Composite quantile regression for GARCH models using high-frequency data

Meng Wang\textsuperscript{a,1}, Zhao Chen\textsuperscript{b,2}, Christina Dan Wang\textsuperscript{c,3,*}

\textsuperscript{a} Department of Statistics and Finance, University of Science and Technology of China, Hefei, Anhui 230026, China
\textsuperscript{b} Department of Statistics, Pennsylvania State University, State College, PA 16802, United States
\textsuperscript{c} Department of Statistics, Columbia University, New York, NY 10027, United States

A R T I C L E   I N F O

Article history:
Received 16 April 2016
Revised 22 November 2016
Accepted 22 November 2016
Available online xxx

Keywords:
GARCH models
Composite quantile regression
Volatility proxy
High-frequency data
Robustness
Asymptotic normality
Efficiency

A B S T R A C T

The composite quantile regression (CQR) method is newly proposed to estimate the generalized autoregressive conditional heteroskedasticity (GARCH) models, with the help of high-frequency data. High-frequency intraday log-return processes are embedded into the daily GARCH models to generate the corresponding volatility proxies. Based on proxies, the parameter estimation of GARCH model is derived through the composite quantile regression. The consistency and the asymptotic normality of the proposed estimator are obtained under mild conditions on the innovation processes. To examine the finite sample performance of our newly proposed method, simulation studies are conducted with comparison to several existing estimators of the GARCH model. From the simulation studies, it can be concluded that the proposed CQR estimator is robust and more efficient. An empirical analysis on high-frequency data is presented to illustrate the new methodology.

© 2017 ECOSTA ECONOMETRICS AND STATISTICS. Published by Elsevier B.V. All rights reserved.

1. Introduction

Generalized autoregressive conditional heteroskedasticity (GARCH) model, introduced by Bollerslev (1986), has been very successful in modeling the financial time series and volatilities. In the past decades, the model has evolved into many GARCH-type models such as the E-GARCH model (Nelson, 1991), the GJR model (Glosten et al., 1993), and the P-GARCH model (Bollerslev and Ghysels, 1996). In recent years, new models based on GARCH have been developed to take advantage of the emerging high-frequency data. High-frequency data have become widely available due to the development of information technology and automatic trading system. The benefits of incorporating high-frequency data into modeling have been discovered by researchers and traders. Visser (2011) proposed the high-frequency GARCH model by embedding intraday log-return processes into daily GARCH process. He showed that, using suitable volatility proxies based on high frequency data, the new model can outperform the GARCH model that only considers daily close-to-close data.

In high-frequency GARCH models, estimation of parameters is the fundamental statistical problem. The traditional maximum likelihood estimation (MLE) is the most efficient method when the conditional distribution of log-returns is Gaussian.
(Straumann and Mikosh, 2006). Without Gaussian assumption, the quasi maximum likelihood estimation performs well for symmetric conditional distribution with finite fourth moment (Aknouche and Bibi, 2007). However, the parameter estimation of the high-frequency GARCH models faces many challenges. First, for a given volatility proxy (Visser, 2011), the conditional distribution of the proxy depends both on the intraday log-return processes and on the structure of the proxy. The distribution may not be Gaussian, and is hard to obtain. Second, high-frequency data bring microstructure noises. Because of the microstructure noise, the volatilities of high-frequency data become larger as data frequency increases. Meanwhile, the accumulative microstructure noises may lead to biased estimators as indicated in many papers. Zhang et al. (2003) found the evidence that the realized variances may not converge to the integrated volatility as sample frequency increases. Third, the distribution of high-frequency financial data is usually heavy-tailed and asymmetric. Due to the challenges, the classic maximum likelihood estimation (MLE) and quasi maximum likelihood fail to get efficient and accurate estimators. MLE suffers from the following problems: outliers, negative skewness, and accumulative errors. These problems are further amplified in volatility process due to the quadratic structure in normal density. To overcome the aforementioned problems, robust statistical procedures should be take into account as a substitute for MLE and QMLE.

Quantile regression (QR) has been considered as a robust estimation and widely applied to GARCH type models and other stochastic volatility models in many studies. Koenker and Zhao (1996) studied QR for linear ARCH model. Peng and Yao (2003) considered the least absolute deviation estimation for GARCH model. Engle and Manganelli (2004) applied QR method to estimate the value at risk of asset returns. Xiao and Koenker (2009) studied the quantile regression for linear GARCH model by a truncated ARCH($\infty$) representation of volatilities. Lee and Noh (2011) proved that the impact of the volatility’s initial value is asymptotically negligible when using QR method for GARCH models. Because the QR method depends on the choice of quantile $\tau$, it may not be as efficient as other methods. To improve the efficiency, Zou and Yuan (2008) introduced composite quantile regression (CQR) as a more efficient and robust method. CQR can take full advantage of the information at different quantiles. Thereafter, Kai et al. (2010) proposed local composite quantile regression smoothing to nonparametric regression. Kai et al. (2011) combined variable selection technique and CQR for the varying coefficients partially linear model. Fan et al. (2013) applied CQR to the single-index model. However, there are few applications of CQR for financial models in the existing literature.

In this paper, we propose using composite quantile regression to estimate the parameters in high-frequency GARCH model. Using volatility proxy, we embed the intraday high-frequency log-return processes into the GARCH model. We estimate the parameters of the high-frequency GARCH model by CQR and establish the corresponding consistency and asymptotic normality under mild conditions. With numerical studies, we also show that the CQR estimators are more robust than MLE.

The rest of this paper is organized as follows: in Section 2, we set up the model, introduce how to embed intraday log-return processes into GARCH model and construct the proxy. We propose the composite quantile regression method with proxies and obtain the estimators of the GARCH parameters. In Section 3, we provide the consistency and asymptotic normality of the proposed estimator. In Section 4, we present Monte Carlo studies and empirical analysis to examine the finite sample performance of newly proposed procedure. All technical details and proofs are provided in the Appendix.

2. Composite quantile regression with proxies

2.1. GARCH model with high-frequency data

Suppose the daily log-return process $\{r_t, t \geq 1\}$ follows the GARCH($p, q$) model:

$$r_t = \psi_t, \eta_t,$$

$$v_t^2 = 1 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_j v_{t-j}^2,$$  \hspace{1cm} (2.1)

where the innovations $\{\eta_t, t \geq 1\}$ are independent and identically distributed random variables. Here $\{\eta_t, t \geq 1\}$ are not necessarily Gaussian distributed. For each day $t$, we observe a sequence of high-frequency returns data. Denote by $R_t(\cdot)$ the intraday log-return process. Thus, $R_t(0)$ means the overnight log-return and $R_t(1) = r_t$ after normalizing the intraday observation time to $[0, 1]$. We naturally consider that $R_t(\cdot)$ is continuous due to highly frequent observations. Furthermore, we assume that the volatility $\psi_t$ is a constant within each day $t$. Let $\Phi_t(\cdot) = R_t(\cdot)/\psi_t$, we embed the continuous time series $R_t(\cdot)$ into GARCH($p, q$) model:

$$R_t(u) = \psi_t, \Phi_t(u),$$

$$v_t^2 = 1 + \sum_{i=1}^{q} \alpha_i \Phi_{t-i}^2 + \sum_{j=1}^{p} \beta_j v_{t-j}^2,$$  \hspace{1cm} (2.2)

where the process $\Phi_t(\cdot)$ is continuous, independent and identically distributed over different trading days. We are able to recover the close-to-close return $r_t$ by setting $u = 1$ in Eq. (2.2). The proxies for volatility $\psi_t$ (Visser, 2011) are defined in the following way.

Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.econsta.2016.11.004
**Definition 1.** The random variable $H_t = H(R_t)$ is a proxy whenever $H$ is positive and positively homogeneous in $R_t$ that is $H(s R_t) = s H(R_t)$, for any constant $s \geq 0$.

There are several choices for the proxy. A canonical method to get the daily volatility from $R_t$ is to use the five-minute realized quadratic variation ($RQV_t$). $RQV_t$ is the sum of squared five-minute increments in the $t$th trading day. The realized volatility, $H_t = \sqrt{RQV_t}$, is a classic volatility proxy and performs well in practice. Other common proxies include the daily high-low range $H_t = \max_u R_t(u) - \min_u R_t(u)$, and the absolute return $H_t = |R_t|$. When $R_t$ is symmetrically distributed, $|R_t|$ is equivalent to $R_t$ under regression setting.

By positive homogeneity $H_t = H(R_t) = v_t H(\Psi_t)$, we rewrite the model (2.2) as:

$$H_t = v_t \epsilon_t,$$

$$v_t^2 = 1 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_j v_{t-j}^2,$$

(2.3)

where $\epsilon_t = H(\Psi_t)$ are independent and identically distributed positive random variables for different trading days. Note that $\sqrt{E\epsilon_t^2}$ and $\sqrt{E\epsilon_t^2}$ are the normalizing parameters mentioned in Visser (2011). Also note that the GARCH coefficients $\{\alpha_i\}_{i=1}^{q}$ and $\{\beta_j\}_{j=1}^{p}$ are invariant when $H_t$ changes to $s H_t$. Hereinafter we only focus on the GARCH proxies model (2.3).

### 2.2. Composite quantile regression for GARCH proxies model

To overcome the problems mentioned in the introduction: unknown conditional distribution of the proxy; accumulated microstructure noises; infinite high order moment, we propose the composite quantile regression to model (2.3). Let $\alpha = (\alpha_1, \ldots, \alpha_q)^T$, $\beta = (\beta_1, \ldots, \beta_p)^T$, $\gamma = (\gamma_1, \gamma_2)^T$ and $\alpha^* = (\alpha_1^*, \ldots, \alpha_q^*)^T$, $\beta^* = (\beta_1^*, \ldots, \beta_p^*)^T$, $\gamma^* = (\gamma_1^*, \gamma_2^*)^T$. The superscript * stands for the true values of parameters throughout the paper. To ensure the existence of a unique strictly stationary and ergodic solution of model (2.1), Bougerol and Picard (1992) provided a sufficient and necessary condition that the top Lyapunov exponent is strictly negative. For GARCH(1,1) model, $E(\log(\beta_1^* + \alpha_1^* \epsilon_t^2)) < 0$ ensures the existence of a unique strictly stationary and ergodic solution. In this paper, we adopt the similar assumption that

$$\omega^* \sum_{i=1}^{q} \alpha_i^* + \sum_{j=1}^{p} \beta_j^* < 1,$$

(2.4)

where $\omega^* = \text{Var}(\eta_t)$. This assumption not only guarantees the existence of a unique strictly stationary solution but also the finite second moment of $R_t$ (Francq and Zakoian, 2004).

We employ the equally spaced $K$ quantiles, $\tau_k = \frac{k}{K}$, for $1 \leq k \leq K$, for the composite quantile regression. Denote $P(\epsilon_t \leq \xi_k^*) = \tau_k$. The conditional $\tau_k$ quantile function of $H_t$ based on the past observations is

$$Q_k(\tau_k \mid \mathcal{F}_{t-1}) = \left(1 + \sum_{i=1}^{q} \alpha_i^* r_{t-i}^2 + \sum_{j=1}^{p} \beta_j^* v_{t-j}^2\right)^{1/2} \xi_k^*.$$

(2.5)

where $\mathcal{F}$ is the $\sigma$-field generated by $(R_t(s), s \leq t)$. Let

$$\xi^* = (\xi_1^*, \ldots, \xi_K^*), \ \theta_k^* = (\xi_k^*, \gamma^T)^T, \ \theta^* = (\xi^*, \gamma^T)^T$$

be the true values of

$$\xi = (\xi_1, \ldots, \xi_K)^T, \ \theta_k = (\xi_k, \gamma^T)^T, \ \theta = (\xi^*, \gamma^T)^T.$$  

We define the following parameter space.

