Conditional Quantile Estimation for GARCH Models

Authors: Zhijie Xiao, Roger Koenker

Boston College Working Papers in Economics, 2009

CONDITIONAL QUANTILE ESTIMATION FOR GARCH MODELS

ZHIJIE XIAO AND ROGER KOENKER

Abstract. Conditional quantile estimation is an essential ingredient in modern risk management. Although GARCH processes have proven highly successful in modeling financial data it is generally recognized that it would be useful to consider a broader class of processes capable of representing more flexibly both asymmetry and tail behavior of conditional returns distributions. In this paper, we study estimation of conditional quantiles for GARCH models using quantile regression. Quantile regression estimation of GARCH models is highly nonlinear; we propose a simple and effective two-step approach of quantile regression estimation for linear GARCH time series. In the first step, we employ a quantile autoregression sieve approximation for the GARCH model by combining information over different quantiles; second stage estimation for the GARCH model is then carried out based on the first stage minimum distance estimation of the scale process of the time series. Asymptotic properties of the sieve approximation, the minimum distance estimators, and the final quantile regression estimators employing generated regressors are studied. These results are of independent interest and have applications in other quantile regression settings. Monte Carlo and empirical application results indicate that the proposed estimation methods outperform some existing conditional quantile estimation methods.

1. Introduction

Distributional information such as conditional quantiles and variances play an essential role in risk measurement. Evaluation of Value-at-Risk, as mandated in many current regulatory contexts, is explicitly a conditional quantile estimation problem. Closely related quantile-based concepts such as expected shortfall, conditional value at risk, and limited expected loss, are also intimately linked to quantile estimation, see, e.g., Artzner, Delbaen, Eber, and Heath (1999), Wang (2000), Wu and Xiao (2002), and Bassett, Koenker, and Kordas (2004).

The literature on estimating conditional quantiles is large. Many existing methods of quantile estimation in economics and finance are based on the assumption that financial returns have normal (or conditional normal) distributions. Under the assumption of a conditionally normal returns distribution, the estimation of conditional quantiles is equivalent to estimating conditional volatility of returns. The massive literature on volatility modeling offers a rich source of parametric methods of this type. However, there is accumulating evidence that financial time series, and returns distributions are *not* well approximated by Gaussian models. In particular, it is frequently found that market returns display negative

Version: March 13, 2009. We thank the editor, an associate editor, three referees and participants at the 2007 JSM, MIT, and the Cass Conference in Econometrics for helpful comments and discussions on an earlier version of this paper. We also thank Chi Wan for excellent research assistance. Research was partially supported by NSF grant SES-05-44673.

skewness and excess kurtosis. Extreme realizations of returns can adversely effect the performance of estimation and inference designed for Gaussian conditions; this is particularly true of ARCH and GARCH models whose estimation of variances are very sensitive to large innovations. For this reason, research attention has recently shifted toward the development of more robust estimators of conditional quantiles.

There is growing interest in non-parametric estimation of conditional quantiles; although local, nearest neighbor and kernel methods are somewhat limited in their ability to cope with more than one or two covariates. Other approaches to estimating VaR include the hybrid method of Boudoukh, Richardson, and Whitelaw (1998) and methods based on extreme value theory see, e.g. Boos (1984), McNeil (1998), and Neftci (2000)

Quantile regression as introduced by Koenker and Bassett (1978) is well suited to estimating conditional quantiles. Just as classical linear regression methods based on minimizing sums of squared residuals enable one to estimate models for conditional mean, quantile regression methods offer a mechanism for estimating models for the conditional quantiles. These methods exhibit robustness to extreme shocks, and facilitate distribution-free inference.

In recent years, quantile regression estimation for time-series models has gradually attracted more attention. Koenker and Zhao (1996) extended quantile regression to linear ARCH models where $\sigma_t = \gamma_0 + \gamma_1 |u_{t-1}| + \cdots + \gamma_q |u_{t-q}|$, and estimate conditional quantiles of u_t by a linear quantile regression of u_t on $(1, |u_{t-1}|, \cdots, |u_{t-q}|)$. However, evidence from financial applications indicates that, comparing to the GARCH models, ARCH type of models can not parsimoniously capture the persistent influence of long past shocks.

Engle and Manganelli (2004) suggest a nonlinear dynamic quantile model where conditional quantiles themselves follow an autoregression. In particular, they propose the following Conditional Autoregressive Value at Risk (CAViaR) specification for the τ-th conditional quantile of y_t:

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \beta_0 + \sum_{i=1}^{p} \beta_i Q_{y_{t-i}}(\tau|\mathcal{F}_{t-i-1}) + \sum_{j=1}^{q} \alpha_j \ell(x_{t-j})$$

where $x_{t-j} \in \mathcal{F}_{t-j}$, $\mathcal{F}_{t-j}$ is the information set at time $t-j$, and $Q_{y_t}(\tau|\mathcal{F}_{t-1})$ is the conditional quantile of y_t given information set $\mathcal{F}_{t-1}$. The CAViaR model has attracted a great deal of research attension in recent years. The focus of Engle and Manganelli (2004) is on introduction of the CAViaR model instead of how to estimate such models. In the CAViaR model, since the regressors $Q_{y_{t-i}}(\tau|\mathcal{F}_{t-i-1})$ are latent and are dependent on the unknown parameters, estimation of the CAViaR model is complicate and conventional nonlinear quantile regression techniques are not directly applicable. Engle and Manganelli (2004) use grid searching combined with recursive iteration of existing Matlab optimization algorithms to obtain an approximation for the conditional quantile. Rossi and Harvey (2009) recently proposed an iterative Kalman filter method to calculate dynamic conditional quantiles that can be applied to calculate the CAViaR model.

There are some recent studies on estimation and applications of estimating conditional quantiles. In particular, based on a relation between expectile and quantile, Taylor (2008a) and Kuan, Yeh, and Hsu (2009) estimate conditional quantiles using asymmetric least squares methods. Taylor (2008b) proposes the exponentially weighted quantile regression for estimating time-varying quantiles. Gourieroux and Jasiak (2008)) proposed the Dynamic Additive Quantile (DAQ) model for calculating conditional quantiles. Giot and Laurent (2003) model Value-at-Risk using parametric ARCH models based on skewed t-distributions. Coroneo and Veredas (2008) analyze the density of high frequency financial returns using quantile regression of ARCH models and conducted intradaily Value-at-Risk assessment.

In this paper, we study quantile regression estimation for a class of GARCH models. GARCH models have proven to be highly successful in modelling financial data, and is arguably the most widely used class of models in financial applications. However, quantile regression GARCH models are highly nonlinear and thus complicated to estimate. As will become apparent in our later discussion, the quantile estimation problem in GARCH models corresponds to a restricted nonlinear quantile regression and conventional nonlinear quantile regression techniques are not directly applicable, adding a new challenge to the already complicated estimation problem. To circumvent these difficulties, we propose a robust and easy-to-implement two-step approach for quantile regression on GARCH models. The proposed estimation procedure consists a global estimation in the first step to incorporate the global restriction on the conditional scale parameter, and a second step local estimation for the conditional quantiles. In particular, although different implementations are possible, we suggest that in the first step, a sieve quantile regression approximation is estimated for multiple quantiles, and combined via minimum distance methods to obtain preliminary estimators for the parameters of the global GARCH model. In the terminology of Aitchison and Brown (1957) this procedure can be viewed as an extension of their "method of quantiles." The second step then focuses on the local behavior at the specific quantile and estimates the conditional quantile based on the first stage results. The proposed method is relatively easy to implement compared to other nonlinear estimation techniques in quantile regression and has good sampling performance in our simulation experiments. The methods that we employ to study the asymptotic behavior of our two-stage procedure: combining information over quantiles via minimum distance estimation, and quantile regression with generated regressors are also of independent interest and applicable in other econometric and statistical applications.

As will be made explicit in section 2, the linear GARCH process has a CAViaR(p, q) representation Instead of focusing on discussion of the model itself, we focus on estimation of this model. In this sense, the estimation procedure that we propose in the current paper provides a method of calculating a class of CAViaR models. Comparing to studies in the existing literature, instead of only looking at local properties at the specified quantile, the proposed procedure takes into account both global model coherence and optimal local approximation. Since the GARCH model has been proved to be highly successful in financial applications, estimates that are globally coherent with the GARCH feature seem appealing in financial applications. Third, the estimation procedure in our paper also provides a robust estimator for the conditional volatility. Such an estimator is not dependent on distributional assumptions, and thus robust to skewed and heavy-tailed innovations.

The remainder of the paper is organized as follows: We discuss the estimation of conditional quantiles in GARCH models and propose the two-stage estimation procedure in the next section; Section 3 studies the asymptotic behavior of the proposed estimators in each stage, including the sieve quantile estimation, the minimum distance estimation that combines information over various quantiles, and the proposed two-step estimator. The results of a small Monte Carlo experiment are reported in Section 4.

2. Quantile Regression for Linear GARCH Models

Since Bollerslev (1986), a variety of GARCH models have been proposed by various researchers, including the EGARCH model of Nelson (1991) and the linear GARCH model of Taylor (1986). In the original quadratic form of the GARCH model we say that : u_t follows a GARCH(p, q) process if

$$\begin{aligned} u_t &= \sigma_t \cdot \varepsilon_t, \\ \sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 + \gamma_1 u_{t-1}^2 + \cdots + \gamma_q u_{t-q}^2, \end{aligned}$$

where ε_t is an iid sequence of mean zero Gaussian random variables. As noted by Pan and Duffie (1997), maximum likelihood estimation of this form of the GARCH model has the potential disadvantage that it is overly sensitivity to extreme returns. For example, if we consider a market crash, extreme daily absolute returns may be 10 to 20 times normal daily fluctuation, so the quadratic form of GARCH model yields a return effect which is 100 to 400 times the normal variance. This not only causes overshooting in volatility forecasting, but also carries this influence far into the future. As an alternative, Taylor (1986) suggested a modified GARCH model: we will say that u_t follows a linear GARCH(p, q) process if

$$u_t = \sigma_t \cdot \varepsilon_t, \tag{1}$$

$$\sigma_t = \beta_0 + \beta_1 \sigma_{t-1} + \cdots + \beta_p \sigma_{t-p} + \gamma_1 \left| u_{t-1} \right| + \cdots + \gamma_q \left| u_{t-q} \right|. \tag{2}$$

The quadratic GARCH model seems computationally more convenient than the linear GARCH model, but linear GARCH may be more appropriate in modelling financial returns. The linear GARCH structure is less sensitive to extreme returns, but it is more difficult to handle mathematically. However, the linear structure is well suited for quantile estimation.

We will consider quantile regression estimation for the linear GARCH model (1) and (2), where $\beta_0 > 0$, $(\gamma_1, \cdots, \gamma_q)^\top \in \Re_+^q$, and ε_t are independent and identically distributed with mean zero and unknown distribution function $F_\varepsilon(\cdot)$. We will admit a general class of distributions for ε_t, including the normal distribution and other commonly used distributions for financial applications with asymmetry and heavier tails. Our primary purpose is to estimate the τ-th conditional quantile of u_t, but we also provide robust estimators for the conditional volatility as well as the GARCH parameters.

2.1. **Conditional Quantiles for the Linear GARCH Model.**

Let $\mathcal{F}_{t-1}$ represents information up to time $t-1$, the τ-th conditional quantile of u_t is given by

$$Q_{u_t}(\tau | \mathcal{F}_{t-1}) = \theta(\tau)^\top z_t,$$

where

$$\begin{aligned} z_t &= (1, \sigma_{t-1}, \cdots, \sigma_{t-p}, \left| u_{t-1} \right|, \cdots, \left| u_{t-q} \right|)^\top, \\ \theta(\tau)^\top &= (\beta_0, \beta_1, \cdots, \beta_p, \gamma_1, \cdots, \gamma_q) F^{-1}(\tau). \end{aligned}$$

Notice that $\sigma_{t-j}F^{-1}(\tau) = Q_{u_{t-j}}(\tau|\mathcal{F}_{t-j-1})$, the conditional quantile $Q_{u_t}(\tau|\mathcal{F}_{t-1})$ has the following CAViaR(p, q) representation

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) = \beta_0^* + \sum_{i=1}^{p} \beta_i^* Q_{u_{t-i}}(\tau|\mathcal{F}_{t-i-1}) + \sum_{j=1}^{q} \gamma_j^* |u_{t-j}| \tag{3}$$

where

$$\beta_0^* = \beta_0(\tau) = \beta_0 F^{-1}(\tau), \beta_i^* = \beta_i, i = 1, \cdots, p \text{ and } \gamma_j^* = \gamma_j(\tau) = \gamma_j F^{-1}(\tau), j = 1, \cdots, q.$$

Remark. More generally, we may consider a time series y_t in a regression model, say,

$$y_t = \mu^\top X_t + u_t, \tag{4}$$

where the residuals u_t follow a linear GARCH process as characterized by (1). Under weak regularity conditions, the τ-th conditional quantile of y_t in the model (4) is given by

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \mu^\top X_t + \theta(\tau)^\top z_t,$$

where $X_t = (1, x_{2,t}, ..., x_{k,t})^\top$. In the above problem, the key component is the estimation of conditional quantiles of the process u_t: $Q_{u_t}(\tau|\mathcal{F}_{t-1}) = \theta(\tau)^\top z_t$. For this reason, we focus our discussion on model (1) and (2). ■

2.2. **Quantile Regression Estimation of GARCH Models.** Quantile regression provides a convenient approach of estimating conditional quantiles. It has the important virtue of robustness to distributional assumptions and makes no prior presumption about the symmetry of the innovation process. Such properties are especially attractive for financial applications since often financial data like portfolio returns or log returns are heavy-tailed and asymmetrically distributed. We begin by considering estimating of the conditional quantiles of u_t given by (1) employing quantile regression.

