\, q(\boldsymbol{\gamma} \to \boldsymbol{\gamma}')\, q(d \to d')\, q(s_Q \to s'_Q),
\end{aligned}
$$

$q(a \to b)$ represents the transition probability density function of b conditional on a, and

$$
\begin{aligned}
&q(\boldsymbol{\beta}' \to \boldsymbol{\beta}) / q(\boldsymbol{\beta} \to \boldsymbol{\beta}') \\
&= \textstyle\prod_{j=1}^{k} \prod_{p=1}^{p_j} \frac{a'_{jp}\phi'_{jp}}{a_{jp}\phi_{jp}} \prod_{q=1}^{q_j} \frac{b'_{jq}}{b_{jq}} \prod_{j=1}^{k} \frac{\phi'_{j0}\left[1 - \Phi\left\{(\ln \phi'_{j0} - \ln \phi_0)/\tilde{\lambda}_{j0}\right\}\right]}{\phi_{j0}\left[1 - \Phi\left\{(\ln \phi_{j0} - \ln \phi_0)/\tilde{\lambda}_{j0}\right\}\right]}.
\end{aligned}
$$

Since d' is simulated uniformly on $\{1, \ldots, d_0\}$, $q(d' \to d)/q(d \to d') = 1$; similarly, $q(s'_Q \to s_Q)/q(s_Q \to s'_Q) = 1$. Finally,

$$
\frac{q(\boldsymbol{\gamma}' \to \boldsymbol{\gamma})}{q(\boldsymbol{\gamma} \to \boldsymbol{\gamma}')} = \prod_{j=1}^{k-1} \frac{\Phi((b - \gamma_j)/\tilde{\xi}_j) - \Phi((\gamma'_{j-1} - \gamma_j)/\tilde{\xi}_j)}{\Phi((b - \gamma'_j)/\tilde{\xi}_j) - \Phi((\gamma_{j-1} - \gamma'_j)/\tilde{\xi}_j)}
$$

where $\gamma_0 = \underline{\gamma}$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function.

References

[1] Brooks, S.P. (1998). Markov chain Monte Carlo method and its application. *The Statistician*, 47, 69–100.

[2] Cai, Y. (2010). Forecasting for quantile self exciting threshold autoregressive time series models. *Biometrika*, **97**, 199–208.

[3] Cai, Y., Stander, J. and Davies, N. (2012). A new Bayesian approach to quantile autoregressive time series model estimation and forecasting. *Journal of Time Series Analysis*, 33, 684–698.

[4] Chen, C.W.S., Chen, M. and Chen, S. (2014). Pairs Trading via Three-Regime Threshold Autoregressive GARCH Models. In: Huynh VN., Kreinovich V., Sriboonchitta S. eds *Modeling Dependence in Econometrics. Advances in Intelligent Systems and Computing* 251, 127-140. Springer, Cham.

[5] Chen, C.W.S., Gerlach, R.H. and Wei, D.C.M. (2009). Bayesian causal effects in quantiles: Accounting for heteroscedasticity. *Computational Statistics and Data Analysis*, 53, 1993–2007.

[6] Davidson, A.C. (2001). *Biometrika* centenary: Theory and general methodology. *Biometrika*, 88, 13–52.

[7] Davison, A. C. (2003). *Statistical Models*. Cambridge: Cambridge University Press.

[8] Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998). Evaluating density forecasts. *International Economic Review*, 39, 863–883.

[9] Engle, R.F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, 22, 367–381.

[10] Faraway, J.J. (2016). *Extending the Linear Model with R*, 2nd edn, FL: Chapman and Hall/CRC.

[11] Firth, D. (1993). Recent developments in quasi-likelihood methods. *Bulletin of the 49th Session of the International Statistical Institute*, 341–358.

[12] Gamerman, D. and Lopes, H.F. (2002). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd edn, FL: Chapman and Hall/CRC.

[13] Geyer, C.J. (2011). Introduction to Markov Chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo*. (eds S. Brooks, A. Gelman, G. K. Jones and Meng X. L.) FL: CRC Press, 3–48.

[14] 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.

[15] Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, 71, 285317.

[16] Heyde, C.C. (1997). *Quasi-likelihood and its Application: A General Approach to Optimal Parameter Estimation*. New York: Springer.

[17] Koenker, R. (2005). *Quantiles Regression*. Cambridge University Press.

[18] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33-50.

[19] Komunjer, I. (2005a). Quasi-maximum likelihood estimation for conditional quantiles. `https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=NASM2002&paper_id=652`.

[20] Komunjer, I. (2005b). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, 128, 137-164.

[21] Kuan, C.M. (2004) The quasi-maximum likelihood method: theory. Chapter 9 of `http://homepage.ntu.edu.tw/~ckuan/pdf/et01/ch9.pdf`.

[22] Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *Journal of Econometrics* 167, 240–253.

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

[24] McCullagh, P. (1991). Quasi-likelihood and estimating functions. In *Statistical Theory and Modelling: In Honour of Sir David Cox, FRS*, eds D. V. Hinkley, N. Reid and E. J. Snell, 265–286. London: Chapman & Hall.

[25] McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn, London: Chapman and Hall/CRC.

[26] McCulloch, R.E. and Tsay, R.S. (1993) Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association*, 88, 968–978.

[27] McCulloch, R.E. and Tsay, R.S. (1994) Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis*, 15, 235–250.

[28] Medeiros, M.C. and Veiga, A. (2009). Modeling Multiple Regimes in Financial Volatility with a Flexible Coefficient GARCH(1, 1) Model. *Econometric Theory*, 25, 117-161.

[29] Mikosch, T. and Straumann, D. (2006). Stable limits of martingale transforms with application to estimation of GARCH parameters. *Annals of Statistics*, 34, 493–522.

[30] Park, J.A., Baek, J.S. and Hwang, S.Y. (2009). Persistent-threshold-GARCH processes: Model and application. *Statistics & Probability Letters*, 79, 907–914.

[31] Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Oxford University Press.

[32] Rosenberg, M.A. and Young, V.R. (1999). A Bayesian approach to understanding time series data. *North American Acturial Journal*, 3, 130–143.

[33] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

[34] Taylor, J.W. (2005). Generating volatility forecasts from Value at Risk estimates. *Management Science*, 51, 712–725.

[35] Timmermann, A. (2000). Density forecasting in economics and finance. *Journal of Forecasting* 19, 231–234.

[36] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalize linear models and the Gauss-Newton method. *Biometrika*, 61, 439–447.

[37] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.

[38] White, H. (1994). *Estimation, Inference, and Specification Analysis*. New York: Cambridge University Press.

[39] 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.

[40] Yu, P.L.H., Li, W.K. and Jin, S. (2010). On some models for Value at Risk. *Econometric Reviews*, 29, 622–641.

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