**Assumption 1.** The true values of model (2.3) satisfy $\theta^* \in \Theta_\mu(\omega^*)$. The parameter space is defined by

$$\Theta_\mu(\omega^*) = \left\{ \theta \in \mathbb{R}^{p+q+K} : |\xi_k^*| \leq \frac{1}{\mu^k}, k = 1, \ldots, K; \ \omega^* \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j \leq 1 - \mu, \ \alpha_i, \beta_j \geq 0, \ \forall i, j \right\},$$

(2.6)

where $\mu \in (0, 1)$ is a real number such that $\theta^* \in \Theta_\mu(\omega^*)$ and $\omega^* = \text{Var}(\eta_t)$.

**Remark 1.** Assumption 1 implies that the parameter space is compact. It seems to be stronger than Eq. (2.4). However, for any values satisfying Eq. (2.4), we can always choose the constant $\mu$ small enough such that $|\xi_k^*| \leq \frac{1}{\mu^k}$ and $\omega^* \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j \leq 1 - \mu$.

Suppose that $R_1(u), \ldots, R_n(u)$ are the observations of intraday log-return process. Because the volatility processes $\{\eta_t\}_{t=1}^{n}$ are unobservable, it is hard to estimate the parameter in GARCH proxy model directly. Xiao and Koenker (2009) proposed a two-step method to deal with this situation. Under certain regularity conditions, $v_t^2$ can be represented as an ARCH(∞):
\[ \nu_t^2 = 1 + \sum_{i=1}^{\infty} c_i r_{t-i}^2 \approx 1 + \sum_{i=1}^{m} c_i r_{t-i}^2, \]  
\tag{2.7} 

where the last term is a linear approximation with truncated time \( t - m \). By choosing a suitable truncation parameter and substituting this approximation into GARCH model (2.1), we obtain

\[ r_t \approx \left( 1 + \sum_{i=1}^{m} c_i r_{t-i}^2 \right)^{1/2} \eta_t. \]  
\tag{2.8} 

Suppose \( \{c_i\} \) is a consistent estimator of \( \{c_i\} \) from approximation model (2.8). Substituting the estimators into Eq. (2.7), we have

\[ \hat{\nu}_t^2 = 1 + \sum_{i=1}^{m} \hat{c}_i r_{t-i}^2. \]  
\tag{2.9} 

However, the truncation of ARCH(\( \infty \)) still brings unpredictable bias to estimators. This is the main reason why the method in Xiao and Koenker (2009) does not always perform well in practice, especially when the sample size is small.

In this paper, we adopt another procedure to deal with the unobservable volatilities. According to Lee and Noh (2011), we replace \( \nu_t^2 \) by

\[ \hat{\nu}_t^2(\gamma) = 1 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_j \hat{\nu}_{t-j}(\gamma). \]  
\tag{2.10} 

where the initial values

\[ r_0 = \cdots = r_{1-q} = r_1, \quad \hat{\nu}_0^2(\gamma) = \cdots = \hat{\nu}_{1-p}^2(\gamma) = r_1. \]

Let \( \hat{\nu}_i(\theta_k) = \hat{\nu}_i(\gamma) \hat{\varepsilon}_k \), where \( \gamma = (\alpha^T, \beta^T)^T \). Now we define the CQR estimator of GARCH proxies model with \( K \) quantiles by

\[ \hat{\theta}_n(\omega^*) = \arg\min_{\theta \in \Theta_\omega(\omega^*)} \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \rho_{\omega_k}(H_i - \hat{\nu}_i(\theta_k)) \]  
\tag{2.11} 

where \( \rho_{\omega_k}(y) = y \tau_k - I(y < 0) \) is the quantile function with fixed quantile \( \tau_k \). Specially when \( K = 1 \) and \( \tau_k = \tau \), the composite quantile regression degenerates to the quantile regression (QR) with \( \tau \). So we treat QR as a special case of CQR, and all conclusions on CQR can be established for QR analogously.

### 3. Asymptotic properties of composite quantile regression

In this section, we investigate the consistency and asymptotic normality of the newly proposed estimator. We also provide the asymptotic relative efficiency (ARE) among the composite quantile regressions based on different proxies. Denote \( A(x) = \sum_{i=1}^{q} \alpha_i x_i \) and \( B(x) = 1 - \sum_{j=1}^{p} \beta_j x_j \). We impose the following regularity assumptions.

**Assumption 2.** For \( p > 0, q > 0 \), \( \alpha_i^2 \) and \( \beta_j^2 \) are both away from zero. The polynomial \( A(x) \) and \( B(x) \) do not have common factors.

**Assumption 3.** Let \( F_\varepsilon \) and \( f_\varepsilon \) be the cumulative distribution and density function of \( \varepsilon_t \) respectively, \( f_\varepsilon \) is differentiable and satisfies \( \sup_{x} |f_\varepsilon(x)| \leq C_1, \sup_{x} \left| f'_\varepsilon(x) \right| \leq C_2 \) for some positive constants \( C_1 \) and \( C_2 \). \( f_\varepsilon(f_\varepsilon^{-1}(\tau_k)) > 0 \) for \( 1 \leq k \leq K \). And \( E(\varepsilon_t^2) < \infty \).

Assumption 2 is for the identifiability of GARCH model. Assumption 3 is a common regularity condition in the literature on quantile regression. The following theorem establishes the strong consistency of CQR estimator \( \hat{\theta}_n(\omega^*) \).

**Theorem 1.** Suppose that Assumptions 1–3 hold, \( \hat{\theta}_n(\omega^*) \) is strongly consistent in the sense that

\[ \hat{\theta}_n(\omega^*) - \theta^* \rightarrow 0, \text{ a.s., as } n \rightarrow \infty, \]

where “a.s.” is short for “almost surely” convergence. To establish the asymptotic normality of \( \hat{\theta}_n(\omega^*) \), we impose a strengthened version of Assumption 1.

**Assumption 1’.** \( \theta^* \) is an interior point of \( \Theta_\omega(\omega^*) \).

Assumptions 1 and 1’ make the roots of \( A(x) \) and \( B(x) \) outside the unit circle on the complex plane. Therefore, they both ensure that the GARCH model has a unique strictly stationary and ergodic solution denoted by \( \{\nu_t(\gamma)\} \) satisfying

\[ \nu_t^2(\gamma) = 1 + \sum_{i=1}^{q} \alpha_i r_{t-i}^2 + \sum_{j=1}^{p} \beta_j \nu_{t-j}^2(\gamma). \]
The following theorem states the asymptotic normality of CQR estimator $\hat{\theta}_n(\omega^*)$.

**Theorem 2.** If Assumptions 1', 2 and 3 hold, it follows that

$$\sqrt{n} \left( \hat{\theta}_n(\omega^*) - \theta^* \right) \xrightarrow{d} N(0, B^{-1} A B^{-1})$$

as $n \to \infty$.

where

$$A = \sum_{k=1}^{K} \sum_{k'=1}^{K} \left( \tau_k \wedge \tau_0 - \tau_k \tau_0 \right) \mathbb{E} \left( \frac{\partial q_i(\theta^*_0)}{\partial \theta} \frac{\partial q_i(\theta^*_0)}{\partial \theta^T} \right) = \sum_{k=1}^{K} A_{nk} \tau_k,$$

$$B = \sum_{k=1}^{K} f(\xi_k^e) \mathbb{E} \left( \frac{1}{\nu_1(\gamma^*)} \frac{\partial q_i(\theta^*_0)}{\partial \theta} \frac{\partial q_i(\theta^*_0)}{\partial \theta^T} \right) = \sum_{k=1}^{K} B_{nk},$$

with $q_i(\theta^*_0) = v_i(\gamma^*) \xi_i^e$. The notation $\wedge$ stands for getting the smaller one between the two.

This conclusion still holds for quantile regression while the asymptotic variance is much simpler.

**Corollary 1.** Suppose $\hat{\theta}_n(\omega^*) = \arg\min_{\theta \in \mathbb{R}^m} \sum_{i=1}^{n} \rho_i(\theta)$ is the quantile regression estimator of model (2.3). Suppose Assumptions 1–3 hold, it follows $\hat{\theta}_n(\omega^*) \to \theta^*$ a.s., as $n \to \infty$. Furthermore, when Assumption 1' is satisfied,

$$\sqrt{n} \left( \hat{\theta}_n(\omega^*) - \theta^* \right) \xrightarrow{d} N(0, B^{-1} A B^{-1})$$

as $n \to \infty$.

where

$$A_{tt} = \tau(1 - \tau) \mathbb{E} \left( \frac{\partial q_i(\theta^*)}{\partial \theta} \frac{\partial q_i(\theta^*)}{\partial \theta^T} \right)$$

and

$$B_{t} = f(\xi^e) \mathbb{E} \left( \frac{1}{\nu_1(\gamma^*)} \frac{\partial q_i(\theta^*)}{\partial \theta} \frac{\partial q_i(\theta^*)}{\partial \theta^T} \right),$$

with $q_i(\theta^*) = v_i(\gamma^*) \xi_i^e$.

**Remark 2.** The quantile regression requires $F^{-1}(\tau) \neq 0$. The proxies are positive random variables by definition. We are able to choose an appropriate quantile $\tau$ satisfying that $F^{-1}(\tau)$ is not close to 0.

**Remark 3.** The matrix $A$ and $B$ are both positive definite. The proof is provided in the appendix.

**Remark 4.** In Theorems 1, 2 and Corollary 1, our arguments are on the parameter space $\Theta_{\mu}(\omega^*)$. Let $\hat{\omega}_n$ be a strong consistent estimator for $\omega^*$. It can be proved that CQR based on $\Theta_{\mu}(\hat{\omega}_n)$ has the same properties as CQR based on $\Theta_{\mu}(\omega^*)$ asymptotically does. Therefore, in practice, we replace $\Theta_{\mu}(\omega^*)$ by $\Theta_{\mu}(\hat{\omega}_n)$ as the parameter space.

Denote by $V$ the $(p + q) \times (p + q)$ lower diagonal blocked submatrix of $B^{-1} A B^{-1}$. From Theorem 2, we have

$$\sqrt{n} \left( \hat{\gamma}_n(\omega^*) - \gamma^* \right) \xrightarrow{d} N(0, V),$$

where $\gamma^* = (\omega^T \tau, \beta^T \tau)^T$ are the true values of GARCH($p$, $q$) parameters. The following lemma proves the expression of $V$, which will be applied to establish the asymptotic relative efficiency of CQR based on different proxies and quantiles.

**Lemma 1.** If Assumptions 1', 2 and 3 hold, it follows that

$$V = \frac{\sum_{k=1}^{K} \sum_{j=1}^{K} (\tau_i \wedge \tau_j - \tau_i \tau_j) \xi_i^e \xi_j^e V^{-1} V^{-1}}{(\sum_{k=1}^{K} f(\xi_k^e) \xi_k^e)^2},$$

where $V_0$ and $V_1$ are independent of the proxy $H$ and quantiles $\{\tau_k\}$.

Let $H_i^e = \nu_i \epsilon_i^e$ be another proxy and $(\xi_1^e, \ldots, \xi_K^e)$ be another $K$ quantiles $(\tau_1^e, \ldots, \tau_K^e)$ of $\epsilon_i^e$. The density function of $\epsilon_i^e$ is $f_{\epsilon_i^e}$. We show the following results.