Since z_t contains σ_{t-k} $(k = 1, \cdots, q)$ which in turn depend on unknown parameters $\theta = (\beta_0, \beta_1, \cdots, \beta_p, \gamma_1, \cdots, \gamma_q)$, we will write z_t as $z_t(\theta)$ whenever it is necessary to emphasize the nonlinearity and its dependence on θ. To estimate the conditional quantiles of the process u_t we consider the following nonlinear quantile regression estimator solving:

$$\min_{\theta} \sum_t \rho_\tau(u_t - \theta^\top z_t(\theta)), \tag{5}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$. However, estimation of (5) for a fixed τ in isolation cannot yield a consistent estimate of θ since it ignores the global dependence of the σ_{t-k}'s on the entire function $\theta(\cdot)$. If the dependence structure of u_t is characterized by (1) and (2), we can consider the following restricted quantile regression[1] instead of (5):

$$\left(\widehat{\pi}, \widehat{\theta}\right) = \begin{cases} \arg\min_{\pi,\theta} \sum_i \sum_t \rho_{\tau_i}(u_t - \pi_i^\top z_t(\theta)) \\ \text{s.t. } \pi_i = \theta(\tau_i) = \theta F^{-1}(\tau_i). \end{cases}$$

[1]If we look at the CAViaR(p,q) representation of the GARCH model (2), the parameters β_i^*, $(i = 1, \cdots, p)$ are global and are not dependent on the specific quantile; however, parameters β_0^* and γ_j^*, $(j = 1, \cdots, q)$ are local in the sense that they vary over different τ. In this sense, the linear GARCH model (1) has a CAViaR(p, q) representation: $Q_{u_t}(\tau_k|\mathcal{F}_{t-1}) = \beta_{0k}^* + \sum_{i=1}^{p} \beta_i^* Q_{u_{t-i}}(\tau_k|\mathcal{F}_{t-i-1}) + \sum_{j=1}^{q} \gamma_{jk}^* |u_{t-j}|$, which satisfies the global restriction: $\beta_{0k}^* = \beta_0 F^{-1}(\tau_k)$, $\gamma_{jk}^* = \gamma_j F^{-1}(\tau_k)$, $j = 1, \cdots, q$, for $k = 1, \cdots, K$.

Estimation of this global restricted nonlinear quantile regression is complicated both computationally and theoretically. In this paper, we propose a simpler two-stage estimator that both incorporates the global restrictions and also focuses on the local approximation around the specified quantile. The proposed procedure is easily implemented, and asymptotic theory as well as Monte Carlo evidence indicates that the proposed estimator has good performance compared to conventionally used methods in estimating conditional quantiles based on parametric GARCH models.

2.3. **A Two-Step Estimator for Conditional Quantiles.** In this section, we describe our two-step estimator for conditional quantiles of the linear GARCH model. The proposed estimation consists the following two steps: (i) We consider a global estimation in the first step to incorporate the global dependence of the latent σ_{t-k}'s on θ. (ii) Then, using results from the first step, we focus on the specified quantile to find the best local estimate for the conditional quantile.

In general, different estimation methods may be used in the first step - see additional discussions on related issues in Section 3.4. We focus our discussion on the following quantile autoregression based approach primarily due to its simplicity and its effectiveness as a preliminary estimator.

We propose the following estimation procedure: In the first stage unrestricted estimates of several quantile autoregressions are combined via minimum distance methods to construct global estimates of the conditional scale parameters; in the second stage local estimates of the conditional quantiles are computed based on the local scale estimates.

Giving the GARCH model (1) and (2), let

$$A(L) = 1 - \beta_1 L - \cdots - \beta_p L^p, \quad B(L) = \gamma_1 + \cdots + \gamma_q L^{q-1},$$

under regularity assumptions presented in Section 3 ensuring the invertibility of $A(L)$, we obtain an ARCH(∞) representation for σ_t :

$$\sigma_t = a_0 + \sum_{j=1}^{\infty} a_j \left| u_{t-j} \right|, \tag{6}$$

where the coefficients a_j satisfy summability conditions implied by the regularity conditions. For identification, we normalize $a_0 = 1$. Substituting the above ARCH(∞) representation into (1) and (2), we have

$$u_t = \left(a_0 + \sum_{j=1}^{\infty} a_j \left| u_{t-j} \right| \right) \varepsilon_t, \tag{7}$$

and

$$Q_{u_t}(\tau | \mathcal{F}_{t-1}) = \alpha_0(\tau) + \sum_{j=1}^{\infty} \alpha_j(\tau) \left| u_{t-j} \right|,$$

where $\alpha_j(\tau) = a_j Q_{\varepsilon_t}(\tau), \quad j = 0, 1, 2, \cdots$.

Under our regularity conditions the coefficients a_j decrease geometrically, so letting $m = m(n)$ denote a truncation parameter we may consider the following truncated quantile autoregression:

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) \approx a_0(\tau) + a_1(\tau)\left|u_{t-1}\right| + \cdots + a_m(\tau)\left|u_{t-m}\right|.$$

See Koenker and Xiao (2006) for a discussion of this class of autoregressive models. By choosing m suitably small relative to the sample size n, but large enough to avoid serious bias, we obtain a sieve approximation for the GARCH model.

One could estimate the conditional quantiles simply using a sieve approximation:

$$\check{Q}_{u_t}(\tau|\mathcal{F}_{t-1}) = \widehat{a}_0(\tau) + \widehat{a}_1(\tau)\left|u_{t-1}\right| + \cdots + \widehat{a}_m(\tau)\left|u_{t-m}\right|,$$

where $\widehat{a}_j(\tau)$ are the quantile autoregression estimates. Under the assumptions of Section 3, we have

$$\check{Q}_{u_t}(\tau|\mathcal{F}_{t-1}) = Q_{u_t}(\tau|\mathcal{F}_{t-1}) + O_p(m/\sqrt{n}).$$

However, as shown in the Monte Carlo experiment, this simple sieve approximation provides a a rather noisey estimator for the GARCH coefficients, but it serves as an adequate preliminary estimator.

Since our first step estimation focuses on the global model, it is desirable to use information over multiple quantiles in estimation. Combining information over multiple quantiles also helps us to obtain globally coherent estimate of the scale parameters. In order to use information over multiple quantiles, we estimate the unrestricted model at various quantiles and assemble independent quantile estimates. (Alternatively, one could try to estimate the first step model over several quantiles jointly). In this paper, we combine information at different quantiles via minimum distance estimation.

Suppose that we estimate the m-th order quantile autoregression

$$\widetilde{\alpha}(\tau) = \arg\min_{\alpha} \sum_{t=m+1}^{n} \rho_\tau\left(u_t - \alpha_0 - \sum_{j=1}^{m} \alpha_j\left|u_{t-j}\right|\right) \tag{8}$$

at quantiles $(\tau_1, \cdots, \tau_K)$, and obtain estimates

$$\widetilde{\alpha}(\tau_k), k = 1, \cdots, K.$$

Let $\widetilde{a}_0 = 1$ in accordance with the identification assumption. Denote

$$\mathbf{a} = \left[a_1, \cdots, a_m, q_1, \cdots, q_K\right]^\top, \tilde{\pi} = \left[\widetilde{\alpha}(\tau_1)^\top, \cdots, \widetilde{\alpha}(\tau_K)^\top\right]^\top,$$

where $q_k = Q_{\varepsilon_t}(\tau_k)$, and

$$\phi(\mathbf{a}) = g \otimes \alpha = \left[\begin{array}{ccc} q_1, a_1q_1, \cdots, a_mq_1, & \cdots, & q_K, a_1q_K, \cdots, a_mq_K \end{array}\right]^\top,$$

where $g = [q_1, \cdots, q_K]^\top$ and $\alpha = [1, a_1, a_2, \cdots, a_m]^\top$, we consider the following estimator for the vector $\mathbf{a}$ that combines information over the K quantile estimates based on the restrictions $\alpha_j(\tau) = a_j Q_{\varepsilon_t}(\tau)$:

$$\widetilde{\mathbf{a}} = \arg\min_{\mathbf{a}} \left(\tilde{\pi} - \phi(\mathbf{a})\right)^\top A_n \left(\tilde{\pi} - \phi(\mathbf{a})\right), \tag{9}$$

where A_n is a $(K(m+1)) \times (K(m+1))$ positive definite matrix. To summarize: We propose the following two-step estimator for the conditional quantiles of u_t:

Step 1: Estimate the following m-th order quantile autoregression (8) at quantiles $(\tau_1, \cdots, \tau_K)$, and obtain $\widetilde{\alpha}(\tau_k)$, $k = 1, \cdots, K$. By setting $\widetilde{a}_0 = 1$ and solving the minimum distance estimation problem (9), we obtain an estimator for $(a_0, \cdots, a_m)$, denoting it as $(\widetilde{a}_0, \cdots, \widetilde{a}_m)$. Thus σ_t can be estimated by

$$\widetilde{\sigma}_t = \widetilde{a}_0 + \sum_{j=1}^{m} \widetilde{a}_j \left| u_{t-j} \right| .$$

Step 2: Quantile regression of u_t on $\widetilde{z}_t = (1, \widetilde{\sigma}_{t-1}, \cdots, \widetilde{\sigma}_{t-p}, \left| u_{t-1} \right|, \cdots, \left| u_{t-q} \right|)^\top$ by

$$\min_{\theta} \sum_t \rho_\tau (u_t - \theta^\top \widetilde{z}_t), \tag{10}$$

the two-step estimator of $\theta(\tau)^\top = (\beta_0(\tau), \beta_1(\tau), \cdots, \beta_p(\tau), \gamma_1(\tau), \cdots, \gamma_q(\tau))$ is then given by solution of (10), $\widehat{\theta}(\tau)$, and the τ-th conditional quantile of u_t can be estimated by

$$\widehat{Q}_{u_t}(\tau|\mathcal{F}_{t-1}) = \widehat{\theta}(\tau)^\top \widetilde{z}_t.$$

Iteration can be applied to the above procedure for further improvement

In the next section, we develop the asymptotic theory of the related estimators.

3. Asymptotic Properties of The Proposed Estimator

This section investigates the asymptotic behavior of the proposed estimators, including the sieve quantile autoregression, the minimum distance estimation and the second stage estimation with generated regressors.

3.1. **A Quantile Autoregression Sieve Approximation.** In this subsection, we study a quantile autoregression approximation for our underlying linear GARCH model. The nature of the sieve approximation used in the first stage of the procedure plays a crucial role in the proposed estimator. There is an extensive literature on the asymptotic behavior of regression estimators with increasing parametric dimensions. Huber (1973) first considered M-estimation of linear regression with continuously differentiable ρ (objective) function, and showed that asymptotic normality can be preserved if $m^3/n \to 0$ as $n \to \infty$. Several subsequent researchers successfully improved on Huber's results, including Portnoy (1985), Mammen (1989), Welsh (1989), and Bai and Wu (1994). Welsh (1989) and He and Shao (2000) studied nonlinear M-estimation with increasing dimension and an objective function with possible nondifferentiability at finitely many points.

The focus of most prior studies is to determine the best possible expansion rate for the number of parameters m as a function of the sample size n, and generally assumed independent observations. Our objectives are somewhat different. Rather than trying to determine the best rate for the truncation parameter m, our focus will be estimation of conditional quantiles in the second step and the sieve regression is only a preliminary step. In fact, as will become clear later in our analysis, under Assumption S1, the error coming from an m-th order truncation is of order $O_p(b^m)$ ($b < 1$) and the approximation error of $\widetilde{\sigma}_t$ is of order $O_p(\sqrt{m/n})$, it would suffice to consider a truncation m as a sufficiently large constant multiple of $\log(n)$. In addition, we consider time dependent data, and treat truncation as an approximation, assuming that the true quantile function is an infinite summation. In prior literature there is typically a sequence of true models with increasing parametric dimension.

For convenience of the asymptotic analysis, we make the following assumptions. We again stress that we are not seeking to achieve the weakest possible regularity conditions for the asymptotic analysis, but instead to focus on the design of a robust, flexible and easy-to-implement procedure for estimation of the GARCH model.

Assumption S1. *The polynomials $A(L)$ and $B(L)$ have no common factors, $A(z) \neq 0$, for $|z| \leq 1$; and $B(z) \neq 0$, for $|z| \leq 1$.*

Assumption S2. *$\{\varepsilon_t\}$ are iid random variables with mean 0 and variance $\sigma^2 < \infty$. The distribution function of ε_t, F_ε, has a continuous density f_ε with $0 < f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right) < \infty$.*

Assumption S3. *Denote the conditional distribution function $\Pr[u_t < \cdot|x_t]$ as $F_{u|x}(\cdot)$ and its derivative as $f_{u|x}(\cdot)$* is continuously differentiable and $0 < f_{u|x}(\cdot) < \infty$ *on it's support.*

Assumption S4: *Let $x_t = (1, |u_{t-1}|, \cdots, |u_{t-m}|)^\top$, and*

$$\overline{D}_n = -\mathrm{E}\left(\frac{1}{n}\sum_{t=m+1}^{n}\frac{x_t x_t^\top}{\sigma_t}\right),$$

and denote the maximum and minimum eigenvalues of $\overline{D}_n$ as $\lambda_{\max}\left(\overline{D}_n\right)$ and $\lambda_{\min}\left(\overline{D}_n\right)$ then

$$\lim\inf_{n\to\infty}\lambda_{\min}\left(\overline{D}_n\right) > 0, \lim\sup_{n\to\infty}\lambda_{\max}\left(\overline{D}_n\right) < \infty.$$

Assumption S5: *There exist (small) positive constants $\delta_1 > 0$ and $\delta_2 > 0$ such that*

$$\Pr\left(\max_{1\leq t\leq n} u_t^2 > n^{\delta_1}\right) \leq \exp(-n^{\delta_2}).$$

Assumption S6: *The truncation parameter m satisfies $m(n) = c\log n$ for some $c > 0$.*

Assumptions S1 and S2 are standard assumptions in the GARCH literature. Assumption S1 is an invertibility condition on the ARCH operator and ensures that u_t is stationary with weak dependence and that appropriate limiting theory can be applied. This condition is useful in our technical development and, no doubt could be weakened, but we do not attempt to do so, or to find minimal conditions under which our results hold. The variance of ε_t is usually standardized to be 1, but we assume that ε_t has variance σ^2 in Assumption S2 because we prefer the slightly different standardization that the first coefficient in the ARCH(∞) representation (7) is 1 ($a_0 = 1$). Assumptions S3 and S4 are similar to those in the previous literature on sieve estimation. Assumption S5 assumes that the maximum of u_t^2 has a generalized extreme value distribution. This is is a higher level assumption and generally holds under weak dependence assumptions. The expansion rate of the truncation parameter given in Assumption S6 is also for convenience and similar results can be expected to hold for a much wider range of m.

Under Assumption S1, $A(L)$ is invertible and we have an ARCH(∞) representation (7) for σ_t, where the coefficients a_j decrease at a geometric rate, i.e. there exits positive constants

$b<1$ and c such that $|a_j| \leq cb^j$. Consequently,

$$\sigma_t = a_0 + a_1 |u_{t-1}| + \cdots + a_m |u_{t-m}| + O_p(b^m).$$

Denoting $\alpha(\tau) = (a_0(\tau), a_1(\tau), \cdots, a_m(\tau))^\top \in \mathcal{R}^{m+1}$, and $x_t = (1, |u_{t-1}|, \cdots, |u_{t-m}|)^\top$, the sieve quantile regression can be written as

$$\widetilde{\alpha}(\tau) = \arg\min_{\alpha} \sum_{t=m+1}^{n} \rho_\tau(u_t - \alpha^\top x_t). \tag{11}$$

The following result establishes consistency and asymptotic normality for the sieve estimator.

Theorem 1. *Let $\widetilde{\alpha}(\tau)$ be the solution of (11), then under Assumptions S1 - S6, we have*

$$\|\widetilde{\alpha}(\tau) - \alpha(\tau)\|^2 = O_p(m/n). \tag{12}$$

and

$$\sqrt{n}\left(\widetilde{\alpha}(\tau) - \alpha(\tau)\right) = -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)} D_n^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=m+1}^{n} x_t \psi_\tau(u_{t\tau})\right) + o_p(1) \tag{13}$$

where

$$\psi_\tau(u) = \tau - I(u<0), and D_n = \left[\frac{1}{n} \sum_{t=m+1}^{n} \frac{x_t x_t^\top}{\sigma_t}\right].$$

For any $\lambda \in \mathcal{R}^{m+1}$,

$$\frac{\sqrt{n}\lambda^\top\left(\widetilde{\alpha}(\tau) - \alpha(\tau)\right)}{\sigma_\lambda} \Rightarrow N(0,1),$$

where $\sigma_\lambda^2 = \frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)^2} \lambda^\top D_n^{-1} \Sigma_n(\tau) D_n^{-1} \lambda$, and $\Sigma_n(\tau) = \frac{1}{n}\sum_{t=m+1}^{n} x_t x_t^\top \psi_\tau^2(u_{t\tau})$.

3.2. **Minimum Distance Estimation of Conditional Scale.** Having estimated the truncated quantile autoregressions on a grid of τ's, we would now like to combine these estimates to obtain estimates of the conditional scale parameters, σ_t. This is accomplished most easily using the minimum distance methods proposed in Section 2.3. The asymptotic properties of this estimator are summarized in the following Theorem.

Theorem 2. *Under assumptions S1 - S6, the minimum distance estimator $\widetilde{\mathbf{a}}$ solving (9) has the following asymptotic representation:*

$$\sqrt{n}(\widehat{\mathbf{a}} - \mathbf{a}_0) = -\left[G^\top A_n G\right]^{-1} G^\top A_n \left[\frac{1}{\sqrt{n}} \sum_{t=m+1}^{n} \Upsilon_{Kt} \otimes \left[D_n^{-1} x_t\right]\right] + o_p(1)$$

where

$$G = \left[g_0 \otimes J_m \vdots I_K \otimes \alpha_0\right], \Upsilon_{Kt} = \begin{bmatrix} \frac{\psi_{\tau_1}(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)} \\ \cdots \\ \frac{\psi_{\tau_m}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \end{bmatrix}, g_0 = \begin{bmatrix} Q_{\varepsilon_t}(\tau_1) \\ \cdots \\ Q_{\varepsilon_t}(\tau_K) \end{bmatrix},$$

where g_0 and α_0 are the true values of vectors $g = [q_1, \cdots, q_K]^\top$ and $\alpha = [1, a_1, a_2, \cdots, a_m]^\top$, and

$$J_m = \begin{bmatrix} 0 & \cdots & 0 \\ 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$

is an $(m+1) \times m$ matrix and I_K is an K-dimensional identity matrix.

When A_n is an identity matrix,

$$G^\top A_n G = G^\top G = \begin{bmatrix} \begin{bmatrix} g_0^\top g_0 \otimes J_m^\top J_m \\ I_K^\top g_0 \otimes \alpha_0^\top J_m \end{bmatrix} & \begin{bmatrix} g_0^\top I_K \otimes J_m^\top \alpha_0 \\ I_K^\top I_K \otimes \alpha_0^\top \alpha_0 \end{bmatrix} \end{bmatrix}.$$

Alternatively, setting $\mathcal{D} = I_K \otimes D_n$, $V_{xt} = x_t x_t^\top$, $V_{\psi t} = \Upsilon_{Kt} \Upsilon_{Kt}^\top$, and

$$\Psi_K = \frac{1}{n} \sum_{t=m+1}^{n} V_{\psi t} \otimes V_{xt}, \Lambda = \mathcal{D}^{-1} \Psi_K \mathcal{D}^{-1},$$

the optimal choice of A is given by

$$A = \Lambda^{-1} = \mathcal{D} \Psi_K^{-1} \mathcal{D}.$$

In this case, $G^\top A_n G = G^\top \mathcal{D} \Psi_K^{-1} \mathcal{D} G$, and $\mathcal{D}G = \left[(I_K g_0 \otimes D_n J_m) \vdots (I_{K^2} \otimes D_n \alpha_0) \right]$.

The first stage estimation immediately delivers an estimator for the conditional variance:

$$\widetilde{\sigma}_t = \widetilde{a}_0 + \sum_{j=1}^{m} \widetilde{a}_j \left| u_{t-j} \right|.$$

For convenience of later analysis, we partition the $(K(m+1)) \times (K(m+1))$ weighting matrix A_n as $[A_{n1}, \cdots, A_{nK}]$, where A_{nk} $(k = 1, \cdots, K)$ are $(K(m+1)) \times (m+1)$ sub-matricies. Let $\underline{x}_t^\top = (|u_{t-1}|, \cdots, |u_{t-m}|)$,

$$L_{m/K} = [I_m, 0_{m \times K}], H_j = L_{m/K} \left[G^\top A_n G \right]^{-1} G^\top \left[A_{n1} D_n^{-1} x_j, \cdots, A_{nK} D_n^{-1} x_j \right],$$

and denote $H = n^{-1} \sum_j H_j \Delta H_j$, where

$$\Delta = \begin{bmatrix} \frac{\tau_1(1-\tau_1)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)^2} & \cdots & \frac{\tau_1 \wedge \tau_K - \tau_1 \tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right) f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\ \cdots & \ddots & \cdots \\ \frac{\tau_1 \wedge \tau_K - \tau_1 \tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right) f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \cdots & \frac{\tau_K(1-\tau_K)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)^2} \end{bmatrix},$$

the asymptotic behavior of this preliminary estimator for the scale parameter is given below.

Corollary 1. *Under assumptions S1-S6, let $\omega_t^2 = \underline{x}_t^\top H \underline{x}_t$, as $n \to \infty$*

$$\frac{\sqrt{n} \left(\widetilde{\sigma}_t - \sigma_t \right)}{\omega_t} \Rightarrow N(0,1).$$

In contrast to most existing estimators of conditional volatility based on Gaussian distributional assumptions, our volatility estimator has the nice property that is relatively robust to assumptions on the error distribution.

3.3. **Asymptotic Distribution of the Second Stage Estimator.** Using the results from the first stage estimation, the second step local estimator $\theta(\tau)$ can be obtained by quantile regression of u_t on $\widetilde{z}_t = (1, \widetilde{\sigma}_{t-1}, \cdots, \widetilde{\sigma}_{t-p}, |u_{t-1}|, \cdots, |u_{t-q}|)^\top$, and the τ-th conditional quantile of u_t can be estimated by

$$\widehat{Q}_{u_t}(\tau|\mathcal{F}_{t-1}) = \widehat{\theta}(\tau)^\top \widetilde{z}_t.$$

The limiting behavior of the second-stage estimator minimizing (10) is described in the following result.

Theorem 3. *Under assumptions S1-S6, the two-step estimator $\widehat{\theta}(\tau)$ based on (10) has the asymptotic representation:*

$$\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) = -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1}\left\{\frac{1}{\sqrt{n}}\sum_t z_t\psi_\tau(u_{t\tau})\right\} + \Omega^{-1}\Gamma\sqrt{n}\left(\widetilde{a} - a\right) + o_p(1)$$

where $a = [a_1, a_2, \cdots, a_m]^\top$, $\Omega = E\left[z_t z_t{}^\top/\sigma_t\right]$, and

$$\Gamma = \sum_{k=1}^{p}\theta_k C_k, C_k = E\left[\left(|u_{t-k-1}|, \cdots, |u_{t-k-m}|\right)\frac{z_t}{\sigma_t}\right].$$

In particular, since the first stage estimation is based on (9) the above asymptotic representation can be rewritten, denoting $L_{m/K} = \left[I_m \vdots 0_{m\times K}\right]$, as,

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) \;\; = \;\; & -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1}\left\{\frac{1}{\sqrt{n}}\sum_t z_t\psi_\tau(u_{t\tau})\right\} \\ & -\Omega^{-1}\Gamma L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_n\left[\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n}\Upsilon_{Kt}\otimes\left[D_n^{-1}x_t\right]\right] \\ & +o_p(1).\end{aligned}$$

The asymptotic distribution of the two-step estimator $\widehat{\theta}(\tau)$ can be immediately obtained from the above Theorem. Let

$$\begin{aligned}\Psi_t \;\; &= \;\; \left[\begin{array}{cccc}\frac{\psi_\tau(u_{t\tau})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}, & \frac{\psi_{\tau_1}(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)}, & \cdots, & \frac{\psi_{\tau_K}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)}\end{array}\right]^\top, \\ M_t \;\; &= \;\; \left[\begin{array}{cccc} z_t, & \Gamma L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_{n1} D_n^{-1} x_t, & \cdots, & \Gamma L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_{nK} D_n^{-1} x_t\end{array}\right],\end{aligned}$$

and define

$$M = \operatorname*{p\lim}_n\left[\frac{1}{n}\sum_t M_t \Xi M_t^\top\right],$$

where

$$\Xi=\begin{bmatrix}\frac{\tau(1-\tau)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)^2} & \frac{\tau\wedge\tau_1-\tau\tau_1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)} & \cdots & \frac{\tau\wedge\tau_K-\tau\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\ \frac{\tau\wedge\tau_1-\tau\tau_1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)} & \frac{\tau_1(1-\tau_1)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)^2} & \cdots & \frac{\tau_1\wedge\tau_K-\tau_1\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\ \cdots & \cdots & \ddots & \cdots \\ \frac{\tau\wedge\tau_K-\tau\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \frac{\tau_1\wedge\tau_K-\tau_1\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \cdots & \frac{\tau_K(1-\tau_K)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)^2}\end{bmatrix}.$$

The limiting distribution of the two stage estimator is summarized in the following corollary.

Corollary 2. *Under assumptions S1-S6, the two-step estimator $\widehat{\theta}(\tau)$ has the following limiting distribution:*

$$\sqrt{n}\left(\widehat{\theta}(\tau)-\theta(\tau)\right)\Rightarrow N\left(0,\Omega^{-1}M\Omega^{-1}\right),\ \textit{as } n\to\infty.$$

In the simple case that we estimate the first stage model at a single quantile τ, let $\widetilde{\alpha}(\tau)=(\widetilde{\alpha}_0(\tau),\cdots,\widetilde{\alpha}_m(\tau))^\top$, by setting $\widetilde{a}_0=1$ and solving the equations $\widetilde{\alpha}_j(\tau)=\widetilde{a}_j\widetilde{Q}_{\varepsilon_t}(\tau)$, we obtain the following estimator for $(a_0,\cdots,a_m)$:

$$\widetilde{a}_0=1,\widetilde{a}_1=\frac{\widetilde{\alpha}_1(\tau)}{\widetilde{\alpha}_0(\tau)},\cdots,\widetilde{a}_m=\frac{\widetilde{\alpha}_m(\tau)}{\widetilde{\alpha}_0(\tau)},$$

In this case, the estimator

$$\widetilde{\sigma}_t=\widetilde{a}_0+\sum_{j=1}^m\widetilde{a}_j\left|u_{t-j}\right|,$$

in Step 1 has the following representation:

$$\widetilde{\sigma}_t=\sigma_t+\frac{1}{\alpha_0(\tau)}\left[\widetilde{\alpha}(\tau)-\alpha(\tau)\right]^\top\breve{x}_t+O_p\left(\frac{m^2}{n}\right)=\sigma_t+O_p\left(\sqrt{\frac{m}{n}}\right)+O_p\left(\frac{m^2}{n}\right),$$

where

$$\breve{x}_t=\left(-\frac{\sum_{j=1}^m\alpha_j(\tau)}{\alpha_0(\tau)},\left|u_{t-1}\right|,\cdots,\left|u_{t-m}\right|\right),$$

and the two-stage estimator has the following simplified asymptotic representation.