**Corollary 2.** Suppose Assumptions 1', 2 and 3 hold. Two CQR estimators respectively based on $H(\tau_1, \ldots, \tau_K)$ and $H'(\tau_1^e, \ldots, \tau_K^e)$. The two estimators have asymptotic relative efficiency:

$$\text{ARE} (H, H') = \frac{\sum_{k=1}^{K} \sum_{j=1}^{K} (\tau_i^e \wedge \tau_j^e - \tau_i^e \tau_j^e) \xi_i^e \xi_j^e}{\sum_{k=1}^{K} \sum_{j=1}^{K} (\tau_i \wedge \tau_j - \tau_i \tau_j) \xi_i^e \xi_j^e} \left( \sum_{k=1}^{K} f_{\epsilon_i^e}(\xi_k^e)^2 \right)^2.$$

Specially, when $H$ and $H'$ have the same single quantile $\tau$, the ARE is:

$$\text{ARE} (H, H') = \frac{\int_{\xi_1^e}^{\xi_K^e} \xi_i^e d\xi_i^e}{\int_{\xi_1^e}^{\xi_K^e} \xi_i^e d\xi_i^e}.$$

Corollary 2 implies that choosing appropriate proxies can improve the efficiency of estimators. The value of $f_{\epsilon_i^e}(\xi_i^e)^2$ is determined by the proxy innovations. From Corollary 2, we see that the efficiency of CQR with proxy $H$ depends on two factors: quantiles and density function of the proxy innovations. So we tend to choose quantiles with larger $f_{\epsilon_i^e}(\xi_i^e)^2$ to get a more efficient CQR estimator for $\gamma$. We further discuss it in the numerical studies.

Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.ecosta.2016.11.004
4. Numerical studies

4.1. Monte Carlo simulation

In this section, we conduct Monte Carlo simulations to examine the finite sample performance of our proposed method. To illustrate the robustness and efficiency, we compare the proposed method with MLE and QR method. We further study the effect of the number of quantiles used in CQR method.

To simulate $r_t$ and $R_t(u)$ processes, we first simulate the intraday diffusion process $\Psi(t)$. The logarithm of the diffusion coefficients in $\Psi(t)$ is simulated by the Ornstein–Uhlenbeck process. The two processes are defined by:

$$
\begin{align*}
    d\Psi_t(u) &= \exp\{X_t(u)\} dB_t^{(1)}(u), \quad u \in [0, 1], \\
    dX_t(u) &= (\theta_1 - \theta_2 X_t(u)) du + \theta_3 dB_t^{(2)}(u),
\end{align*}
$$

where $B_t^{(1)}$ and $B_t^{(2)}$ are independent standard Brownian motions. Since the Ornstein–Uhlenbeck process $X_t$ is stationary if and only if $\theta_2$ is positive, we set

$$
\theta_1 = -\frac{1}{32}, \quad \theta_2 = \frac{1}{2}, \quad \theta_3 = \frac{1}{4}.
$$

$X(0)$ is randomly sampled from its stationary process. Without loss of generality, we assume the overnight return is 0, that is $\Psi(0) = 0$. We approximate the Eq. (4.1) by Riemann’s summation with 480 grid points equally spaced on $[0, 1]$.

As Visser (2011) mentioned, both the quadratic variation over the unit interval and the realized variance $RQV(\Psi_t)$ have expectation 1, when $2\theta_1 + \theta_2^2 = 0$. Therefore, let $\eta_t = \Psi_t(1)$, where $\eta_t$ is the innovation process in the GARCH model, and then $\omega^* \equiv \text{Var}(\eta_t) = 1$. $\{\eta_t\}$ and $\{t_t\}$ follow the stationary GARCH(1, 1) model,

\begin{align*}
    r_t &= v_t \eta_t, \\
    v_t^2 &= 1 + 0.3r_{t-1}^2 + 0.6v_{t-1}^2.
\end{align*}

The order selection of GARCH model is still an open question in the literature. And it is beyond the scope of this paper. Thus, we suppose the GARCH order $(p, q) = (1, 1)$ is fixed in the whole numerical studies.

Daily log-return process satisfies $R_t(1) = r_t$ and $R_t(u) = v_t \Psi_t(u)$ for $u \in [0, 1]$. 48 intraday log-returns for each day are generated for 1000-days realization, because China stock market opens 4 h everyday, and the prices are recorded every five minutes. The daily log-returns are denoted as $\{R_t(u_t)\}_{t=1}^{48}$, with $u_t = \frac{t}{4}$ and $t = 1, \ldots, 1000$.

We first investigate the efficiencies of the CQR estimators with different volatility proxies. The realized volatility, defined by

$$
H_t^{(1)} = \sqrt{RQV_t} = \left(\sum_{i=1}^{48} [R_t(u_i) - R_t(u_{i-1})]^2\right)^{1/2},
$$

is the most popular one to characterize the volatility. Another commonly used proxy is the high-low range defined by $H_t^{(2)} = \max_i R_t(u) - \min_i R_t(u)$. A third rarely used proxy is $|r_t|$, since $|r_t|$ is positive and homogeneous. $r_t$ is symmetrically distributed under our setting, so the regression on $|r_t|$ is equivalent that on $r_t$. Based on aforementioned volatility proxies, we compare the finite sample performance of the following three methods:

1. maximum likelihood estimate (MLE);
2. quantile regression with $\tau$, denoted by QR($\tau$);
3. composite quantile regression with $k$ quantiles, denoted by CQR$_k$.

Among the three estimation methods, QR, CQR based on proxies are newly proposed in this paper. Each of the three methods will be applied to both the original process $r_t$ and the processes incorporating 3 proxies. MLE based on $r_t$ is widely used to estimate GARCH parameters. MLE’s based on proxies $H_t^{(1)}$ and $H_t^{(2)}$ are discussed by Visser (2011), which demonstrates that the efficiency is greatly improved by using proxies. Quantile regression on $r_t$ is discussed by Xiao and Koenker (2009), and Lee and Noh (2011).

We assess the efficiency by the root mean squared error (RMSE) as presented in Table 1. Table 1 lists the RMSE’s of MLE, QR and CQR based on different proxies for parameter $(\alpha, \beta) = (0.3, 0.6)$ in the GARCH model with three sample sizes: $T = 500, 750$ and 1000, respectively, over 1000 replications of the sample path.

The table is separated into three panels. The first three rows in the top panel list RMSE of MLE, QR and CQR using daily close-to-close data. For the results in these three rows, we choose $\tau = 0.1$ for QR and $\tau = 0.1, 0.3, 0.7, 0.9$ for CQR. The subscript number in CQR stands for how many quantiles are adopted in the composite quantile regression. The 0.5 quantile was not employed in this case, since the median of $r_t$ is 0. To compare the efficiency gained by proxies, the bottom two panels present the RMSE’s based on the high-low range and the realized volatility. In these two panels, the CQR employs $\tau = 0.2, 0.8$. CQR employs $\tau = 0.2, 0.4, 0.6, 0.8$, and CQR employs $\tau = 0.2, 0.3, \ldots, 0.8$. The number of quantiles is increased gradually in order to explore the impact of the number of quantiles on CQR.

For the results in Table 1, we will start with the discussion of three main findings to demonstrate the advantage of our newly proposed CQR method. First, all methods based on proxies produce smaller RMSE’s than those based on $r_t$, and
Table 1
RMSE of MLE, QR and CQR based on \( r_t \), high-low range and realized volatility. CQR\(_4\) for \( r_t \) employs quantiles \( \tau = 0.1, 0.3, 0.7, 0.9 \). For high-low range and realized volatility, CQR\(_8\) employs \( \tau = 0.2, 0.8 \). CQR\(_4\) employs \( \tau = 0.2, 0.4, 0.6, 0.8 \), and CQR\(_7\) employs \( \tau = 0.2, 0.3, \ldots, 0.8 \).

<table>
<thead>
<tr>
<th>Proxy</th>
<th>Estimator</th>
<th>( T = 500 )</th>
<th>( T = 750 )</th>
<th>( T = 1000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha )</td>
<td>( \beta )</td>
<td>( \alpha )</td>
<td>( \beta )</td>
</tr>
<tr>
<td>( r_t )</td>
<td>MLE</td>
<td>0.0830</td>
<td>0.0868</td>
<td>0.0704</td>
</tr>
<tr>
<td></td>
<td>QR(0.1)</td>
<td>0.1256</td>
<td>0.1689</td>
<td>0.1153</td>
</tr>
<tr>
<td></td>
<td>CQR(_4)</td>
<td>0.0971</td>
<td>0.1063</td>
<td>0.0819</td>
</tr>
<tr>
<td>High-low range</td>
<td>MLE</td>
<td>0.0575</td>
<td>0.0482</td>
<td>0.0466</td>
</tr>
<tr>
<td></td>
<td>QR(0.2)</td>
<td>0.0663</td>
<td>0.0574</td>
<td>0.0564</td>
</tr>
<tr>
<td></td>
<td>QR(0.5)</td>
<td>0.0610</td>
<td>0.0523</td>
<td>0.0519</td>
</tr>
<tr>
<td></td>
<td>QR(0.8)</td>
<td>0.0683</td>
<td>0.0593</td>
<td>0.0570</td>
</tr>
<tr>
<td></td>
<td>CQR(_8)</td>
<td>0.0583</td>
<td>0.0481</td>
<td>0.0483</td>
</tr>
<tr>
<td></td>
<td>CQR(_4)</td>
<td>0.0549</td>
<td>0.0446</td>
<td>0.0452</td>
</tr>
<tr>
<td></td>
<td>CQR(_7)</td>
<td>0.0545</td>
<td>0.0443</td>
<td>0.0452</td>
</tr>
<tr>
<td>Realized volatility</td>
<td>MLE</td>
<td>0.0381</td>
<td>0.0346</td>
<td>0.0296</td>
</tr>
<tr>
<td></td>
<td>QR(0.2)</td>
<td>0.0360</td>
<td>0.0357</td>
<td>0.0281</td>
</tr>
<tr>
<td></td>
<td>QR(0.5)</td>
<td>0.0302</td>
<td>0.0215</td>
<td>0.0242</td>
</tr>
<tr>
<td></td>
<td>QR(0.8)</td>
<td>0.0347</td>
<td>0.0249</td>
<td>0.0287</td>
</tr>
<tr>
<td></td>
<td>CQR(_2)</td>
<td>0.0287</td>
<td>0.0193</td>
<td>0.0231</td>
</tr>
<tr>
<td></td>
<td>CQR(_4)</td>
<td>0.0264</td>
<td>0.0182</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>CQR(_7)</td>
<td>0.0262</td>
<td>0.0181</td>
<td>0.0207</td>
</tr>
</tbody>
</table>

CQR method has the smallest RMSE. This improvement of RMSE can be observed by comparing the results in the top panel and the ones in the bottom two panels. Second, almost within each panel, the RMSE’s from CQR methods are significantly smaller than those from MLE and QR methods except in the top panel. In the top panel of Table 1, MLE is the most efficient method among the three. This is expected since the \( r_t \) from the ordinary GARCH model is adopted in the estimation without intraday information which provides essential advantage for the other two methods. When taking the intraday information into account, however, even the efficiency of MLE based on the other proxies is greatly improved. Therefore, we should focus on the comparison in the bottom two panels and clearly the newly proposed CQR method always produces much smaller RMSE’s when applied with two proxies. Third, between the two proxies, the realized volatility seems to be a better choice, since all methods based on realized volatility improve the RMSE’s more than based on high-low range do. Because the realized volatility can accurately characterize the volatility, more efficiency is gained by adopting realized volatility than by high-low range. As we mentioned in the introduction, however, the realized volatility brings more cumulative errors and outliers; the innovations of proxies depart from the Gaussian assumption which is the core assumption for MLE. Thus, compared with MLE, the improvement of CQR and QR methods based on realized volatility proxy is much more than that based on high-low range.