Corollary 3. *Under our assumptions S1 - S6, if we estimate the first stage model at same single quantile τ, the second stage quantile regression estimator $\widehat{\theta}(\tau)$ based on (10) has the following Bahadur representation:*

$$\sqrt{n}\left(\widehat{\theta}(\tau)-\theta(\tau)\right)=-\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1}\left\{\frac{1}{\sqrt{n}}\sum_t\left[z_t+R^\top D_n^{-1}x_t\right]\psi_\tau(u_{t\tau})\right\}+o_p(1)$$

where $R^\top=\frac{1}{\alpha_0(\tau)}\left(-\sum_{j=1}^m\frac{\alpha_j(\tau)}{\alpha_0(\tau)}r_j,r_1,\cdots,r_m\right)$, and $r_j=\sum_{k=1}^p\theta_kE\left[\left|u_{t-k-j}\right|\frac{z_t}{\sigma_t}\right]$. Consequently,

$$\sqrt{n}\left(\widehat{\theta}(\tau)-\theta(\tau)\right)\Rightarrow N\left(0,\frac{\tau(1-\tau)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)^2}\Omega^{-1}\underline{M}\Omega^{-1}\right),$$

where $\underline{M} = \underline{M}_1 + \underline{M}_2 + \underline{M}_3$, *with* $\underline{M}_1 = E\left[z_t z_t^\top\right]$, $\underline{M}_2 = \lim \frac{1}{n}\sum_t \left[R^\top D_n^{-1} x_t z_t^\top + z_t x_t^\top D_n^{-1} R\right]$, *and* $\underline{M}_3 = \lim \frac{1}{n}\sum_t R^\top D_n^{-1} x_t x_t^\top D_n^{-1} R$.

Remark. We may compare the quantile regression estimator $\widehat{\theta}(\tau)$ based on generated regressors $\widetilde{z}_t$ with the infeasible quantile regression estimator $\widetilde{\theta}(\tau)$ based on unobserved regressors z_t. Note that the infeasible estimator $\widetilde{\theta}(\tau)$ has the following Bahadur representation:

$$\sqrt{n}\left(\widetilde{\theta}(\tau) - \theta(\tau)\right) = -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1}\left\{\frac{1}{\sqrt{n}}\sum_t z_t \psi_\tau(u_{t\tau})\right\} + o_p(1),$$

and

$$\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) \Rightarrow N\left(0, \frac{\tau(1-\tau)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)^2}\Omega^{-1}\underline{M}_1\Omega^{-1}\right).$$

Comparing it with the Bahadur representation of $\widehat{\theta}(\tau)$ given in Corollary 1, we see that the Bahadur representation (and thus the variance) of $\widehat{\theta}(\tau)$ contains an additional term that arises from the preliminary estimation. ■

The proposed estimation procedure in this paper can be extended in several different directions. First, like many other nonlinear estimation procedures, the propsed estimation procedure may be iterated to achive further improvement. From the two step estimation, we obtain estimates of $\theta(\tau)$ and $Q_{u_t}(\tau|\mathcal{I}_{t-1})$ at different quantiles. Consequently, estimates of θ and $F^{-1}(\tau)$ can be derived immediately, and updated estimates of σ_t can also be obtained. The updated estimator of σ_t can then be used to re-estimate $\theta(\tau)$ and $Q_{u_t}(\tau|\mathcal{I}_{t-1})$. The above procedure can be iterated to obtain estimators for both the conditional quantiles and the conditional volatility σ_t. Second, different estimation methods may be used in the first step global estimation. The current paper considers QAR estimation in the first stage due to its convenience in implementation and effectiveness for a wide range of time series. We conjecture that when the process are nearly integrated, a different first step estimation methods may be preferred since the autoregression representation is obtained for invertible ARMA models. Third, the basic idea of the two-step method can be applied to other types of GARCH processes.

4. Monte Carlo Results

In this section, we report on a Monte Carlo experiment designed to examine the sampling performance of the proposed estimation procedures and compare them with existing methods. In particular, we compare the proposed quantile regression GARCH estimation procedures with the simple quantile autoregression approximation; the RiskMetrics method that is widely used in industry; and the CAViaR model proposed by Engle and Manganelli (2004). The CAViaR model is estimated using the Matlab code of Manganelli (2002).

As measures of performance we report bias and mean square error (MSE) of the various estimators of the 0.05 conditional quantile of the response *as averaged over the sample.* For comparison purpose, we consider the following eight estimation procedures:

(1) RiskM: The conventional RiskMetrics method RiskMetrics Group (1996), based on Gaussian GARCH(1,1) with fixed parameters, that is widely used in financial applications for estimation of Value-at-Risk;
(2) GGARCH: The Gaussian GARCH(1,1) with estimated parameters.
(3) ARCH: Sieve ARCH quantile regression approximation with $m = 3n^{1/4}$.
(4) QGARCH1: The proposed two-step estimation method using information at the specified quantile in the first step estimation.
(5) QGARCH2: The proposed two-step estimation method using information over multiple quantiles in the first step estimation. In particular, we estimate the sieve ARCH quantile regression at each *percentile* ($\tau_k = 5k\%,\ k = 1, \cdots, 19.$), and estimate the GARCH parameters using the Minimum distance estimation ($A_n = I$) coupled with trimming to avoid the random denominator going to zero.
(6) QGARCH3: The proposed estimation method using information at the specified quantile in the first step estimation and iterate for potential improvements. Thus, following Step 1 in our procedure, we estimate a sieve quantile autoregression and obtain estimates of σ_t, then we run quantile regression of u_t based on the estimated regressors and obtain the two-step estimator of $\theta(\tau)^\top = (\beta_0(\tau), \beta_1(\tau), \gamma_1(\tau))$. Estimates of parameters of the GARCH model can then be derived from the quantile regression estimates by solving

$$\frac{\beta_1(\tau)}{\beta_0(\tau)} = \frac{\beta_1}{\beta_0}, \frac{\gamma_1(\tau)}{\beta_0(\tau)} = \frac{\gamma_1}{\beta_0}, \frac{\beta_0}{1-\beta_1} = 1.$$

Finally, we recompute the estimates of σ_t and iterate the process to convergence.
(7) CAViaR1: CAViaR estimator using the Matlab code of Manganelli (2002), the number of grid points is chosen to be n (=sample size).
(8) CAViaR2: CAViaR estimator using the Matlab code of Manganelli (2002), the number of grid points is chosen to be 10000.

The data were generated from a linear GARCH(1,1) process with several choices of parameter values and error distributions. Two different choices for the distribution of ε_t are considered: (i) i.i.d Normal; (ii) i.i.d. $t(4)$ - Student-t distribution with 4 degrees of freedom; The first design of ε_t actually has normal distribution and we expect the traditional methods based on normal assumption should be reasonable. The second design of ε_t has a heavier tail. Two sample sizes $n = 100$, $n = 500$, are examined in the simulation, and number of repetitions is 50. In each instance we estimate the 0.05 quantile. We consider the following three sets of parameter values:

P1. $\beta_0 = 0.1, \beta_1 = 0.5, \gamma_1 = 0.3.$
P2. $\beta_0 = 0.1, \beta_1 = 0.8, \gamma_1 = 0.1$
P3. $\beta_0 = 0.1, \beta_1 = 0.9, \gamma_1 = 0.05.$

The first set of parameter values (P1) satisfies the regularity conditions and the generates a stationary linear GARCH process. Table 1 reports result of bias and mean squared error of different estimation procedures for this case. The Monte Carlo results in Table 1 provide some baseline evidence in evaluating the sampling performance of the proposed method when the regularity conditions are satisfied. In addition to (P1), we also consider parameter values that are close to nonstationary GARCH, and examine the performance of

	Normal				Student with 4 df.			
	n=100		n=500		n=100		n=500	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
RiskM	0.1522	0.0382	0.1362	0.0305	0.6479	0.6884	0.8299	1.1840
GGARCH	0.2653	0.1433	0.1315	0.0301	0.7961	0.9813	0.5092	0.3968
ARCH	0.2545	0.1119	0.1348	0.0308	0.5140	0.5834	0.2985	0.2418
CAViaR2	0.1693	0.0549	0.1310	0.0282	0.3582	0.2883	0.2208	0.1390
CAViaR5	0.1647	0.0518	0.1306	0.0283	0.3311	0.2576	0.2208	0.1390
QGARCH1	0.1278	0.0286	0.0685	0.0083	0.3038	0.1917	0.1687	0.0757
QGARCH2	0.1257	0.0267	0.0715	0.0087	0.3116	0.2096	0.1643	0.0865
QGARCH3	0.1003	0.0185	0.0576	0.0064	0.2954	0.2041	0.1248	0.0477

TABLE 1. Bias and MSE for Estimates at $\tau = 0.05$, $\beta_0 = 0.1, \beta_1 = 0.5, \gamma_1 = 0.3$.

	Normal				Student with 4 df.			
	n=100		n=500		n=100		n=500	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
RiskM	0.2653	0.1433	0.1315	0.0301	0.7961	0.9813	0.5092	0.3968
GGARCH	0.3557	0.2957	0.2323	0.0959	1.6455	3.5759	1.8645	4.1445
ARCH	0.5395	0.4834	0.3014	0.1497	1.1976	4.2614	0.5966	0.6739
CAViaR2	0.2643	0.1239	0.1262	0.0275	0.6441	0.8417	0.2417	0.1078
CAViaR5	0.2290	0.0883	0.1213	0.0256	0.6260	0.7680	0.2388	0.1045
QGARCH1	0.2570	0.1143	0.1405	0.0318	0.6233	0.7460	0.3101	0.2002
QGARCH2	0.2528	0.1091	0.1546	0.0400	0.6197	0.7283	0.3427	0.2093
QGARCH3	0.2266	0.0844	0.1236	0.0265	0.5235	0.5404	0.2749	0.1352

TABLE 2. Bias and MSE for Estimates at $\tau = 0.05$, $\beta_0 = 0.1, \beta_1 = 0.8, \gamma_1 = 0.1$.

the estimation procedures in this situation. In the second and third sets (P2 and especially P3) of parameter values, β_1 are large and $\beta_1 + \gamma_1$ are close to 1. When $\beta_1 + \gamma_1 = 1$, the process becomes nonstationary and the regularity assumption S1 no longer holds. When $\beta_1 + \gamma_1$ is close to 1, the process becomes nearly integrated and the ARCH approximation used in the first step becomes poor. Monte Carlo results confirm this. In particular, Table 2 gives results corresponding to $\beta_1 = 0.8, \gamma_1 = 0.1$. Table 3 corresponds to the case $\beta_1 = 0.9, \gamma_1 = 0.05$.

The Monte Carlo results indicate that in general, the proposed GARCH quantile estimator has reasonably good performance for a wide range of time series. They generally have better performance over other estimation procedures in the stationary case. As the data becomes more nonstationary, the performance of all these estimation procedures decreases. In table 2, the proposed quantile regression GARCH estimation procedures still have relatively better performance in general, but the difference between the CAViaR and GARCH quantile estimation becomes smaller. In table 3, when the data are generated from a nearly integrated GARCH process with $\beta_1 = 0.9, \gamma_1 = 0.05$, the CAViaR model has relatively better performance than the two-step estimators.

	Normal				Student with 4 df.			
	n=100		n=500		n=100		n=500	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
RiskM	0.3557	0.2957	0.2323	0.0959	1.6455	3.5759	1.8645	4.1445
GGARCH	0.3557	0.2957	0.2323	0.0959	1.6455	3.5759	1.8645	4.1445
ARCH	1.0650	1.8698	0.5960	0.5856	2.1401	9.4193	1.2103	2.7186
CAViaR1	0.5181	0.4475	0.2464	0.1037	1.1364	2.3682	0.4788	0.4044
CAViaR2	0.4995	0.4222	0.2149	0.0785	1.1704	2.5111	0.4858	0.4193
QGARCH1	0.4387	0.3233	0.2608	0.1084	1.0568	2.0841	0.6293	0.6849
QGARCH2	0.4482	0.3462	0.2651	0.1154	1.0035	1.8752	0.5924	0.6046
QGARCH3	0.4623	0.3657	0.2466	0.1087	0.9436	1.7429	0.5506	0.5300

Table 3. Bias and MSE for Estimates at $\tau = 0.05, \beta_0 = 0.1, \beta_1 = 0.9, \gamma_1 = 0.05$.

	S&P 500	Nikkei 225	FTSE 100	Hang Seng
Mean	0.00166	0.0010	0.0017	0.0023
Std. Dev.	0.0201	0.0240	0.0240	0.0369
Max	0.1002	0.1213	0.1389	0.1547
Min	-0.1566	-0.1322	-0.2486	-0.4061
Skewness	-0.4676	-0.3125	-1.8523	-1.7155
Kurtosis	3.3561	3.2958	16.198	11.376
AC(1)	-0.0709	-0.0211	-0.0219	0.0899
AC(2)	0.0501	0.0527	0.0928	0.0776
AC(3)	-0.0211	0.0333	-0.0511	-0.0241
AC(4)	-0.0031	-0.0088	-0.0118	-0.0110
AC(5)	-0.0199	0.0143	-0.0596	-0.0396
AC(10)	-0.0535	-0.0716	-0.0165	-0.0285

Table 4. Table 4: Summary Statistics: Weekly returns (in decimal) of four major equity indexes. AC(k) denotes autocorrelation of order k. The sample period is from July 1981 to March 2008. The source of the data is the online data service Datastream.

5. An Empirical Application To International Equity Markets

We employ the proposed estimation procedure to study returns in international equity markets. The data that we use are the weekly return series, from July 1981 to March 2008, for four major world equity market indexes: the U.S. S&P 500 Composite Index, the Japanese Nikkei 225 Index, the U.K. FTSE 100 Index, and the Hong Kong Hang Seng Index.

While the U.S. and U.K. equity markets are mature and appreciated significantly over the sample period, the emerging market in Hong Kong experienced much higher volatility and more dramatic jumps in prices. The Japanese market, though mature, generated somewhat lower returns over the sample period, although it went through a bubbly period in the late 1980's and then a bursting of the bubble in the 1990's. The rather different risk dynamics of these markets provide a rich ground for analyzing the risk management performance of various estimators of Value at Risk.

Table 4 reports some summary statistics of the data. The mean weekly returns of the four indexes ranges from 0.16% to 0.23% per week, or about 8.32% to 11.96% annually. The Hong Kong Hang Seng Index returned an average 0.23%, a 10-fold increase in the index level over the 20-year sample period. In comparison, the Nikkei 225 index only increased by 6-fold. The U.S. S&P 500 Index and the FTSE 100 Index on average return about 0.17% per week, slightly below that of the Hong Kong Hang Seng Index. However, the Hang Seng's phenomenal rise come with much higher risk than the S&P 500 or the FTSE 100. The weekly sample standard deviation of the index is 3.69%, the highest of the four indexes, as compared to 2.01% for the S&P 500 and 2.40% for the FTSE 100. The Nikkei 225 Index exhibit a weekly standard deviation of 2.40%. As has been documented extensively in the literature, all four indexes display negative skewness and excess kurtosis. The autocorrelation coefficients for all four indexes are quite small. Prior to estimation of the GARCH model, we demean each of the return series using a parsimonious autoregression. Since mean returns at this frequency are small and autocorrelation coefficients are also very modest this step has little impact on the results.

We estimate the Value at Risk for several distinct quantiles, $\{.01, .03, .05, .10, .15\}$ for each of the indices, employing the proposed quantile regression estimation procedure based on the GARCH(1,1) model. We compare the results estimated by the proposed method with results estimated by the CAViaR model and the ARCH models as described in the previous section.

To compare the relative performance, we compute the coverage ratios, that is, the percentage of realized returns that fall below the estimated quantiles. These results are reported in Tables 5-9. Since VaR is an out of sample concept, we consider prediction of VaR for the last 500 periods. Thus, at each time point t (in the last 500 periods), we estimate the model based on data up to time t, and predict the next period $(t+1)$ conditional quantiles (using estimates based on this period information). We compute the coverage ratios based on the percentage of next period realized returns that are below the predicted quantiles .

Formal tests for the out of sample evaluation have been studied in the literature (see e.g. Berkowitz, Christoffersen, and Pelletier (2009) for related literature). A widely used test is the Kupiec (1995) proportion of failure test. The Kupiec test is a likelihood ratio test and has asymptotic χ^2 distribution with one degree of freedom.

Other tests have also been proposed in the literature. For example, if we consider the indicator function: $I_{t+1}(\tau) = 1(u_t \leq Q_{u_t}(\tau|\mathcal{F}_{t-1}))$, then $I_{t+1}(\tau) - \tau$ has mean zero and is a martingale difference sequence, thus

$$Z_n = \frac{1}{\sqrt{n\tau(1-\tau)}} \sum (I_{t+1}(\tau) - \tau) \Rightarrow N(0,1), \text{ as } n \to \infty.$$

(see, e.g., Campbell (2005)). A two-sided test can be constructed based on the above asymptotic normal staistic.

We conduct both the Kupiec test and the Z_n test in our applications. The calculated testing statistics are also reported in Tables 5-9. The 5% level critical values for the Kupiec

% VaR	1%	3%	5%	10%	15%
			SP500		
Coverage Rate	0.0320	0.0500	0.0660	0.1160	0.1880
Kupiec Test	15.4671	5.7489	2.4592	1.3598	5.3140
Z_n Test	4.9441	2.6216	1.6416	1.1926	2.3797
			Nikkei225		
Coverage Rate	0.0380	0.0680	0.0920	0.148	0.2280
Kupiec Test	23.1298	18.3993	15.0408	11.3256	21.1596
Z_n Test	6.2925	4.9811	4.3091	3.5777	4.8845
			FTSE100		
Coverage Rate	0.0280	0.0560	0.0860	0.1320	0.1840
Kupiec Test	10.9940	22.3282	11.3308	5.2231	4.2805
Z_n Test	4.0452	5.7499	3.6935	2.3851	2.1292
			Hang Seng		
Coverage Rate	0.018	0.020	0.038	0.1120	0.174
Kupiec Test	2.6126	1.9421	1.6469	0.7732	2.1671
Z_n Test	1.7979	-1.3108	-1.2312	0.8944	1.5029

TABLE 5. Coverage Rates and Testing Results for QAR Model

% VaR	1%	3%	5%	10%	15%
			SP500		
Coverage Rate	0.0140	0.0380	0.0600	0.1200	0.1780
Kupiec Test	0.7187	1.0159	0.9921	2.1025	2.9307
Z_n Test	0.8989	1.0486	1.0260	1.4907	1.7534
			Nikkei225		
Coverage Rate	0.0180	0.0480	0.086	0.1580	0.2380
Kupiec Test	2.6126	4.7282	11.3308	16.1835	26.5904
Z_n Test	1.7979	2.3595	3.6935	4.3231	5.5108
			FTSE100		
Coverage Rate	0.0180	0.0420	0.0840	0.1240	0.1680
Kupiec Test	2.6126	2.2064	10.1945	2.9967	1.2312
Z_n Test	1.7979	1.5730	3.4883	1.7889	1.1272
			Hang Seng		
Coverage Rate	0.0040	0.0120	0.034	0.0940	0.1620
Kupiec Test	2.3530	7.1705	3.0215	0.2037	0.5528
Z_n Test	-1.3484	-2.3595	-1.6416	-0.4472	0.7515

TABLE 6. Coverage Rates and Testing Results for QGARCH1 Model

test and the Z_n test are 3.841 and 1.96 respectively. The testing results indicates that both the CAViaR method and the quantile GARCH method provide reasonable coverage rates, and are substantially better than the ARCH based estimation.

% VaR	1%	3%	5%	10%	15%
		SP500			
Coverage Rate	0.0160	0.0340	0.0640	0.1240	0.1660
Kupiec Test	1.5383	0.2638	1.9027	2.9967	0.9761
Z_n Test	1.3484	0.5243	1.4364	1.7889	1.0020
		Nikkei225			
Coverage Rate	0.0120	0.052	0.060	0.1160	0.2360
Kupiec Test	0.1899	6.8538	0.9921	1.3598	25.4596
Z_n Test	0.4495	2.8838	1.0260	1.1926	5.3855
		FTSE100			
Coverage Rate	0.020	0.032	0.0700	0.0840	0.1580
Kupiec Test	3.9136	0.0673	3.7651	1.4957	0.2474
Z_n Test	2.2473	0.2622	2.0520	-1.1926	0.5010
		Hang Seng			
Coverage Rate	0.004	0.018	0.0360	0.096	0.1720
Kupiec Test	2.3530	2.8791	2.2765	0.0900	1.8270
Z_n Test	-1.3484	-1.5730	-1.4364	-0.2981	1.3777

Table 7. Coverage Rates and Testing Results for QGARCH2 Model

% VaR	1%	3%	5%	10%	15%
		SP500			
Coverage Rate	0.0140	0.0260	0.0620	0.1200	0.1940
Kupiec Test	0.7187	0.2876	1.4130	2.1025	7.0602
Z_n Test	0.8989	-0.5243	1.2312	1.4907	2.7554
		Nikkei225			
Coverage Rate		0.034	0.0627	0.118	
Kupiec Test		0.2638	1.4215	1.7119	
Z_n Test		0.5243	1.2425	1.3416	
		FTSE100			
Coverage Rate	0.0180	0.0520	0.0700	0.1060	
Kupiec Test	2.6126	6.8538	3.8651	0.1965	
Z_n Test	1.7979	2.8838	2.0520	0.4472	
		Hang Seng			
Coverage Rate	0.0020	0.0250	0.0425	0.0667	0.1160
Kupiec Test	4.8134	0.3639	0.4980	3.0736	4.8539
Z_n Test	-1.7979	-0.5862	-0.6882	-2.3608	-2.1292

Table 8. Coverage Rates and Testing Results for CAViaR1 Model

Appendix A. Proofs

A.1. **Proof of Theorem 1.** Our proofs rely heavily on the theory of empirical processes as in Welsh (1989) and employ exponential inequalities for weakly dependent and martingale difference sequences. We use the notation E_t to signify the conditional expectation $\mathrm{E}(\cdot|x_t)$.

% VaR	1%	3%	5%	10%	15%
SP500					
Coverage Rate	0.0140	0.0260	0.0600	0.1240	0.1940
Kupiec Test	0.7187	0.2876	0.9921	2.9967	7.0602
Z_n Test	0.8989	-0.5243	1.0260	1.7889	2.7554
Nikkei225					
Coverage Rate	0.0175	0.0575	0.0620	0.120	0.186
Kupiec Test	1.8574	8.2419	1.4130	2.1025	7.1598
Z_n Test	1.5076	3.2242	1.2312	1.4907	2.8333
FTSE100					
Coverage Rate	0.0200	0.0520	0.0680	0.1140	0.1825
Kupiec Test	3.9136	6.8538	3.0806	2.0879	3.1363
Z_n Test	2.2473	2.8838	1.8468	1.2981	1.8204
Hang Seng					
Coverage Rate	0.0020	0.029	0.0425	0.0658	0.1160
Kupiec Test	4.8134	0.3277	0.4980	4.749	4.8539
Z_n Test	-1.7979	-0.4389	-0.6882	-2.667	-2.1292

TABLE 9. Coverage Rates and Testing Results for CAViaR2 Model

Let

$$\psi_\tau(u) = \tau - I(u < 0),$$

then $\psi_\tau(u)$ is the right-hand derivative of $\rho_\tau(u)$. ($\rho_\tau(u)$ is differentiable everywhere except at $u = 0$.) The derivative of $\rho_\tau(u_t - \alpha^\top x_t)$ w.r.t. α (except at point $u_t = \alpha^\top x_t$) is then

$$\varphi_{t\tau}(\alpha) = \psi_\tau(u_t - \alpha^\top x_t)x_t = \left[\tau - I(u_t < \alpha^\top x_t)\right] x_t.$$

Notice that

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) = \alpha(\tau)^\top x_t + R_m(\tau).$$

where $R_m(\tau) = \left(\sum_{j=m+1}^{\infty} a_j \left|u_{t-j}\right|\right) Q_\varepsilon(\tau) = O_p(b^m)$ under Assumption S1. Let $u_{t\tau}^* = u_t - \alpha(\tau)^\top x_t$, and

$$u_{t\tau} = u_t - Q_{u_t}(\tau|\mathcal{F}_{t-1}) = \left(a_0 + \sum_{j=1}^{\infty} a_j \left|u_{t-j}\right|\right) [\varepsilon_t - Q_\varepsilon(\tau)] = \sigma_t \varepsilon_{t\tau},$$

then $u_{t\tau}^* = u_{t\tau} + R_m(\tau)$, and

$$\mathrm{E}_t\left[\psi_\tau(u_{t\tau})\right] = 0.$$

Under Assumption S3,

$$\begin{aligned}
\mathrm{E}_t\left(\varphi_{t\tau}(\alpha(\tau))\right) &= \mathrm{E}_t\left[\tau - I(u_t < \alpha(\tau)^\top x_t)\right] x_t \\
&= \left[\tau - F_{u|x}(Q_{u_t}(\tau|x_t) + R_m(\tau))\right] x_t \\
&= O_p\left(b^m \cdot \|x_t\|\right),
\end{aligned}$$

where we define $\|\alpha\|$ to be the L_2 norm of α.