Even both CQR and QR based on proxies are better than MLE method, the CQR method is more efficient than QR method. The better efficiency of CQR method is due to the fact that CQR is less impacted by the choice of quantiles. From the table, we notice that the choice of quantile highly impacts the efficiency of QR. We specially choose the extreme quantile \( \tau = 0.1 \) to illustrate this impact. This impact can be explained by Corollary 2. Corollary 2 shows that the asymptotic variance of \( QR(\tau) \) is determined by \( \frac{g(\tau)}{g(\tau)^2} \). Asymptotic variance will be bigger if \( g(\tau) \) is smaller. Since \( g(0.2) \) and \( g(0.8) \) are relatively smaller than \( g(0.5) \), \( QR(0.5) \) outperforms \( QR(0.2) \) and \( QR(0.8) \). Therefore, the efficiency of QR is essentially determined by the choice of quantile \( \tau \). But which \( \tau \) gives better \( QR(\tau) \) is hard to know when we have little information about the quantiles of log-return process. Compared with QR, composite quantile regression uses more information at different quantiles. As more quantiles are employed in CQR, its efficiency grows, but the growth decelerates, gradually. The deceleration of efficiency is shown in the table as the difference between CQR\(_4\) and CQR\(_7\) is almost negligible. While little efficiency is gained by CQR\(_7\), the CQR\(_7\) requires much more computing time than CQR\(_4\) from our simulation experience.

All the aforementioned advantages of CQR methods still hold when we increase the sample size \( T \). As the sample size increases, the RMSE’s of three methods decrease uniformly. This improvement along sample size is expected as all methods are consistent.

Except Table 1, we also visualize the difference among methods by plotting the estimators in Fig. 1. 1000 estimators by each method are plotted with sample size 1000 and parameters \( (\alpha, \beta) = (0.3, 0.6) \). The parameters are chosen according to Assumption 1 that the parameter space is \( \Theta = \{ (\alpha, \beta) : \alpha + \beta \leq 0.999 \} \). Therefore, the parameter space \( \Theta \) retains these estimators below the straight line \( \{ (\alpha, \beta) : \alpha + \beta = 0.999 \} \). In each plot, parameter \( \alpha \) are indicated on the x coordinate and parameter \( \beta \) on the y coordinate. The left three plots present the MLE based on three proxies. The right three plots present the CQR\(_4\) based on three proxies. Subplot (e) and (f) show the best performance of MLE and CQR estimators using realized volatility as proxy. In subplot (e), however, many estimators by MLE are far from the true value. This deviation is expected since the realized volatility leads cumulative errors and outliers. The cumulations of errors violates the underlying assumptions for MLE. Compared to MLE, the CQR method can eliminate the side-effect of the realized volatility proxy. Therefore, the CQR estimators based on realized volatility are almost concentrated on the true value \( (0.3, 0.6) \).

Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.ecosta.2016.11.004
Next we examine the robustness of CQR methods. We impose random perturbation to \( \{r_t\} \) and \( \{H_t\} \). Define the contaminated \( \{r'_t\} \) and \( \{H'_t\} \) by:

\[
\begin{align*}
    r'_t & = \begin{cases} 
    r_t + z_t & \text{if } 1 \leq t \leq \lfloor \Delta T \rfloor, \\
    r_t & \text{elsewhere};
    \end{cases} \\
    H'_t & = \begin{cases} 
    H_t + z_t & \text{if } 1 \leq t \leq \lfloor \Delta T \rfloor, \\
    H_t & \text{elsewhere};
    \end{cases}
\end{align*}
\]

Fig. 1. Scatters of estimators of MLE and CQR of \((\alpha, \beta) = (0.3, 0.6)\) based on \(r_t\), high-low range and realized volatility.

Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.ecosta.2016.11.004
Table 2
RMSE of MLE, QR and CQR based on high-low range and the realized volatility, when contaminated data appear. CQR\(_2\) employs \(\tau = 0.2\), 0.8. CQR\(_4\) employs \(\tau = 0.2, 0.4, 0.6, 0.8\), and CQR\(_7\) employs \(\tau = 0.2, 0.3, \ldots, 0.8\).

<table>
<thead>
<tr>
<th>Proxy</th>
<th>Method</th>
<th>0%</th>
<th>5%</th>
<th>10%</th>
<th>15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-low range</td>
<td>MLE</td>
<td>0.0414</td>
<td>0.0310</td>
<td>0.0529</td>
<td>0.0386</td>
</tr>
<tr>
<td></td>
<td>QR(0.2)</td>
<td>0.0475</td>
<td>0.0370</td>
<td>0.0505</td>
<td>0.0391</td>
</tr>
<tr>
<td></td>
<td>QR(0.5)</td>
<td>0.0456</td>
<td>0.0341</td>
<td>0.0471</td>
<td>0.0351</td>
</tr>
<tr>
<td></td>
<td>QR(0.8)</td>
<td>0.0514</td>
<td>0.0403</td>
<td>0.0548</td>
<td>0.0410</td>
</tr>
<tr>
<td></td>
<td>CQR(_2)</td>
<td>0.0424</td>
<td>0.0313</td>
<td>0.0463</td>
<td>0.0336</td>
</tr>
<tr>
<td></td>
<td>CQR(_4)</td>
<td>0.0397</td>
<td>0.0287</td>
<td>0.0432</td>
<td>0.0313</td>
</tr>
<tr>
<td></td>
<td>CQR(_7)</td>
<td>0.0397</td>
<td>0.0285</td>
<td>0.0428</td>
<td>0.0309</td>
</tr>
<tr>
<td>Realized volatility</td>
<td>MLE</td>
<td>0.0267</td>
<td>0.0231</td>
<td>0.0425</td>
<td>0.0340</td>
</tr>
<tr>
<td></td>
<td>QR(0.2)</td>
<td>0.0249</td>
<td>0.0177</td>
<td>0.0345</td>
<td>0.0201</td>
</tr>
<tr>
<td></td>
<td>QR(0.5)</td>
<td>0.0224</td>
<td>0.0153</td>
<td>0.0258</td>
<td>0.0169</td>
</tr>
<tr>
<td></td>
<td>QR(0.8)</td>
<td>0.0256</td>
<td>0.0184</td>
<td>0.0294</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>CQR(_2)</td>
<td>0.0204</td>
<td>0.0147</td>
<td>0.0263</td>
<td>0.0171</td>
</tr>
<tr>
<td></td>
<td>CQR(_4)</td>
<td>0.0189</td>
<td>0.0153</td>
<td>0.0243</td>
<td>0.0153</td>
</tr>
<tr>
<td></td>
<td>CQR(_7)</td>
<td>0.0189</td>
<td>0.0152</td>
<td>0.0240</td>
<td>0.0151</td>
</tr>
</tbody>
</table>

where \(z_t\) is independent and identically distributed \(N(0, \text{std}^2_1)\) random variable and \(z_t'\) is independent and identically distributed absolute value of \(N(0, \text{std}^2_2)\) random variable. std\(_1\) is the standard deviation of \(r_t\) and std\(_2\) is the standard deviation of \(H_t\). \(\Delta\) % stands for the percentage of contaminated data. Based on the contaminated data, the estimation results are listed in Table 2. Table 2 presents the RMSE of MLE, QR and CQR based on high-low range and the realized volatility over 1000 days. The percentages of contamination are 0%, 5%, 10% and 15%, respectively.

As more data are contaminated, the RMSE of MLE increases. Especially, when \(\Delta\) % changes from 0% to 5%, the RMSE of MLE increases sharply. The increase of RMSE is due to more outliers brought by contamination. Although the RMSE of CQR also increases along the percentage of contamination, the RMSE’s of CQR with the highest contamination percentage are even smaller than the RMSE’s of other methods with lowest contamination percentage. Therefore, CQR methods based on two proxies perform more stably and are more efficient.

In summary, through the Monte Carlo simulation studies, we conclude that using the realized volatility proxies, the CQR method can improve the efficiency and robustness of the parameter estimation of high-frequency GARCH model. The CQR method overcomes the non-normality, outliers and cumulative errors brought by the realized volatility proxy. Employing more quantiles, CQR can obtain higher efficiency but consumes more computation time. And the efficiency gain is not dramatically significant as indicated by the difference of RMSE between CQR\(_4\) and CQR\(_7\). Hence CQR\(_4\) is a good choice when considering both estimate accuracy and computation burden.

4.2. Empirical analysis

In the empirical analysis, we will apply our newly proposed method to Value at Risk management. Value at Risk (VaR) refers to the maximal loss of a portfolio for a given confidence level. Since VaR is essentially the 95th conditional quantile of a financial sequence, quantile regression is naturally considered as a useful method in VaR analysis. In this section, VaR is studied with the high-frequency log-returns of HuShen 300 Index of China stock market from April 13, 2006 to April 10, 2014. During this period, there are totally 1940 trading days. 48 × 1940 observations are evenly recorded. Observations are denoted by \(p_t(u)\) and the high-frequency log-returns at time \(u\) of day \(t\) by \(R_t(u) = 100 \cdot \log(p_t(u)/p_{t-1}(1))\). \(r_t = R_t(1)\) is the daily close-to-close log-return. According to the existing literature, \(r_t\) is assumed to follow GARCH(1,1) model and \(R_t(u)\) follows model (2.3). For these models, we use the data of the first 1000 trading days as training samples to obtain the estimator \((\hat{\alpha}, \hat{\beta})\). Naturally, we estimate \(\hat{\tau}_t(\gamma)\) by \(\hat{\tau}_t(\gamma)\) in (2.10) and estimate \(\xi_k = F_{\gamma}^{-1}(\tau_k)\) by sample \(\tau_k\) quantile of sequence \(\{\tau_t/\hat{\tau}_t(\gamma)\} : 1 \leq t \leq 1000\). We use the remaining data as testing samples to predict VaR. VaR \(q_t\) at time \(t\) with prefixed quantile \(\tau\) is defined by:

\[
q_t = q_t(\tau) = \inf\{x : \tau_{t-1}(x) \geq \tau\},
\]

where \(\tau_{t-1}\) is the conditional probability given \(\sigma\)-filtration \(\mathcal{F}_{t-1}\).

We forecast the one-step-ahead VaR at the coverage probability \(\tau = 10\%\) by

\[
\hat{q}_t = F^{-1}_\gamma((1 + \hat{\alpha})_{\tau_{t-1}}^2 + \hat{\beta}\bar{\tau}_{t-1}(\hat{\alpha}, \hat{\beta}))^{1/2}; \quad 1001 \leq t \leq 1940.
\]

In this forecast procedure, we adopt realized volatility as the proxy and apply MLE, LADE and CQR with \(\tau = 0.2, 0.4, 0.6, 0.8\) based on the realized volatility proxy. We evaluate the forecast according to two tests. One test is the unconditional likelihood ratio test proposed by Kupiec (1995): let \(n = 940\), \(n' = \sum_{t=1001}^{t=1940} I(r_t \leq \hat{q}_t)\), and \(p' = n'/n\).

\[
\text{LR}_{\text{uc}} = 2(n - n') \ln \left( \frac{1 - p'}{1 - p} \right) + 2n' \ln \left( \frac{p'}{p} \right).
\]

Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.econsta.2016.11.004
Table 3
Kupiec’s test and Christoffersen’s test for MLE, LAD and CQR4 with \( \tau = (0.2, 0.4, 0.6, 0.8) \), based on \( H_0 = \sqrt{RQV} \).