We first show that $\|\widehat{\alpha}(\tau)-\alpha(\tau)\|^2 = O_p(m/n)$. Let $\lambda \in S = \{\lambda \in \mathcal{R}^{m+1} : \|\lambda\| = 1\}$, by convexity of the objective function, it suffices to show that for any $\epsilon > 0$, there exists $B < \infty$ such that, for sufficiently large n,

$$\Pr\left\{\inf_{\lambda\in S}\sum_t \lambda^\top \varphi_{t\tau}(\alpha(\tau)+B(mn)^{1/2}\lambda) > 0\right\} > 1-\epsilon.$$

For notational convenience, we next define

$$\eta_t(v) = \varphi_{t\tau}(\alpha(\tau)+v) - \varphi_{t\tau}(\alpha(\tau))$$

then

$$\begin{aligned}
(14)\quad \sum_t \lambda^\top \varphi_{t\tau}(\alpha(\tau)+B(mn)^{1/2}\lambda) &= \sum_t \lambda^\top \varphi_{t\tau}(\alpha(\tau)) \\
(15)\qquad &+ \sum_t \lambda^\top \mathrm{E}_t\left\{\eta_t(B(mn)^{1/2}\lambda)\right\} \\
(16)\qquad &+ \sum_t \lambda^\top \left[\eta_t(B(mn)^{1/2}\lambda) - \mathrm{E}_t\left\{\eta_t(B(mn)^{1/2}\lambda)\right\}\right]
\end{aligned}$$

we analyze each of the right-hand-side terms (14), (15) and (16), and show that

$$\sum_t \lambda^\top \varphi_{t\tau}(\alpha(\tau)+B(m/n)^{1/2}\lambda) \approx \sum_t \lambda^\top \varphi_{t\tau}(\alpha(\tau)) + B(mn)^{1/2} f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\lambda^\top D_n\lambda$$

For (15), notice that if $\|\alpha-\alpha(\tau)\| \le B(m/n)^{1/2}$,

$$(17)\quad \mathrm{E}_t\left[I(u_t<\alpha(\tau)^\top x_t) - I(u_t<\alpha^\top x_t)\right] = -f_{u|x}(Q_{u_t}(\tau|x_t))x_t^\top[\alpha-\alpha(\tau)] + O_p(m^2/n).$$

Thus, given the GARCH structure (1) and (2),

$$F_{u|x}(u) = F_\varepsilon(u/\sigma_t), f_{u|x}(u) = f_\varepsilon(u/\sigma_t)/\sigma_t$$

thus

$$Q_{u_t}(\tau|x_t) = \sigma_t F_\varepsilon^{-1}(\tau), f_{u|x}(Q_{u_t}(\tau|x_t)) = \frac{1}{\sigma_t} f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right),$$

and

$$\frac{1}{n}\sum_{t=m+1}^n f_{u|x}(Q_{u_t}(\tau|x_t))x_t x_t^\top = f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\left[\frac{1}{n}\sum_{t=m+1}^n \frac{x_t x_t^\top}{\sigma_t}\right] = f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)D_n,$$

so by (17) we have

$$\sum_t \lambda^\top \mathrm{E}_t\left\{\eta_t(B(mn)^{1/2}\lambda)\right\} \approx B(mn)^{1/2} f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\lambda^\top D_n\lambda.$$

To show that the third term (16) is of smaller order of magnitude and can be dropped, we need stochastic equicontinuity corresponding to $\eta_t(v) - \mathrm{E}_t\{\eta_t(v)\}$ using weak dependence property of u and the martingale difference sequence property of the term, as well as the moment condition on x. In particular, we want to show that,

$$\sup_{\|v\|\le B(m/n)^{1/2}}\left|\sum_t \lambda^\top\left[\eta_t(v) - E_t\{\eta_t(v)\}\right]\right| = o_p\left(\frac{1}{\sqrt{nm}}\right).$$

Covering the ball $\left\{\|v\| \leq B(m/n)^{1/2}\right\}$ with cubes $\mathcal{C} = \{\mathcal{C}_k\}$ where $\mathcal{C}_k$ is a cube with center v_k, side length $(m/n^5)^{1/2}B$, so $\mathrm{card}(\mathcal{C}) = (2n^2)^m = N(n)$, and for $v \in \mathcal{C}_k$, $|v - v_k| \leq (m/n^{5/2})B$. Thus, since $I(u_t < z)$ is nondecreasing in z,

$$\begin{aligned}
&\sup_{\|v\| \leq B(m/n)^{1/2}} \left| \sum_t \lambda^\top \left[\eta_t(v) - E_t\left\{\eta_t(v)\right\}\right] \right| \\
&\leq \max_{1 \leq k \leq N(n)} \left| \sum_t \lambda^\top \left[\eta_t(v_k) - E_t\left\{\eta_t(v_k)\right\}\right] \right| \qquad (18)\\
&\quad + \max_{1 \leq k \leq N(n)} \left| \sum_t \left|\lambda^\top x_t\right| \left\{b_{t\tau}(v_k) - \mathrm{E}_t\left[b_{t\tau}(v_k)\right]\right\} \right| \qquad (19)\\
&\quad + \max_{1 \leq k \leq N(n)} \sum_t \left|\lambda^\top x_t\right| \mathrm{E}_t\left[d_{t\tau}(v_k)\right] \qquad (20)
\end{aligned}$$

where

$$\begin{aligned}
b_{t\tau}(v_k) &= I(u_t < (\alpha(\tau) + v_k)^\top x_t) - I(u_t < (\alpha(\tau) + v_k)^\top x_t + (m/n^{5/2})B\|x_t\|),\\
d_{t\tau}(v_k) &= I(u_t < (\alpha(\tau) + v_k)^\top x_t + (m/n^{5/2})B\|x_t\|)\\
&\quad - I(u_t < (\alpha(\tau) + v_k)^\top x_t - (m/n^{5/2})B\|x_t\|).
\end{aligned}$$

The analysis of terms (19) and (20) are similar to Welsh (1989). We focus on the first term (18). Notice that $\mathrm{card}(\mathcal{C}) = (2n^2)^m$, an exponential inequality is needed to control the rate. Since $\|v_k\| \leq B(m/n)^{1/2}$, by calculation of moments, we have

$$\omega_n^2 = \sum_t \mathrm{E}_t \left[\lambda^\top \left[\eta_t(v_k) - E_t\left\{\eta_t(v_k)\right\}\right]\right]^2 = O_p((mn)^{1/2}m^{3/2},$$

and

$$S_n^2 = \sum_t \left[\lambda^\top \left[\eta_t(v_k) - E_t\left\{\eta_t(v_k)\right\}\right]\right]^2 = O_p((mn)^{1/2}m^{3/2}.$$

Let $M = (mn)^{1/2}$, noting that $\xi_t = [\eta_t(v_k) - E_t\{\eta_t(v_k)\}]$ is a martingale difference sequence we have

$$\begin{aligned}
&\Pr\left[\max_{1 \leq k \leq N(n)} \left| \frac{1}{\sqrt{n}} \sum_t \lambda^\top \left[\eta_t(v_k) - E_t\left\{\eta_t(v_k)\right\}\right] \right| > \epsilon\right]\\
&\leq N(n) \max_k \Pr\left(\left| \frac{1}{\sqrt{n}} \sum_t \lambda^\top \left[\eta_t(v_k) - E_t\left\{\eta_t(v_k)\right\}\right] \right| > \varepsilon\right)\\
&\leq N(n) \max_k \Pr\left(\left| \sum_t \lambda^\top \xi_t \right| > \sqrt{n}\varepsilon;\ S_n^2 + \omega_n^2 \leq M\right)\\
&\quad + N(n) \max_k \Pr\left(\left| \sum_t \lambda^\top \xi_t \right| > \sqrt{n}\varepsilon;\ S_n^2 + \omega_n^2 > M\right)
\end{aligned}$$

For the first term, by exponential inequality for martingale difference sequences (see, e.g., Bercu and Touati (2008)), we have

$$N(n)\max_{k}\Pr\left(\left|\sum_{t}\lambda^{\top}\xi_{t}\right|>\sqrt{n}\varepsilon;\ S_{n}^{2}+\omega_{n}^{2}\leq M\right)\leq 2N(n)\exp\left(-\frac{n\epsilon^{2}}{2M}\right).$$

For the second term,

$$\Pr\left(\ S_{n}^{2}+\omega_{n}^{2}>M\right)\leq\Pr\left(\ S_{n}^{2}>M/2\right)+\Pr\left(\omega_{n}^{2}>M/2\right),$$

and each term can be bounded exponentially under assumptions S1 and S5. Thus,

$$\begin{aligned}\sum_{t}\lambda^{\top}\varphi_{t\tau}(\alpha(\tau)+B(m/n)^{1/2}\lambda)=\sum_{t}\lambda^{\top}\varphi_{t\tau}(\alpha(\tau))\quad&+\quad B(mn)^{1/2}f_{\varepsilon}\left(F_{\varepsilon}^{-1}(\tau)\right)\lambda^{\top}D_{n}\lambda\\&+\quad o_{p}\left((nm)^{1/2}\right).\end{aligned}$$

By Assumption S4, the minimum eigenvalue of $\overline{D}_n$ is bounded from below, and

$$\sum_{t}\varphi_{t\tau}(\alpha(\tau))=O_{p}(\sqrt{nm})$$

so for large n,

$$\begin{aligned}&\left\{\inf_{\lambda\in S}\sum_{t}\lambda^{\top}\varphi_{t\tau}(\alpha(\tau)+B(mn)^{1/2}\lambda)>0\right\}\\\supseteq\quad&\left\{\frac{1}{\sqrt{nm}}\inf_{\lambda\in S}\sum_{t}\lambda^{\top}\varphi_{t\tau}(\alpha(\tau))>-\frac{B}{2}\lambda_{\min}\left[f_{\varepsilon}\left(F_{\varepsilon}^{-1}(\tau)\right)D_{n}\right]\right\}\end{aligned}$$

whose probability goes to 1 as B and $n\to\infty$. Thus we have proved (12).

If $\widetilde{\alpha}(\tau)$ is the solution of (11), let $\widehat{v}=\sqrt{n}(\widetilde{\alpha}(\tau)-\alpha(\tau))$, from the above analysis we have,

$$\sqrt{n}\left(\widehat{\alpha}(\tau)-\alpha(\tau)\right)=-\frac{1}{f_{\varepsilon}\left(F_{\varepsilon}^{-1}(\tau)\right)}D_{n}^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n}x_{t}\psi_{\tau}(u_{t\tau})\right)+o_{p}(1),$$

thus for any $\lambda\in\mathcal{R}^{m+1}$,

$$\frac{\sqrt{n}\lambda^{\top}\left(\widehat{\alpha}(\tau)-\alpha(\tau)\right)}{\sigma_{\lambda}}\Rightarrow N(0,1),$$

where

$$\sigma_{\lambda}^{2}=\lambda^{\top}D_{n}^{-1}\Sigma_{n}D_{n}^{-1}\lambda,\Sigma_{n}=\frac{1}{n}\sum_{t=m+1}^{n}x_{t}x_{t}^{\top}\psi_{\tau}^{2}(u_{t\tau}).\blacksquare$$

A.2. **Proof of Theorem 2.** To analyze the asymptotic behavior of our estimators, we need to first derive the asymptotic representation for $\tilde{\pi}$. Notice that

$$\sqrt{n}\left(\widetilde{\alpha}(\tau_k)-\alpha(\tau_k)\right)=-\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_k)\right)}D_n^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n}x_t\psi_{\tau_k}(u_{t\tau_k})\right)+o_p(1),$$

thus,

$$\begin{aligned}\sqrt{n}(\tilde{\pi}-\pi) &= -\begin{bmatrix}\left(\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n}D_n^{-1}x_t\frac{\psi_{\tau_1}(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)}\right)\\ \cdots \\ \left(\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n}D_n^{-1}x_t\frac{\psi_{\tau_K}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)}\right)\end{bmatrix}+o_p(1)\\ &= -\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n}\Upsilon_{Kt}\otimes\left[D_n^{-1}x_t\right]+o_p(1)\end{aligned}$$

Let $\mathcal{D}=I_K\otimes D_n$, $V_{xt}=x_tx_t^\top$, and

$$V_{\psi t}=\Upsilon_{Kt}\Upsilon_{Kt}^\top=\begin{bmatrix}\frac{\psi_{\tau_1}^2(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)^2} & \cdots & \frac{\psi_{\tau_1}(u_{t\tau_1})\psi_{\tau_1}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)}\\ \cdots & \ddots & \cdots\\ \frac{\psi_{\tau_1}(u_{t\tau_1})\psi_{\tau_1}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \cdots & \frac{\psi_{\tau_K}^2(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)^2}\end{bmatrix}.$$

And set,

$$\begin{aligned}\Psi_K &= \frac{1}{n}\sum_{t=m+1}^{n}V_{\psi t}\otimes V_{xt}\\ &= \begin{bmatrix}\frac{1}{n}\sum x_tx_t^\top\frac{\psi_{\tau_1}^2(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)^2} & \cdots & \frac{1}{n}\sum x_tx_t^\top\frac{\psi_{\tau_1}(u_{t\tau_1})\psi_{\tau_1}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)}\\ \cdots & \ddots & \cdots\\ \frac{1}{n}\sum x_tx_t^\top\frac{\psi_{\tau_1}(u_{t\tau_1})\psi_{\tau_1}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \cdots & \frac{1}{n}\sum x_tx_t^\top\frac{\psi_{\tau_1}^2(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)^2}\end{bmatrix}.\end{aligned}$$

Define $\Lambda=\mathcal{D}^{-1}\Psi_K\mathcal{D}^{-1}$, and denote

$$G=\left.\frac{\partial\phi(\mathbf{a})}{\partial\mathbf{a}^\top}\right|_{\mathbf{a}=\mathbf{a}_0}=\dot{\phi}(\mathbf{a}_0)=\left[g\otimes J_m\vdots I_K\otimes\alpha\right].$$

The objective function may be equivalently written as

$$\begin{aligned}Q_n(\mathbf{a}) &= \left(\tilde{\pi}-\phi(\mathbf{a})\right)^\top A_n\left(\tilde{\pi}-\phi(\mathbf{a})\right)\\ &= \left([\tilde{\pi}-\pi]-[\phi(\mathbf{a})-\phi(\mathbf{a}_0)]\right)^\top A_n\left([\tilde{\pi}-\pi]-[\phi(\mathbf{a})-\phi(\mathbf{a}_0)]\right)\end{aligned}$$

and the first order condition is given by:

$$\frac{1}{2}\frac{\partial Q_n(\widehat{\mathbf{a}})}{\partial\mathbf{a}}=-\left(\tilde{\pi}-\pi\right)^\top A_n\dot{\phi}(\widehat{\mathbf{a}})+\dot{\phi}(\widehat{\mathbf{a}})^\top A_n\left(\phi(\widehat{\mathbf{a}})-\phi(\mathbf{a}_0)\right)=0$$