<table>
<thead>
<tr>
<th>p-value</th>
<th>CQR4</th>
<th>Distribution of statistic</th>
<th>Rejection region under 5% level</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR_{cc}</td>
<td>Statistic</td>
<td>MLE</td>
<td>LAD</td>
</tr>
<tr>
<td>0.001</td>
<td>0.007</td>
<td>0.074</td>
<td>( \chi^2_{(1)} )</td>
</tr>
<tr>
<td>LR_{cc}</td>
<td>Statistic</td>
<td>13.78</td>
<td>7.91</td>
</tr>
</tbody>
</table>

Kupiec (1995) used the unconditional likelihood ratio test to examine whether the occurrence frequency of violations equals the corresponding probability, and proved that it asymptotically follows \( \chi^2_{(1)} \). Another test was proposed by Christoffersen (1998), which tests the independence of violations. Let \( t_i = t(r_i \leq \hat{q}_i) \) for 1001 \( \leq t \leq 1940 \) be the indicator of a violation, and \( n_{ij} \) \( (i, j = 0, 1) \) be the number of days satisfying \( \{ t : k_{i-1} = i, h_j = j, 1002 \leq t \leq 1940 \} \). Let \( p_{ij} = n_{ij}/(n_{i0} + n_{i1}) \). Christoffersen’s test statistic is

\[
LR_{cc} = 2 \left[ \ln(\hat{p}_0^{p}\hat{p}_1^{p} \hat{p}_0^{n}\hat{p}_1^{n}) - \ln((1 - \hat{p})^{n_{i0}} \hat{p}^{n_{i1}}) \right],
\]

which is asymptotically \( \chi^2_{(2)} \) distributed. Christoffersen’s test will not reject the null hypothesis when the occurrences of violations would be spread out over the sample and not come in cluster.

Table 3 provides the testing results and the corresponding \( p \)-values of MLE, LAD and CQR4 based on \( H_0 = \sqrt{RQV} \). For MLE and LAD, both Kupiec’s test and Christoffersen’s test are significant under level 5%, which means the VaR forecasts by MLE and LAD are not accurate. On the contrary, both tests of VaR prediction by CQR4 are not significant at 5% significant level. Therefore, CQR provides a better VaR forecast method.

The empirical analysis advocates the newly proposed methodology and fully supports the conclusion that CQR estimation based on suitable proxies are highly efficient and robust for high-frequency GARCH model.

5. Conclusion

In this paper, we presented new methods, CQR and QR based on proxies, to estimate the parameters in high-frequency GARCH model. With the newly proposed methods, we take advantage of the intraday high frequency data through proxies. Proxies can be constructed in many ways according to Definition 1. With proxies, CQR estimation of the parameters in GARCH model is both consistent and asymptotically normal as indicated in Theorems 1 and 2. Not only do the CQR methods have nice asymptotic properties, but they also perform better in terms of accuracy which is measured by RMSE. The RMSE’s of CQR methods with proxies are uniformly smaller than those of MLE method with proxies as observed from simulation studies. While comparing the RMSE’s of methods with or without proxies, RMSE’s are always smaller for methods applied with proxies. This is also true for the conventional MLE estimators, even though the improvement of RMSE is more dramatic with CQR method applied with proxies. The improvement of RMSE’s implies the higher efficiency of CQR methods. Besides the higher efficiency, the CQR method is also more robust. The robustness is reflected through the study of contaminated data. With the contamination level increases, the CQR methods are still more efficient than other methods. And the efficiency of CQR method is even higher than that of other methods at lower noise level. Both efficiency and robustness can be achieved by the new CQR method applied with proxies. Among the proxies we studied, the realized volatility proxy is preferred due to the smallest RMSE achieved with all estimation methods. Hence, CQR method with realized volatility proxy is applied to an empirical study of Value at Risk forecast. The empirical results are also in favor of our newly proposed CQR methods based on proxies because of the more accurate forecast. Therefore, as long as the intraday high frequency returns are available, our newly proposed CQR method should be considered over other methods when estimating the parameters in stochastic volatility model for higher efficiency and robustness.

Appendix

In this section, we prove the asymptotic properties of CQR estimate. Denote the Euclidean norm of vector \( a \) by \( |a| \). We define \( \|A\| = \sum_{i,j} |a_{ij}| \) for matrix \( A = (a_{ij}) \). \( \mathcal{F}_t \) represents information up to time \( t \). \( V \) denotes a generic integrable random variable, \( \{S_t\} \) denotes a generic nonnegative, stationary, ergodic and square-integrable stochastic process such that \( S_t \in \mathcal{F}_t \) and \( 0 < \rho < 1 \). We first provide several properties of \( v_t(\gamma) \), \( q_t(\theta_k) \), \( \tilde{v}_t(\gamma) \), \( \tilde{q}_t(\theta_k) \) and their derivatives. Lemmas 2–5 are corollaries of Lemmas A1–A4 in Lee and Noh (2011).

**Lemma 2.** Under Assumptions 1–3, for any small \( \mu \) and \( \theta \in \Theta_\mu(\omega^*) \),

\[
\left\{ \frac{\partial q_t(\theta_k)}{\partial \theta} ; t \in \mathbb{Z} \right\} \quad \text{and} \quad \left\{ \frac{\partial^2 q_t(\theta_k)}{\partial \theta \partial \theta^T} ; t \in \mathbb{Z} \right\}
\]
are strictly stationary and ergodic. There exist measurable functions $f_1$ and $f_2$ satisfying
\[
\frac{\partial q_t(\theta_k)}{\partial \theta} = f_1(\theta_k; r_{t-1}, r_{t-2}, \ldots) \quad \text{and} \quad \frac{\partial^2 q_t(\theta_k)}{\partial \theta \partial \theta^T} = f_2(\theta_k; r_{t-1}, r_{t-2}, \ldots).
\]

**Lemma 3.** Under Assumptions 1–3, for any small $\mu$ and $\theta \in \Theta_\mu(\omega^*)$,\[
\sup_k \sup_{\theta} \left| \frac{\partial q_t(\theta_k)}{\partial \theta} \right| \leq S_t \quad \text{and} \quad \sup_k \sup_{\theta} \left\| \frac{\partial^2 q_t(\theta_k)}{\partial \theta \partial \theta^T} \right\| \leq S_t.
\]
The same conclusion is valid for $\widetilde{q}_t(\theta_k)$.

**Lemma 4.** If Assumptions 1’, 2 and 3 hold,\[
E \left[ \frac{1}{n \nu_t^2(\gamma^*)} \left| \frac{\partial q_t(\theta_k)}{\partial \theta} \right|^4 \right] < \infty \quad \text{and} \quad E \left[ \frac{1}{n \nu_t^2(\gamma^*)} \left\| \frac{\partial^2 q_t(\theta_k)}{\partial \theta \partial \theta^T} \right\|^4 \right] < \infty.
\]

**Lemma 5.** Under Assumptions 1–3, for any small $\mu$ and $\theta \in \Theta_\mu(\omega^*)$,\[
\sup_k \sup_{\theta} \left| q_t(\theta_k) - \widetilde{q}_t(\theta_k) \right| \leq V_{1/2} \rho^t,
\]
\[
\sup_k \sup_{\theta} \left| \frac{\partial q_t(\theta_k)}{\partial \theta} - \frac{\partial \widetilde{q}_t(\theta_k)}{\partial \theta} \right| \leq S_t \rho^t,
\]
\[
\sup_k \sup_{\theta} \frac{1}{n \nu_t^2(\gamma)} \left| \frac{\partial q_t(\theta_k)}{\partial \theta} - \frac{\partial \widetilde{q}_t(\theta_k)}{\partial \theta} \right| \leq S_t \rho^t.
\]

**Proof of Theorem 1**

Let \[
\bar{S}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \sum_{k=1}^K \rho_{t_k} (H_t - \bar{q}_t(\theta_k))
\]
and
\[
S_n(\theta) = \frac{1}{n} \sum_{t=1}^n \sum_{k=1}^K \rho_{t_k} (H_t - q_t(\theta_k)).
\]

For simplicity, we define $d_t(\theta_k) = q_t(\theta_k) - q_t(\theta_k^*)$ and $\bar{d}_t(\theta_k) = \bar{q}_t(\theta_k) - \bar{q}_t(\theta_k^*)$. Let\[
\bar{G}_n(\theta) = \bar{S}_n(\theta) - \bar{S}_n(\theta^*) \quad \text{and} \quad G_n(\theta) = S_n(\theta) - S_n(\theta^*).
\]

Obviously we see $\overline{\theta}_n(\omega^*) = \arg\min_{\theta \in \Theta_\mu(\omega^*)} G_n(\theta)$. Following the line of Francq and Zakoia (2004), Lee and Noh (2011), following four properties are enough to establish the consistency.

First we verify that, $\lim_{n \to \infty} G_n(\theta) = G(\theta)$ for any $\theta \in \Theta_\mu(\omega^*)$, where $G(\theta) = E[G_n(\theta)]$. Since $|\rho_t(x) - \rho_t(y)| \leq |x - y|$ for any $\tau$, we have:
\[
\sup_{\theta \in \Theta_\mu(\omega^*)} \left| G_n(\theta) - \bar{G}_n(\theta) \right| \leq \frac{2K}{n} \sum_{t=1}^n \sup_{\theta \in \Theta_\mu(\omega^*)} \left| q_t(\theta_k) - \bar{q}_t(\theta_k) \right| \leq \frac{2K V_{1/2}}{n} \sum_{t=1}^n \rho^t.
\]

The last inequality is guaranteed by Lemma 5. So
\[
\lim_{n \to \infty} \sup_{\theta \in \Theta_\mu(\omega^*)} \left| G_n(\theta) - \bar{G}_n(\theta) \right| = 0.
\]

We define $g_t(\theta_k) = \sum_{\ell=1}^n \{ \rho_{t_k} (H_{t-\ell} - q_{t-\ell}(\theta_k)) - \rho_{t_k} (H_{t-\ell} - q_{t-\ell}(\theta_k^*)) \}$, then $G_n(\theta) = \frac{1}{n} \sum_{\ell=1}^n g_{t_k}(\theta)$. According to Theorem 36.4 in Billingsley (1995), $\{g_t(\theta)\}_t$ is strictly stationary and ergodic. Besides, $g_t(\theta)$ is integrable by the finite second order moment of $\tau$. According to the ergodic theorem, we have:
\[
\lim_{n \to \infty} \bar{G}_n(\theta) = \lim_{n \to \infty} G_n(\theta) = \lim_{n \to \infty} \frac{1}{n} \sum_{\ell=1}^n g_{t_k}(\theta) = G(\theta).
\]

Second, we can get $\gamma = \gamma^*$ when $\nu_t(\gamma)$ equals $\nu_t(\gamma^*)$. This is true by the Corollary 2.1 of Berkes et al. (2003). Third, we verify that $G(\theta)$ is non-negative, and $G(\theta) = 0$ if and only if $\theta = \theta^*$. This proof needs (Knight’s, 1998) identity:
\[
\rho_t(x - y) - \rho_t(x) = -y \psi_t(x) + y \int_0^{\infty} [I(x \leq y s) - I(x \leq 0)] ds.
\]
where $\rho_t(x) = x(\tau - I(x < 0))$, $\psi_t(x) = \tau - I(x < 0)$ for $x \neq 0$ and $0 < \tau < 1$. Here we define $e_{tk} = e_t - \xi^*_k$ and $H_{tk} = H_t - \xi^*_k v_t(y^*)$. Then we have:

$$g_t(\theta) = \sum_{k=1}^{K} \{ \rho_t(e_{tk}v_t(y^*) - d_t(\theta_k)) - \rho_t(e_{tk}v_t(y^*)) \}$$

$$= -\sum_{k=1}^{K} d_t(\theta_k)\psi_t(e_{tk}) + \sum_{k=1}^{K} d_t(\theta_k) \int_0^1 \left\{ I(e_{tk} \leq \frac{d_t(\theta_k)}{v_t(y^*)} s) - I(e_{tk} \leq 0) \right\} ds.$$

So

$$G(\theta) = E[E[g_t(\theta) | F_{t-1}]] = \sum_{k=1}^{K} E \left[ d_t(\theta_k) \int_0^1 \left\{ F\left( \frac{\xi_k + \frac{d_t(\theta_k)}{v_t(y^*)} s}{\sigma_k} \right) - F(\xi_k) \right\} ds \right].$$

Since $d_t(\theta_k)[F(\xi_k + \frac{d_t(\theta_k)}{v_t(y^*)} s) - F(\xi_k)]$ is non-negative, so $G(\theta)$ is. Further we see $G(\theta) = 0$ if and only if $d_t(\theta_k) = 0$ a.s., for any $1 \leq k \leq K$. Notice the innovations of proxies are positive random variables, thus $\xi_k > 0$ for any $0 < \tau_k < 1$. By the second property aforementioned, $d_t(\theta_k) = 0$ implies $\xi_k = \xi^*_k$ and $y = y^*$. Now we have proved the third property. Forth, we need prove that, for any $\theta(\neq \theta^*)$ we can find a neighbourhood $\mathcal{V}(\theta) \subseteq \Theta(\omega^*)$ satisfying

$$\liminf_{n \to \infty} \inf_{\theta \in \mathcal{V}(\theta)} G_n(\theta) > 0 \text{ a.s..}$$

Notice that $\rho_t(x)$ is a measurable function and the ergodicity can be kept under measurable transformation. We can prove this property by similar arguments of Francq and Zakoian (2004). Finally, the consistency of CQR can be got by above four properties.

**Proof of Theorem 2**

Let $\hat{H}_{tk} = H_t - \tilde{\theta}(y^*)\xi^*_k$. the objective function can be expressed as

$$\tilde{S}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \sum_{k=1}^{K} \rho_t(\hat{H}_{tk} - \tilde{d}_t(\theta_k)).$$

Define a $p + q + k$ dimensional vector $\delta = \sqrt{n}(\theta - \theta^*)$ and $p + q + 1$ dimensional vectors $\delta_k = \sqrt{n}(\theta_k - \theta_k^*)$. Let $\Theta_n = \sqrt{n}(\Theta_{\mu}(\omega^*) - \theta^*)$. For any $\delta \in \Theta_n$, we define $\tilde{Z}_n(\delta)$ as follows,

$$\tilde{Z}_n(\delta) = n\tilde{S}_n(\theta) - nS_n(\theta^*)$$

$$= n \sum_{t=1}^{n} \sum_{k=1}^{K} \{ \rho_t(\hat{H}_{tk} - \tilde{d}_t(\theta_k)) - \rho_t(\hat{H}_{tk}) \}.$$

For $\delta \in \Theta_n$, $\tilde{Z}_n(\delta)$ is set to be $\infty$. It holds trivially that

$$\hat{\delta}_n = \sqrt{n}(\hat{\theta}_n(\omega^*) - \theta^*) = \arg\min_{\delta \in \Theta_n} \tilde{Z}_n(\delta).$$

Thus the asymptotic behavior of $\hat{\delta}_n$ follows from the consideration of $\tilde{Z}_n(\delta)$. By Knight's (1998) identity, we have

$$\tilde{Z}_n(\delta) = -\sum_{t=1}^{n} \sum_{k=1}^{K} \tilde{d}_t(\theta_k)\psi_t(\hat{H}_{tk}) + \sum_{t=1}^{n} \sum_{k=1}^{K} \tilde{d}_t(\theta_k) \int_0^1 \left\{ I(\hat{H}_{tk} \leq \tilde{d}_t(\theta_k)s) - I(\hat{H}_{tk} \leq 0) \right\} ds$$

$$\triangleq -\tilde{Z}_{1n}(\delta) + \tilde{Z}_{2n}(\delta).$$

We will show the asymptotic behaviors of $\tilde{Z}_{1n}(\delta)$ and $\tilde{Z}_{2n}(\delta)$ in following lemmas to figure out the asymptotic normality of $\sqrt{n}(\hat{\theta}_n(\omega^*) - \theta^*)$.

**Lemma 6.** Under Assumptions 1, 2 and 3, for $\delta = O_p(1)$, we have

$$|\tilde{Z}_{1n}(\delta) - Z_{1n}(\delta)| = o_p(1),$$

where $Z_{1n}(\delta) = \sum_{t=1}^{n} \sum_{k=1}^{K} \psi_t(\hat{H}_{tk})[q_t(\theta_k^* + \delta_k/\sqrt{n}) - q_t(\theta_k^*)].$

**Proof.** By Taylor’s theorem, $Z_{1n}(\delta)$ is expressed as
\[ Z_{1n}(\delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*)}{\partial \theta} \psi_{t_k}(H_{ik}) + \frac{1}{2n} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial^2 q_t(\theta^*)}{\partial \theta \partial \theta^T} \delta \psi_{t_k}(H_{ik}) \]
\[ \triangleq \Pi_{1n}(\delta) + \Pi_{2n}(\delta), \]
where \( \theta^* \) is a \( p + q + K \) dimensional vector between \( \theta^* \) and \( \theta^* + \delta^T / \sqrt{n} \), then \( \theta^*_k \) is \( p + q + 1 \) subvector of \( \theta^* \) with entries \( \xi^*_k \) and \( \gamma^* \). Note that
\[
\text{Var}[\Pi_{2n}(\delta)] = \frac{1}{4n^2} \sum_{t=1}^{n} E \left[ \left( \sum_{k=1}^{K} \psi_{t_k}(H_{ik}) \right)^2 \right] \leq \frac{\| \delta \|^4}{4n^2} \sum_{t=1}^{n} E[S_t^2] \sum_{i=1}^{K} \left( \tau_i \wedge \tau_j \right) \to 0.
\]

So \( \Pi_{2n}(\delta) = o_p(1) \). With Taylor’s theorem we also give an expression of \( \tilde{Z}_{1n}(\delta) \),
\[
\tilde{Z}_{1n}(\delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial \tilde{q}_t(\theta^*_k)}{\partial \theta} \psi_{t_k}(H_{ik}) + \frac{1}{2n} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial^2 \tilde{q}_t(\theta^*_k)}{\partial \theta \partial \theta^T} \delta \psi_{t_k}(H_{ik})
\]
\[ \triangleq \Pi_{3n}(\delta) + \Pi_{4n}(\delta) + \Pi_{5n}(\delta). \]
By similar technique above, we can prove \( \Pi_{4n}(\delta) = o_p(1) \). To complete our proof, we verify \( |\Pi_{1n}(\delta) - \Pi_{3n}(\delta)| = o_p(1) \) and \( \Pi_{5n}(\delta) = o_p(1) \). Actually
\[
|\Pi_{1n}(\delta) - \Pi_{3n}(\delta)| \leq \frac{\| \delta \|^4}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \left| \frac{\partial q_t(\theta^*_k)}{\partial \theta} - \frac{\partial \tilde{q}_t(\theta^*_k)}{\partial \theta} \right| + \frac{\| \delta \|^4}{\sqrt{n}} \max_{1 \leq i \leq n} \max_{1 \leq k \leq K} \left| \frac{\partial q_t(\theta^*_k)}{\partial \theta} \right| \sum_{t=1}^{n} \sum_{k=1}^{K} \left| \psi_{t_k}(H_{ik}) - \psi_{t_k}(\tilde{H}_{ik}) \right|.
\]
The first term is \( o_p(1) \) by Lemma 5. For the second term, by using mean value theorem and Lemma 5, we have
\[
E[|I(H_{ik} \leq 0) - I(\tilde{H}_{ik} \leq 0)| \mid \mathcal{F}_{t-1}] = E \left[ F\left( \frac{\tilde{q}_t(\theta^*_k)}{q_t(\theta^*_k)} \right) - F\left( \xi^*_k \right) \right] \leq C_1 E \left[ \frac{\tilde{q}_t(\theta^*_k) - q_t(\theta^*_k)}{q_t(\theta^*_k)} \right] \leq C_2 \rho^t.
\]
where \( C \) is a constant. Thus
\[
\sum_{t=1}^{n} \sum_{k=1}^{K} |\psi_{t_k}(H_{ik}) - \psi_{t_k}(\tilde{H}_{ik})| = \sum_{t=1}^{n} \sum_{k=1}^{K} |I(H_{ik} \leq 0) - I(\tilde{H}_{ik} \leq 0)| < \infty \quad \text{a.s.} \quad (5.1)
\]
Hence \( |\Pi_{1n}(\delta) - \Pi_{3n}(\delta)| = o_p(1) \). By similar technique, we can verify \( \Pi_{5n} = o_p(1) \) with Lemma 3 and Eq. (5.1). Now we have completed this proof. □

**Lemma 7.** Under Assumptions 1’, 2 and 3, for \( \delta = o_p(1) \), we have
\[
Z_{1n}(\delta) \to N(0, \delta^T A \delta),
\]
where
\[
A = E \left[ \sum_{k=1}^{K} \sum_{k' = 1}^{K} (\tau_k \wedge \tau_k' - \tau_k \tau_k') \frac{\partial q_t(\theta_k)}{\partial \theta} \frac{\partial q_t(\theta_k')}{\partial \theta^T} \right].
\]

**Proof.** By the proof of Lemma 6, we have verified that
\[
Z_{1n}(\delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*_k)}{\partial \theta} \psi_{t_k}(H_{ik}) + \frac{1}{2n} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial^2 q_t(\theta^*_k)}{\partial \theta \partial \theta^T} \delta \psi_{t_k}(H_{ik})
\]
\[ = \Pi_{1n}(\delta) + \Pi_{2n}(\delta). \]
where $\Pi_{2n}(\delta) = O_p(1)$. Notice that $\sum_{k=1}^{K} \delta^T \frac{\partial q_{t}(\theta^*_t)}{\partial \theta} \psi_{r_t}(H_{tk})$ is strictly stationary and ergodic by Lemma 2. By the independence of $\varepsilon_t$ on $F_{t-1}$, we have

$$E[\Pi_{1n}(\delta)] = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \left[ \frac{\partial q_{t}(\theta^*_t)}{\partial \theta} \right] \psi_{r_t}(H_{tk}) E[\psi_{s_t}(H_{tk})] = 0,$$

$$\text{Var}[\Pi_{1n}(\delta)] = E \left[ \sum_{k=1}^{K} \delta^T \frac{\partial q_{t}(\theta^*_t)}{\partial \theta} \psi_{r_t}(H_{tk}) \right]^2 = \delta^T \Lambda \delta.$$

Hence $Z_{1n}(\delta) \to N(0, \delta^T \Lambda \delta)$ according to the martingale central limit theorem. $A$ is non-singular because $A$ is the variance of a non-degenerate variable. Lemma 3 guarantees $\|A\| < \infty$. This lemma is established.