Thus,

$$\begin{aligned}\sqrt{n}(\widehat{\mathbf{a}}-\mathbf{a}_0) &= \left[\dot{\phi}(\mathbf{a}_0)^\top A_n \dot{\phi}(\mathbf{a}_0)\right]^{-1} \dot{\phi}(\widehat{\mathbf{a}}) A_n \sqrt{n}\,(\tilde{\pi}-\pi)+o_p(1)\\ &= \left[G^\top A_n G\right]^{-1} G^\top A_n \sqrt{n}\,(\tilde{\pi}-\pi)+o_p(1)\\ &= -\left[G^\top A_n G\right]^{-1} G^\top A_n \left[\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n} \Upsilon_{Kt}\otimes\left[D_n^{-1}x_t\right]\right]+o_p(1).\blacksquare\end{aligned}$$

A.3. **Proof of Corollary 3.** Notice that

$$\begin{aligned}\Psi_K &\sim \begin{bmatrix} \frac{\tau_1(1-\tau_1)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)^2} & \cdots & \frac{\tau_1\wedge\tau_K-\tau_1\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\ \cdots & \ddots & \cdots \\ \frac{\tau_1\wedge\tau_K-\tau_1\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \cdots & \frac{\tau_K(1-\tau_K)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)^2}\end{bmatrix} \otimes \left[\frac{1}{n}\sum x_t x_t^\top\right]\\ &= \Delta\otimes\left[\frac{1}{n}\sum x_t x_t^\top\right]\end{aligned}$$

where

$$\Delta = \begin{bmatrix} \frac{\tau_1(1-\tau_1)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)^2} & \cdots & \frac{\tau_1\wedge\tau_K-\tau_1\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\ \cdots & \ddots & \cdots \\ \frac{\tau_1\wedge\tau_K-\tau_1\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \cdots & \frac{\tau_K(1-\tau_K)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)^2}\end{bmatrix},$$

thus, for $\lambda\in\mathcal{R}^{m+K}$,

$$\begin{aligned}\lambda^\top\sqrt{n}(\widehat{\mathbf{a}}-\mathbf{a}_0) &= \lambda^\top\left[G^\top A_n G\right]^{-1} G^\top A_n\sqrt{n}\,(\tilde{\pi}-\pi)+o_p(1)\\ &= \lambda^\top\left[G^\top A_n G\right]^{-1} G^\top A_n N\left(0,\Delta\otimes D^{-1}\Sigma D^{-1}\right)+o_p(1)\end{aligned}$$

By definition,

$$\widetilde{\sigma}_t=\widetilde{a}_0+\sum_{j=1}^{m}\widetilde{a}_j\left|u_{t-j}\right|,$$

we have

$$\begin{aligned}\sqrt{n}\left(\widetilde{\sigma}_t-\sigma_t\right) &= \sum_{j=1}^{m}\sqrt{n}\left(\widetilde{a}_j-a_j\right)\left|u_{t-j}\right|+o_p\left(\frac{m^2}{n}\right)\\ &= \left(\left|u_{t-1}\right|,\cdots,\left|u_{t-m}\right|\right)\begin{bmatrix}\sqrt{n}\left(\widetilde{a}_1-a_1\right)\\ \cdots\\ \sqrt{n}\left(\widetilde{a}_m-a_m\right)\end{bmatrix}+o_p\left(1\right)\end{aligned}$$

Notice that

$$\begin{bmatrix}\sqrt{n}\left(\widetilde{a}_1-a_1\right)\\ \cdots\\ \sqrt{n}\left(\widetilde{a}_m-a_m\right)\end{bmatrix}=\frac{1}{\sqrt{n}}\sum_j H_j\Upsilon_{Kj},$$

let $H = \frac{1}{n}\sum_j H_j \left(\mathrm{E}\left[\Upsilon_{Kj}\Upsilon_{Kj}\right]\right) H_j = \frac{1}{n}\sum_j H_j \Delta H_j$, $\underline{x}_t = \left(|u_{t-1}|, \cdots, |u_{t-m}|\right)^\top$, and $\omega_t^2 = \underline{x}_t^\top H \underline{x}_t$, we have, conditional on information prior to t,

$$\begin{aligned} & \sqrt{n}\left(\widetilde{\sigma}_t - \sigma_t\right) \\ = & -\left(|u_{t-1}|, \cdots, |u_{t-m}|\right) L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_n \left[\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n} \Upsilon_{Kt} \otimes \left[D_n^{-1} x_t\right]\right] + o_p(1) \\ = & \left(|u_{t-1}|, \cdots, |u_{t-m}|\right) \frac{1}{\sqrt{n}} \sum_j H_j \Upsilon_{Kj} + o_p(1) \\ \Rightarrow & \; N\left(0, \underline{x}_t^\top H \underline{x}_t\right), \end{aligned}$$

thus the result can be obtained. ∎

A.4. **Proof of Theorem 4.** We consider quantile regression of u_t on

$$\widetilde{z}_t = (1, \widetilde{\sigma}_{t-1}, \cdots, \widetilde{\sigma}_{t-p}, |u_{t-1}|, \cdots, |u_{t-q}|).$$

For convenience of analysis, we may rewrite $\widetilde{z}_t = z_t(\widetilde{a})$ since it contains elements of

$$\widetilde{\sigma}_{t-k} = \widetilde{\sigma}_{t-k}\left(\widetilde{a}\right) = a_0 + \sum_{j=1}^{m} \widetilde{a}_j \left|u_{t-k-j}\right|.$$

The second stage estimation can then be rewritten as

$$\min_\theta \sum_t \rho_\tau(u_t - \theta^\top z_t(\widetilde{a})).$$

Denote

$$G_n(\theta, a) = \frac{1}{n}\sum_t \psi_\tau(u_t - \theta^\top z_t\left(a\right)) z_t\left(a\right) = \frac{1}{n}\sum_t \left[\tau - I(u_t < \theta^\top z_t\left(a\right))\right] z_t\left(a\right),$$

and

$$G(\theta, a) = \mathrm{E}\left[\psi_\tau(u_t - \theta^\top z_t\left(a\right)) z_t\left(a\right)\right].$$

By iterated expectations

$$\begin{aligned} G(\theta, a) &= \mathrm{E}\left\{\mathrm{E}_t\left[\tau - I(u_t < \theta^\top z_t\left(a\right))\right] z_t\left(a\right)\right\} \\ &= \mathrm{E}\left[\left\{\tau - F_{u|x}(\theta^\top z_t\left(a\right))\right\} z_t\left(a\right)\right]. \end{aligned}$$

Under our conditions, the asymptotic behavior of the second stage estimator $\widehat{\theta}\left(\tau\right)$ is the same as that of $\arg\min_\theta \left\|G_n(\theta, \widetilde{a})\right\|$, and $\theta(\tau)$ solves $\min_\theta \left\|G(\theta, a_0)\right\|$.

We first establish $\sqrt{n}$-consistency of $\widehat{\theta}\left(\tau\right)$ to $\theta\left(\tau\right)$. Let

$$\Gamma_1(\theta, a) = \frac{\partial G(\theta, a)}{\partial \theta} = -\mathrm{E} f_{u|x}(\theta^\top z_t\left(a\right)) z_t\left(a\right) z_t\left(a\right)^\top,$$

and notice that, denoting the vector of true values of a as a^*,

$$\Gamma_{10} = \Gamma_1(\theta, a)|_{\theta=\theta(\tau), a=a^*} \approx -\mathrm{E} f_{u|x}(Q_{u_t}(\tau|x_t)) z_t z_t{}^\top = -f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\Omega,$$

under Assumption S3, $\Gamma_1(\theta,\alpha)$ is continuous at $\theta = \theta(\tau)$ and Γ_{10} is nonsingular, thus there exists a constant $C > 0$ such that $C\|\widehat{\theta}(\tau) - \theta(\tau)\|$ is bounded by $\|G(\widehat{\theta}(\tau), a_0)\|$ with probability going to 1. Define the $(p+q+1)\times m$ matrix,

$$\begin{aligned}\Gamma_2(\theta,a) &= \frac{\partial G(\theta,a)}{\partial a^\top}\\ &= \mathrm{E}\left[\left\{\tau - F_{u|x}(\theta^\top z_t(a))\right\}\frac{\partial z_t(a)}{\partial a^\top}\right] - \mathrm{E}\left[f_{u|x}(\theta^\top z_t(a)) z_t(a)\sum_{j=1}^{p}\theta_j\frac{\partial \sigma_{t-j}(a)}{\partial a^\top}\right]\end{aligned}$$

Notice that $\|G(\theta(\tau), \alpha^*)\| = O_p(b^m)$ and $\|G_n(\theta(\tau), a^*)\| = O_p(n^{-1/2})$, so by the triangle inequality we have

$$\begin{aligned}\|G(\widehat{\theta}(\tau), a^*)\| \le\ & \|G(\widehat{\theta}(\tau), a^*) - G(\widehat{\theta}(\tau), \widetilde{a})\| && (21)\\ & +\|G(\widehat{\theta}(\tau), \widetilde{a}) - G(\theta(\tau), a^*) - G_n(\widehat{\theta}(\tau), \widetilde{a}) + G_n(\theta(\tau), a^*)\| && (22)\\ & +\|G_n(\widehat{\theta}(\tau), \widetilde{a})\| && (23)\\ & +O_p(n^{-1/2}).\end{aligned}$$

We now analyze the terms (21), (22) and (23).

First, for (21), again, by triangle inequality

$$\begin{aligned}\|G(\widehat{\theta}(\tau), a^*) - G(\widehat{\theta}(\tau), \widetilde{a})\| \le\ & \|G(\widehat{\theta}(\tau), \widetilde{a}) - G(\widehat{\theta}(\tau), a^*) - \Gamma_2(\widehat{\theta}(\tau), a^*)(\widetilde{\alpha} - a^*)\|\\ & +\|\Gamma_2(\widehat{\theta}(\tau), a^*)(\widetilde{a} - a^*) - \Gamma_2(\theta(\tau), a^*)(\widetilde{a} - a^*)\|\\ & +\|\Gamma_2(\theta(\tau), a^*)(\widetilde{a} - a^*)\|.\end{aligned}$$

Under Assumption S1, S2 and S3, we have

$$\|G(\widehat{\theta}(\tau), \widetilde{a}) - G(\widehat{\theta}(\tau), a^*) - \Gamma_2(\widehat{\theta}(\tau), a^*)(\widetilde{a} - a^*)\| = O_p\left(\|\widetilde{a} - a^*\|^2\right)$$

and

$$\|\Gamma_2(\widehat{\theta}(\tau), a^*)(\widetilde{a} - a^*) - \Gamma_2(\theta(\tau), a^*)(\widetilde{a} - a^*)\| = O_p\left(\|\widehat{\theta}(\tau) - \theta(\tau)\|\right) o_p(1).$$

Thus,

$$\begin{aligned}\|G(\widehat{\theta}(\tau), a^*) - G(\widehat{\theta}(\tau), \widetilde{a})\| \le\ & O_p\left(\|\widetilde{a} - a^*\|^2\right) + O_p\left(\|\widehat{\theta}(\tau) - \theta(\tau)\|\|\widetilde{a} - a^*\|\right)\\ & +\|\Gamma_2(\theta(\tau), a^*)(\widetilde{a} - a^*)\|. && (24)\end{aligned}$$

In addition,

$$\begin{aligned}\Gamma_2(\theta(\tau), a^*) &\approx -\mathrm{E}\left[f_{u|x}(Q_{u_t}(\tau|x_t)) z_t\sum_{j=1}^{p}\theta_j\frac{\partial \sigma_{t-j}(a)}{\partial a^\top}\right]\\ &= -f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\mathrm{E}\left[\sum_{j=1}^{p}\theta_j\frac{z_t}{\sigma_t}\frac{\partial \sigma_{t-j}(a)}{\partial a^\top}\right]\end{aligned}$$

since

$$\mathrm{E}_t\left[\left\{\tau - F_{u|x}(\theta(\tau)^\top z_t(a^*))\right\}\frac{\partial z_t(a^*)}{\partial a}\right] \approx \mathrm{E}_t\left[\left\{\tau - F_{u|x}(Q_{u_t}(\tau|\mathcal{F}_{t-1}))\right\}\frac{\partial z_t(a^*)}{\partial a}\right] = 0.$$

Thus,

$$\|G(\widehat{\theta}(\tau), a^*) - G(\widehat{\theta}(\tau), \widetilde{a})\| \leq \|\Gamma_2(\theta(\tau), a^*)(\widetilde{a} - a^*)\|(1 + o_p(1)),$$

and

$$\|G(\widehat{\theta}(\tau), \widetilde{a})\| \leq \|G(\widehat{\theta}(\tau), a^*)\|(1 + o_p(1)).$$

For (22), we need to verify stochastic equicontinuity. If we denote

$$m_\tau(Z_t, \theta, a) = \psi_\tau(u_t - \theta^\top z_t(a)) z_t(a),$$

for each element $m_{j\tau}(Z_t, \theta, a) = \psi_\tau(u_t - \theta^\top z_t(a)) z_{jt}(a)$ of $m_\tau(Z_t, \theta, a)$,

$$\begin{aligned} \left| m_{j\tau}(Z_t, \overline{\theta}, \overline{a}) - m_{j\tau}(Z_t, \theta, a) \right| \leq\ & \tau \left| z_{jt}(\overline{a}) - z_{jt}(a) \right| \\ & + \left| I\left(u_t < \overline{\theta}^\top z_t(\overline{a})\right) z_{jt}(\overline{a}) - I\left(u_t < \theta^\top z_t(a)\right) z_{jt}(a) \right|. \end{aligned}$$

For $\tau \left| z_{jt}(\overline{a}) - z_{jt}(a) \right|$, if $\|\overline{a} - a\| \leq \delta$,

$$\tau^r \mathrm{E} \left| z_{jt}(\overline{a}) - z_{jt}(a) \right|^r \leq C_{j1} (\delta / m)^r .$$