**Lemma 8.** Under Assumptions 1', 2 and 3, for $\delta = O_p(1)$, we have

$$|\tilde{Z}_{2n}(\delta) - Z_{2n}(\delta)| = o_p(1),$$

where

$$\tilde{Z}_{2n}(\delta) = \sum_{t=1}^{n} \sum_{k=1}^{K} \tilde{d}_t(\theta_k) \int_0^1 \left[ \left( \sum_{i=k}^{K} \tilde{d}_k(\theta_k) - \tilde{d}_t(\theta_k) \right) s \right] ds,$$

$$Z_{2n}(\delta) = \sum_{t=1}^{n} \sum_{k=1}^{K} d_t(\theta_k) \int_0^1 \left[ \left( \sum_{i=k}^{K} d_k(\theta_k) - d_t(\theta_k) \right) s \right] ds.$$

**Proof.** Let

$$\tilde{B}_{tk} = \int_0^1 \left[ \left( \sum_{i=k}^{K} \tilde{d}_k(\theta_k) - \tilde{d}_t(\theta_k) \right) s \right] ds,$$

and

$$B_{tk} = \int_0^1 \left[ \left( \sum_{i=k}^{K} d_k(\theta_k) - d_t(\theta_k) \right) s \right] ds.$$

Note that $d_t(\theta_k) = q_t(\theta^*_t + \delta_k/\sqrt{n}) - q_t(\theta^*_k)$, so

$$Z_{2n}(\delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*_t)}{\partial \theta} B_{tk} + \frac{1}{2n} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial^2 q_t(\theta^*_t)}{\partial \theta \partial \theta^T} \delta B_{tk},$$

where $\theta'$ is a $p + q + K$ dimensional vector between $\theta^*$ and $\theta^* + \delta/\sqrt{n}$, then $\theta'_k$ is a $p + q + 1$ subvector of $\theta'$ with entries $\xi_k'$ and $\gamma'$. By similar discussion in Lemma 6, we have

$$\frac{1}{2n} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial^2 q_t(\theta^*_t)}{\partial \theta \partial \theta^T} \delta B_{tk} = O_p(1).$$

Hence we can get

$$Z_{2n}(\delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*_t)}{\partial \theta} B_{tk} + o_p(1)$$

and

$$\tilde{Z}_{2n}(\delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*_t)}{\partial \theta} \tilde{B}_{tk} + o_p(1).$$

Then

$$Z_{2n}(\delta) - \tilde{Z}_{2n}(\delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \left( \frac{\partial q_t(\theta^*_t)}{\partial \theta} - \frac{\partial \tilde{q}_t(\theta^*_t)}{\partial \theta} \right) B_{tk}$$

$$+ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial \tilde{q}_t(\theta^*_t)}{\partial \theta} \int_0^1 \left[ \left( \sum_{i=k}^{K} \tilde{d}_k(\theta_k) - \tilde{d}_t(\theta_k) \right) s \right] ds$$

$$+ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial \tilde{q}_t(\theta^*_t)}{\partial \theta} \int_0^1 \left[ \left( \sum_{i=k}^{K} d_k(\theta_k) - d_t(\theta_k) \right) s \right] ds + O_p(1)$$

Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.eosta.2016.11.004
\[ \Delta \equiv \Pi_{\theta M}(\delta) + \Pi_{\theta N}(\delta) + \Pi_{\theta b}(\delta) + o_p(1). \]

It can be seen that \( \Pi_{\theta M}(\delta) = o_p(1) \) by Lemma 5. Let \( \Pi_{\theta N}(\delta) = \frac{1}{n} \sum_{t=1}^{n} \frac{T_{1t}}{T_{1t}}. \) By Lemma 3 and Fubini’s theorem, we have

\[ E[|T_{1n}| |F_{1n-1}|] \leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E[|T_{1t}|] \leq \frac{2}{\sqrt{n}} |K_1| E \left[ \frac{S_{1t}V^{1/2}}{v_t} \right] \sum_{t=1}^{n} \rho_t \to 0. \]

Hence

\[ E[||\Pi_{\theta N}(\delta)||] \leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E[|T_{1n}|] \leq \frac{2}{\sqrt{n}} |K_1| E \left[ \frac{S_{1t}V^{1/2}}{v_t} \right] \sum_{t=1}^{n} \rho_t\to 0. \]

Besides by Lemma 3 and Eq. (5.1), we prove \( \Pi_{\theta M}(\delta) = o_p(1) \) with similar technique in Lemma 6. This proof is completed. \( \square \)

**Lemma 9.** Under Assumptions 1', 2 and 3, for \( \delta = O_p(1) \), we have

\[ Z_{2\delta_2}(\delta) \to \frac{1}{2} \delta^T B \delta. \]

where \( B = \sum_{k=1}^{K} f(\xi^*_k) E[ \frac{1}{v_t(\gamma^*)} \frac{\partial q_t(\theta^*_k)}{\partial \theta} \frac{\partial q_t(\theta^*_k)}{\partial \theta} ] \).

**Proof.** Note that \( d_t(\theta_k) = q_t(\theta_k) + \delta_k/\sqrt{n} - q_t(\theta_k^*) \). By the proof of Lemma 8, we have

\[ Z_{2\delta_2}(\delta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*_k)}{\partial \theta} B_{tk} + o_p(1) \Delta \sum_{t=1}^{n} J_t + o_p(1) \]

\[ = \sum_{t=1}^{n} E[J_{nt}|F_{1t-1}] + \sum_{t=1}^{n} [J_{nt} - E[J_{nt}|F_{1t-1}] + o_p(1) \Delta \sum_{t=1}^{n} \Pi_{\theta M} + \Pi_{\theta b} + o_p(1). \]

By Fubini’s theorem and Taylor’s theorem,

\[ E[J_{nt}|F_{1t-1}] = \frac{1}{\sqrt{n}} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*_k)}{\partial \theta} E \left[ \int_{0}^{1} [I(H_{tk} \leq d_t(\theta_k)s) - I(H_{tk} \leq 0)]ds|F_{1t-1} \right] \]

\[ = \frac{1}{2\sqrt{n}} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*_k)}{\partial \theta} \frac{d_t(\theta_k)}{v_t(\gamma^*)} f(\xi^*_k) + \frac{1}{6\sqrt{n}} \sum_{k=1}^{K} \delta^T \frac{\partial q_t(\theta^*_k)}{\partial \theta} \frac{d_t^2(\theta_k)}{v_t^2(\gamma^*)} f'(\xi^*_k), \]

where \( \xi_k^* \) is between \( \xi_k \) and \( \xi_k^* \) for any \( 0 \leq k \leq K \). Note that \( f(\xi^*_k) \) and \( f'(\xi^*_k) \) are uniformly bounded by Assumption 3, and

\[ \frac{d_t^2(\theta_k)}{v_t^2(\gamma^*)} = \frac{1}{v_t^2(\gamma^*)} \left( q_t(\theta_k^* + \frac{\delta}{\sqrt{n}}) - q_t(\theta_k^*) \right)^2 \]

\[ = \frac{1}{v_t^2(\gamma^*)} \left( \frac{\delta^T}{\sqrt{n}} \frac{\partial q_t(\theta_k^*)}{\partial \theta} + \frac{1}{2n} \delta^T \frac{\partial^2 q_t(\theta_k^*)}{\partial \theta \partial \theta} \delta \right)^2 = O_p \left( \frac{1}{n} \right). \]
So we can get \( \frac{1}{\sqrt{n}} \sum_{k=1}^K \delta^T \left( \frac{\partial q_t(\theta^*_t)}{\partial \theta} \right) \frac{d_i(\theta_k)}{v_t(\gamma^*)} f(\xi_k^*) = o_p(1) \). Hence

\[
E[U_{it} | \mathcal{F}_{t-1}] = \frac{1}{2\sqrt{n}} \sum_{k=1}^K \delta^T \frac{\partial q_t(\theta^*_t)}{\partial \theta} d_i(\theta_k) f(\xi_k^*) = \frac{1}{2\sqrt{n}} \sum_{k=1}^K \delta^T \frac{\partial q_t(\theta^*_t)}{\partial \theta} d_i(\theta_k) f(\xi_k^*) \]

\[
= \frac{1}{2n} \sum_{k=1}^K f(\xi_k^*) \delta^T \left[ \sum_{k=1}^K f(\xi_k^*) \frac{1}{v_t(\gamma^*)} \frac{\partial q_t(\theta^*_t)}{\partial \theta} \frac{\partial q_t(\theta^*_t)}{\partial \theta^T} \right] + \frac{1}{4n^{3/2}} \sum_{k=1}^K f(\xi_k^*) \delta^T \left[ \frac{\partial q_t(\theta^*_t)}{\partial \theta} \frac{\partial^2 q_t(\theta^*_t)}{\partial \theta \partial \theta^T} \right] \delta + o_p\left( \frac{1}{n^{3/2}} \right).
\]

Together with the strictly stationary and ergodic of \( \{ \frac{\partial q_t(\theta^*_t)}{\partial \theta}, t \in \mathbb{Z} \} \) in Lemma 2, we have

\[
\sum_{t=1}^n E[U_{it} | \mathcal{F}_{t-1}] = \frac{1}{2n} \sum_{t=1}^n \sum_{k=1}^K f(\xi_k^*) \delta^T \left[ \sum_{k=1}^K f(\xi_k^*) \frac{1}{v_t(\gamma^*)} \frac{\partial q_t(\theta^*_t)}{\partial \theta} \frac{\partial q_t(\theta^*_t)}{\partial \theta^T} \right] \delta + o_p(1)
\]

\[
= \frac{1}{2} \delta^T B \delta.
\]

As to \( \Pi_{10n} \), we have

\[
\sum_{t=1}^n E\left[ \frac{U_{it} - E[U_{it} | \mathcal{F}_{t-1}]}{\mathcal{F}_{t-1}} \right]^2 | \mathcal{F}_{t-1} = \sum_{t=1}^n E[|U_{it}|^2 | \mathcal{F}_{t-1}]
\]

\[
\leq \max_{1 \leq t \leq n, 1 \leq k \leq K} \frac{K}{\sqrt{n}} \left| \frac{\partial q_t(\theta^*_t)}{\partial \theta} \right| \sum_{t=1}^n E[U_{it} | \mathcal{F}_{t-1}] = o_p(1).
\]

Hence \( \text{Var}(\Pi_{10n}) \rightarrow 0 \) as \( n \rightarrow \infty \). By this way, we have prove \( Z_{2n} \rightarrow \frac{1}{2} \delta^T B \delta \). Next we verify \( B \) is non-singular. If \( p + q + K \) dimensional vector \( \lambda \) satisfies

\[
\lambda^T B \lambda = K \sum_{k=1}^K f(\xi_k^*) \left[ \frac{1}{v_t(\gamma^*)} \left( \lambda^T \frac{\partial q_t(\theta^*_t)}{\partial \theta} \right)^2 \right] = 0,
\]

then \( \lambda^T \frac{\partial q_t(\theta^*_t)}{\partial \theta} = 0 \) a.s. for any \( 1 \leq k \leq K \). Let \( \lambda^T = (\lambda_0^T, \lambda_1^T) = (\lambda_{01}, \ldots, \lambda_{0K}, \lambda_{11}, \ldots, \lambda_{1(p+q)}) \) and \( \epsilon_k \) is a \( K \) dimensional vector with the \( k \)th element 1 and others 0. By (2.1), we have

\[
2\nu^2 \lambda_0^T \epsilon_k + \xi_k \lambda_1^T \frac{\partial u^2}{\partial \gamma} = x^T \left( \begin{array}{c} e_k \\ r_{t-1}^2 \\ \vdots \\ r_{t-1}^{2q} \\ \vdots \\ v_{t-1}^2 \\ \vdots \\ v_{t-1}^{2p} \\ \vdots \\ v_{t-p}^2 \end{array} \right) + \sum_{j=1}^p \beta_j \left( 2\nu^2 \lambda_0^T \epsilon_k + \xi_k \lambda_1^T \frac{\partial u_{t-j}^2}{\partial \gamma} \right).
\]

where \( x^T = 2\lambda_0^T \epsilon_k (e_k^T, \alpha^T, 0_p^T) + (0_p, \xi_k \lambda_1^T_1) \triangleq (x_1, \ldots, x_{p+q+p}) \). Note that \( 2\nu^2 \lambda_0^T \epsilon_k + \xi_k \lambda_1^T_1 \frac{\partial u_{t-j}^2}{\partial \gamma} = 0 \) for any \( j \geq 0 \), so

\[
x^T \left( \begin{array}{c} e_k \\ r_{t-1}^2 \\ \vdots \\ r_{t-1}^{2q} \\ \vdots \\ v_{t-1}^2 \\ \vdots \\ v_{t-p}^2 \end{array} \right) = 0.
\]

Considering that \( \eta_{t-1} \) is independent on \( \mathcal{F}_{t-2} \), we have \( x_2 = 0 \). If \( x \neq 0 \), this model degenerates into a GARCH(\( p-1, q-1 \)) model at least, which is against our assumption. \( x = 0 \) implies \( \lambda = 0 \) when \( k \) takes all its values. Thus \( B \) is non-singular. \( \square \)
Note that Lemmas 6–9 is valid for $\delta = O_p(1)$. We first verify $\sqrt{n}(\hat{\theta}_n(\mu^*) - \theta^*) = O_p(1)$. In the proof of Theorem 1, we have
\[
G_n(\theta) = \tilde{G}_n(\theta) + O_p\left(\frac{1}{n}\right).
\]
By the proof of Lemmas 6–9, we see
\[
\frac{1}{n}Z_{1n}(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^{n} \left( \hat{\theta}_n - \theta^* \right) \frac{\partial q_t(\theta^*_k)}{\partial \theta} \psi_t(H_{tk}) + o_p\left(\frac{|\hat{\theta}_n - \theta^*|}{\sqrt{n}}\right)
\]
\[
\triangleq \frac{(\hat{\theta}_n - \theta^*)}{\sqrt{n}} A_n + o_p\left(\frac{|\hat{\theta}_n - \theta^*|}{\sqrt{n}}\right),
\]
\[
\frac{1}{n}Z_{2n}(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^{n} E[U_{it} | F_{t-1}] + \frac{1}{n} \sum_{t=1}^{n} E[U_{it} - E[U_{it} | F_{t-1}]] + o_p\left(\frac{|\hat{\theta}_n - \theta^*|}{\sqrt{n}}\right),
\]
where $A_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sum_{k=1}^{K} \frac{\partial q_k(\theta^*_k)}{\partial \theta} \psi_t(H_{tk})$. And
\[
\frac{1}{n} \sum_{t=1}^{n} E[U_{it} | F_{t-1}] = \frac{1}{2n} \sum_{t=1}^{n} (\hat{\theta}_n - \theta^*) \sum_{k=1}^{K} f(\xi^*_k) \frac{1}{v_t(\gamma^*)} \frac{\partial q_k(\theta^*_k)}{\partial \theta} \frac{\partial q_k(\theta^*_k)}{\partial \theta^*} \frac{(\hat{\theta}_n - \theta^*)}{\sqrt{n}} + o_p\left(\frac{|\hat{\theta}_n - \theta^*|^2}{\sqrt{n}}\right),
\]
and
\[
\frac{1}{n} \sum_{t=1}^{n} E[E[U_{it} | F_{t-1}]] = o_p\left(\frac{|\hat{\theta}_n - \theta^*|^2}{\sqrt{n}}\right).
\]
So we can get
\[
\tilde{G}_n(\hat{\theta}_n) = -\frac{1}{n}Z_{1n}(\hat{\theta}_n) + \frac{1}{n}Z_{2n}(\hat{\theta}_n)
\]
\[
= -\frac{(\hat{\theta}_n - \theta^*)}{\sqrt{n}} A_n + \frac{1}{2} (\hat{\theta}_n - \theta^*)^T B_n (\hat{\theta}_n - \theta^*) + o_p\left(\frac{|\hat{\theta}_n - \theta^*|^2}{\sqrt{n}}\right) + o_p\left(\frac{1}{n}\right).
\]
where $A_n \stackrel{d}{=} N(0, A)$ by Lemma 7 and $B_n$ converges to $B$ by Lemma 9. Note that $\hat{\theta}_n$ minimizes $\tilde{S}_n(\theta)$, so
\[
0 \geq \tilde{S}_n(\hat{\theta}_n) - \tilde{S}_n(\theta^*)
\]
\[
\geq -\frac{(\hat{\theta}_n - \theta^*)}{\sqrt{n}} |A_n| + \frac{1}{2} (\hat{\theta}_n - \theta^*)^T B_n (\hat{\theta}_n - \theta^*) - o_p\left(\frac{|\hat{\theta}_n - \theta^*|^2}{\sqrt{n}}\right) - o_p\left(\frac{|\hat{\theta}_n - \theta^*|}{\sqrt{n}}\right) - O_p\left(\frac{1}{n}\right),
\]
where $b_n$ is the smallest eigenvalue of $B_n$ and converges to the smallest eigenvalue of $B$ in probability. So
\[
\frac{|A_n| + o_p(1)}{b_n - o_p(1)^2} + \frac{2|o_p(1)|}{b_n - o_p(1)} \geq \left(\frac{\sqrt{n}|\hat{\theta}_n - \theta^*| - |A_n| + o_p(1)}{b_n - o_p(1)}\right)^2,
\]
which implies $\sqrt{n}(\hat{\theta}_n - \theta^*) = O_p(1)$. By combining above lemmas, we can see, for $\delta = O_p(1)$,
\[
\tilde{Z}_n(\delta) \rightarrow Z_0(\delta) \triangleq -\delta^T N(0, A) + \frac{1}{2} \delta^T B \delta.
\]
The convexity of the limiting objective function, $Z_0(\delta)$, ensures its unique minimizer and, consequently that
\[
\hat{\delta}_n = \sqrt{n}(\hat{\theta}_n(\omega^*) - \theta^*) = \text{argmin} \tilde{Z}_n(\delta) \rightarrow \delta_0 \triangleq \text{argmin} Z_0(\delta).
\]
Finally we have proved that
\[
\sqrt{n}(\hat{\theta}_n(\omega^*) - \theta^*) \rightarrow N(0, B^{-1}AB^{-1}),
\]
where
\[
A = E\left[\sum_{k=1}^{K} \sum_{k'=1}^{K} (T_k \cap T_{k'} - T_k T_{k'}) \frac{\partial q_k(\theta^*_k)}{\partial \theta} \frac{\partial q_{k'}(\theta^*_k)}{\partial \theta^*}\right] \quad \text{and} \quad B = \sum_{k=1}^{K} f(\xi^*_k) E\left[\frac{1}{v_t(\gamma^*)} \frac{\partial q_k(\theta^*_k)}{\partial \theta} \frac{\partial q_{k'}(\theta^*_k)}{\partial \theta^*}\right].
\]

Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.ecosta.2016.11.004
Proof of Lemma 1

We divide matrix A into four submatrices:

\[ A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \]

where \( A_{11} \) is \( K \times K \) order submatrix with \((i,j)\) element \((\tau_i \wedge \tau_j - \tau_i \tau_j)F[v_i^2]; A_{12} \) is \( K \times (p+q) \) order submatrix with the ith row vector \( \sum_{j=1}^{K}(\tau_i \wedge \tau_j - \tau_i \tau_j) \xi E[v_i] \frac{\partial \eta}{\partial \gamma^T} \); \( A_{21} = A_{12}^T \) and \( A_{22} = \sum_{i=1}^{K} \sum_{j=1}^{K}(\tau_i \wedge \tau_j - \tau_i \tau_j) \xi E[v_i] \frac{\partial \eta}{\partial \gamma^T} \). Similarly for matrix \( B \) we have

\[ B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \]

where \( B_{11} \) is \( K \times K \) order diagonal matrix with ith element on the diagonal \( f(\xi_k)E[v_i]; B_{12} \) is \( K \times (p+q) \) order submatrix with the ith row vector \( f(\xi_k)\xi E[v_i] \frac{\partial \eta}{\partial \gamma^T} \); \( B_{21} = B_{12}^T \) and \( B_{22} = \sum_{k=1}^{K} f(\xi_k)\xi E[v_i] \frac{\partial \eta}{\partial \gamma^T} \). Note that \( A \) and \( B \) is strictly positive and reversible. The inverse of block matrix \( B \) is

\[ B^{-1} = \begin{pmatrix} B_{11}^{-1} + B_{11}^{-1}B_{12}B_{22}^{-1}B_{11}^{-1} & -B_{11}^{-1}B_{12}B_{22}^{-1} \\ -B_{22}^{-1}B_{21}B_{11}^{-1} & B_{22}^{-1} \end{pmatrix} \]

where \( B_{22,1} = B_{22} - B_{21}B_{11}^{-1}B_{12} \). So the lower right block of \( B^{-1}AB^{-1} \) is

\[ V = B_{22,1}^{-1}B_{21}B_{11}^{-1}A_{11}^{-1}B_{12,22,1}^{-1}B_{22}^{-1}B_{21}^{-1}A_{12}^{-1}B_{12,22,1}^{-1} + B_{22,1}^{-1}A_{22}^{-1} \]

Now we calculate \( V \) term by term:

\[ B_{22,1}^{-1} = \frac{K}{\sum_{k=1}^{K} f(\xi_k)\xi_k^2} \left\{ E \left[ \frac{1}{v_i} \frac{\partial v_i}{\partial \gamma} \frac{\partial v_i}{\partial \gamma^T} \right] - E[v_i]^{-1}E \left[ \frac{\partial v_i}{\partial \gamma} \frac{\partial v_i}{\partial \gamma^T} \right] \right\} \]

\[ B_{21}B_{11}^{-1}A_{11}^{-1}B_{12}^{-1} = \frac{K}{\sum_{i=1}^{K} (\sum_{j=1}^{K} (\tau_i \wedge \tau_j - \tau_i \tau_j) \xi_j E[v_i]^{-1}E[v_i^2])} \left\{ \sum_{i=1}^{K} (\tau_i \wedge \tau_j - \tau_i \tau_j) \xi_j E[v_i]^{-1}E[v_i^2] \right\} \]

\[ B_{21}B_{11}^{-1}A_{12}^{-1} = \frac{K}{\sum_{i=1}^{K} (\sum_{j=1}^{K} (\tau_i \wedge \tau_j - \tau_i \tau_j) \xi_j E[v_i]^{-1}E[v_i^2])} \left\{ \sum_{i=1}^{K} (\tau_i \wedge \tau_j - \tau_i \tau_j) \xi_j E[v_i]^{-1}E[v_i^2] \right\} \]

\[ A_{21}^{-1}B_{22}^{-1} = (B_{21}B_{11}^{-1}A_{12})^T. \]

Hence

\[ V = \frac{\sum_{i=1}^{K} (\sum_{j=1}^{K} (\tau_i \wedge \tau_j - \tau_i \tau_j) \xi_j E[v_i]^{-1}E[v_i^2])}{(\sum_{k=1}^{K} f(\xi_k)\xi_k^2)^2} V_1^{-1}V_1, \]

where

\[ V_1 = E \left[ \frac{1}{v_i} \frac{\partial v_i}{\partial \gamma} \frac{\partial v_i}{\partial \gamma^T} \right] - E[v_i]^{-1}E \left[ \frac{\partial v_i}{\partial \gamma} \frac{\partial v_i}{\partial \gamma^T} \right] \]

\[ V_0 = E[v_i]^{-2}E[v_i^2] \left\{ \frac{\partial v_i}{\partial \gamma} \frac{\partial v_i}{\partial \gamma^T} - E[v_i]^{-1}E \left[ \frac{\partial v_i}{\partial \gamma} \frac{\partial v_i}{\partial \gamma^T} \right] \right\} \]

\[ - E[v_i]^{-1}E \left[ \frac{\partial v_i}{\partial \gamma} \frac{\partial v_i}{\partial \gamma^T} \right] + E \left[ \frac{\partial v_i}{\partial \gamma} \frac{\partial v_i}{\partial \gamma^T} \right]. \]

Here we can see that \( V_0 \) and \( V_1 \) only depend on \( \gamma \) and are unrelated to proxies or quantiles.

References


Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.ecosta.2016.11.004

Please cite this article as: M. Wang et al., Composite quantile regression for GARCH models using high-frequency data, Econometrics and Statistics (2017), http://dx.doi.org/10.1016/j.ecosta.2016.11.004