For the second term,

$$\begin{aligned} & \left| I\left(u_t < \overline{\theta}^\top z_t(\overline{a})\right) z_t(\overline{a}) - I\left(u_t < \theta^\top z_t(a)\right) z_t(a) \right| \leq \\ & \quad \left| I\left(u_t < \overline{\theta}^\top z_t(\overline{a})\right) z_{jt}(\overline{a}) - I\left(u_t < \theta^\top z_t(a)\right) z_{jt}(\overline{a}) \right| \\ & \quad + \left| I\left(u_t < \theta^\top z_t(a)\right) z_{jt}(\overline{a}) - I\left(u_t < \theta^\top z_t(a)\right) z_{jt}(a) \right| \end{aligned}$$

Since $I(u_t < \cdot)$ is a monotonic function,

$$\begin{aligned} & \mathrm{E} \left| I\left(u_t < \overline{\theta}^\top z_t(\overline{a})\right) z_{jt}(\overline{a}) - I\left(u_t < \theta^\top z_t(a)\right) z_{jt}(\overline{a}) \right| \\ \leq\ & \sup_{\|\overline{a} - a\| \leq \delta, \|\overline{\theta} - \theta\| \leq \delta} \mathrm{E} \left[F_{u|x}\left(\overline{\theta}^\top z_t(\overline{a})\right) - F_{u|x}\left(\theta^\top z_t(a)\right) \right] \mathrm{E} \left| z_{jt}(\overline{a}) \right| \\ \leq\ & C_{j2}(\delta / m), \end{aligned}$$

under our smoothness assumption on $F_{u|x}(\cdot)$ and the moment condition on u. Thus, by Lemma 4.2 of Chen (2008), we have,

$$\sup_{\|a - a^*\| \leq \delta, \|\theta - \theta(\tau)\| \leq \delta} \frac{\sqrt{n} \|G_n(\theta, a) - G(\theta, a) - G_n(\theta(\tau), a^*) + G(\theta(\tau), a^*)\|}{1 + \sqrt{n} \left\{ \|G_n(\theta, a)\| + \|G(\theta, a)\| \right\}} = o_p(1),$$

consequently,

$$\begin{aligned} & \|G(\widehat{\theta}(\tau), \widetilde{a}) - G(\theta(\tau), a^*) - G_n(\widehat{\theta}(\tau), \widetilde{a}) + G_n(\theta(\tau), a^*)\| \\ \leq\ & o_p(1) \times \left\{ \|G_n(\widehat{\theta}(\tau), \widetilde{a})\| + \|G(\widehat{\theta}(\tau), \widetilde{a})\| \right\} \\ \leq\ & o_p(1) \times \left\{ \|G_n(\widehat{\theta}(\tau), \widetilde{a})\| + \|G(\widehat{\theta}(\tau), a^*)\|(1 + o_p(1) \right\}, \end{aligned}$$

where the last inequality comes from (24). Thus,

$$\begin{aligned}\|G(\widehat{\theta}(\tau), a^*)\| &\leq \|\Gamma_2(\theta(\tau), a^*)(\widetilde{a}-a^*)\| + O_p\left(\|\widetilde{a}-a^*\|^2\right) + O_p\left(\|\widehat{\theta}(\tau)-\theta(\tau)\|\|\widetilde{a}-a^*\|\right)\\ &\quad + o_p(1)\times\left\{\|G_n(\widehat{\theta}(\tau), \widetilde{a})\| + \|G(\widehat{\theta}(\tau), a^*)\|(1+o_p(1)\right\}\\ &\quad + \|G_n(\widehat{\theta}(\tau), \widetilde{a})\|,\end{aligned}$$

and

$$\begin{aligned}\|G(\widehat{\theta}(\tau), a^*)\|(1-o_p(1)) &\leq \|G_n(\widehat{\theta}(\tau), \widetilde{a})\|(1+o_p(1)) + O_p(n^{-1/2})\\ &= \inf_{\theta}\|G_n(\theta, \widetilde{a})\| + O_p(n^{-1/2}).\end{aligned}$$

We only need to show that

$$\inf_{\theta}\|G_n(\theta, \widetilde{a})\| = O_p(n^{-1/2}),$$

which is true since

$$\begin{aligned}\|G_n(\theta, \widetilde{a})\| &\leq \|G_n(\theta, \widetilde{a}) - G(\theta, \widetilde{a}) - G_n(\theta(\tau), a^*)\|\\ &\quad + \|G(\theta, \widetilde{a}) - G(\theta, a^*)\| + \|G(\theta, a^*)\| + \|G_n(\theta(\tau), a^*)\|\\ &\leq o_p(1)\times\{\|G_n(\theta, \widetilde{a})\| + \|G(\theta, \widetilde{a})\|\} + \|G(\theta, a^*)\| + O_p(n^{-1/2}).\end{aligned}$$

Thus,

$$\|G_n(\theta, \widetilde{a})\|(1-o_p(1)) \leq o_p(1)\times\{\|G(\theta, \widetilde{a})\|\} + \|G(\theta, a^*)\| + O_p(n^{-1/2}),$$

and

$$\inf_{\theta}\|G_n(\theta, \widetilde{a})\| = O_p(n^{-1/2}),$$

since $\|G(\theta(\tau), a^*)\| = 0$ and

$$\|G(\theta, \widetilde{a})\| \leq \|G(\theta, a^*)\| + \|\Gamma_2(\theta(\tau), a^*)(\widetilde{a}-a^*)\|(1+o_p(1)).$$

And consequently,

$$C\|\widehat{\theta}(\tau) - \theta(\tau)\| \leq \|G(\widehat{\theta}(\tau), a^*)\| = O_p(n^{-1/2}).$$

Now define the linearization

$$L_n(\theta, \widetilde{a}) = G_n(\theta(\tau), a^*) + G(\theta, a^*) + \Gamma_2(\theta(\tau), a^*)(\widetilde{a}-a^*),$$

and note that

$$\begin{aligned}G_n(\theta, \widetilde{a}) &= G_n(\theta(\tau), a^*) + \Gamma_1(\theta-\theta(\tau)) + \Gamma_2(\theta(\tau), a^*)(\widetilde{a}-a^*)\\ &\quad + G(\theta, a^*) - G(\theta(\tau), a^*) - \Gamma_1(\theta-\theta(\tau))\\ &\quad + \Gamma_2(\theta, a^*)(\widetilde{a}-a^*) - \Gamma_2(\theta(\tau), a^*)(\widetilde{a}-a^*)\\ &\quad + G(\theta, \widetilde{a}) - G(\theta, a^*) - \Gamma_2(\theta, a^*)(\widetilde{a}-a^*)\\ &\quad + G_n(\theta, \widetilde{a}) - G(\theta, \widetilde{a}) - G_n(\theta(\tau), a^*) + G(\theta(\tau), a^*)\\ &\quad - G(\theta(\tau), a^*).\end{aligned}$$

Under Assumptions S1 - S6,

$$\begin{aligned}
\|G_n(\widehat{\theta},\widetilde{a})-L_n(\widehat{\theta},\widetilde{a})\| &\leq \|G(\widehat{\theta},a^*)-G(\theta(\tau),a^*)-\Gamma_1(\widehat{\theta}-\theta(\tau))\| \\
&\quad +\|\Gamma_2(\theta,a^*)(\widetilde{a}-a^*)-\Gamma_2(\theta(\tau),a^*)(\widetilde{a}-a^*)\| \\
&\quad +\|G(\widehat{\theta},\widetilde{a})-G(\widehat{\theta},a^*)-\Gamma_2(\widehat{\theta},a^*)(\widetilde{a}-a^*)\| \\
&\quad +\|G_n(\widehat{\theta},\widetilde{a})-G(\widehat{\theta},\widetilde{a})-G_n(\theta(\tau),a^*)+G(\theta(\tau),a^*)\| \\
&\quad +\|G(\theta(\tau),a^*)\| \\
&= o_p(n^{-1/2}),
\end{aligned}$$

because

$$\|G(\widehat{\theta},a^*)-G(\theta(\tau),a^*)-\Gamma_1(\widehat{\theta}-\theta(\tau))\| = O_p(\|\widehat{\theta}-\theta(\tau)\|^2)=o_p(n^{-1/2}),$$

$$\|\Gamma_2(\theta,a^*)(\widetilde{a}-a^*)-\Gamma_2(\theta(\tau),a^*)(\widetilde{a}-a^*)\| = o_p(1)\|\widehat{\theta}-\theta(\tau)\| = o_p(n^{-1/2}),$$

by root-n consistency;

$$\|G(\widehat{\theta},\widetilde{a})-G(\widehat{\theta},a^*)-\Gamma_2(\widehat{\theta},a^*)(\widetilde{a}-a^*)\| \leq C\left(\|\widetilde{a}-a^*\|^2\right)=o_p(n^{-1/2}),$$

$$\|G_n(\widehat{\theta},\widetilde{a})-G(\widehat{\theta},\widetilde{a})-G_n(\theta(\tau),a^*)+G(\theta(\tau),a^*)\| = o_p(n^{-1/2}),$$

by stochastic equicontinuity, and

$$\|G(\theta(\tau),a^*)\| = o_p(n^{-1/2}),$$

by definition. Thus

$$\min_{\theta}\|G_n(\theta,\widetilde{a})\| = \min_{\theta}\|L_n(\theta,\widetilde{a})\| + o_p(n^{-1/2}), \tag{25}$$

and

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\theta}(\tau)-\theta(\tau)\right) \\
&= -\left(\Gamma_1^\top\Gamma_1\right)^{-1}\Gamma_1^\top\sqrt{n}\left[G_n(\theta(\tau),a^*)+\Gamma_2(\theta(\tau),a^*)(\widetilde{a}-a^*)\right] \\
&= -\left(f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)^2\Omega^2\right)^{-1} f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\Omega\sqrt{n}\left[G_n(\theta(\tau),a^*)+\Gamma_2(\theta(\tau),a^*)(\widetilde{a}-a^*)\right] \\
&= -\left(f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\right)^{-1}\Omega^{-1}\sqrt{n}\left[G_n(\theta(\tau),a^*)+\Gamma_2(\theta(\tau),a^*)(\widetilde{a}-a^*)\right] \\
&= -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1}\left\{\frac{1}{\sqrt{n}}\sum_t z_t\psi_\tau(u_{t\tau})+\Gamma_2(\theta(\tau),a^*)\sqrt{n}(\widetilde{a}-a^*)\right\}.
\end{aligned}$$

In addition,

$$\Gamma_2(\theta(\tau),a^*) \approx -\mathrm{E}\left[f_{u|x}(Q_{u_t}(\tau|x_t))z_t\sum_{j=1}^p\theta_j\frac{\partial\sigma_{t-j}(a)}{\partial a^\top}\right] = -f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\mathrm{E}\left[\sum_{j=1}^p\theta_j\frac{z_t}{\sigma_t}\frac{\partial\sigma_{t-j}(a)}{\partial a^\top}\right]$$

since

$$\frac{\partial\sigma_{t-k}(a)}{\partial a_j} = |u_{t-k-j}|,$$

we have

$$\Gamma_2(\theta(\tau), a^*) \approx -f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right) \mathrm{E}\left[\sum_{j=1}^{p} \theta_j \left[\frac{z_t |u_{t-j-1}|}{\sigma_t}, \cdots, \frac{z_t |u_{t-j-m}|}{\sigma_t}\right]\right] = \Gamma_{20} = f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\Gamma.$$

Let $L_{m/K} = \left[I_m \vdots 0_{m\times K}\right]$, then the minimum distance estimator of $[a_1, a_2, \cdots, a_m]$ has asymptotic representation:

$$-L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_n \left[\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n} \Upsilon_{Kt} \otimes \left[D_n^{-1} x_t\right]\right] + o_p(1)$$

Thus, the two-step estimator of $\theta(\tau)$ has the following Bahadur representation:

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) \\
&= \Gamma_1(\theta(\tau), \alpha_0)^{-1}\left\{\sqrt{n} G_n(\theta(\tau), a^*) + \Gamma_{20}\sqrt{n}(\widetilde{a} - a^*)\right\} + o_p(1) \\
&= -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1}\left\{\frac{1}{\sqrt{n}}\sum_t z_t \psi_\tau(u_{t\tau})\right\} \\
&\quad -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1}\Gamma_{20} L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_n\left[\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n}\Upsilon_{Kt}\otimes\left[D_n^{-1}x_t\right]\right] \\
&\quad + o_p(1).
\end{aligned}$$

■

A.5. **Proof of Corollary 5.** For the asymptotic distribution of $\widehat{\theta}(\tau)$, notice that

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) \\
&= \Gamma_1(\theta(\tau), \alpha_0)^{-1}\left\{\sqrt{n} G_n(\theta(\tau), a^*) + \Gamma_{20}\sqrt{n}(\widetilde{a} - a^*)\right\} + o_p(1) \\
&= -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1}\left\{\frac{1}{\sqrt{n}}\sum_t z_t \psi_\tau(u_{t\tau})\right\} \\
&\quad -\Omega^{-1}\Gamma L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_n\left[\frac{1}{\sqrt{n}}\sum_{t=m+1}^{n}\Upsilon_{Kt}\otimes\left[D_n^{-1}x_t\right]\right] + o_p(1),
\end{aligned}$$

let

$$\Psi_t = \left[\begin{array}{cccc} \frac{\psi_\tau(u_{t\tau})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}, & \frac{\psi_{\tau_1}(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)}, & \cdots, & \frac{\psi_{\tau_K}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \end{array}\right]^\top,$$

and

$$M_t = \left[\begin{array}{cccc} z_t, & \Gamma L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_{n1} D_n^{-1} x_t, & \cdots, & \Gamma L_{m/K}\left[G^\top A_n G\right]^{-1} G^\top A_{nK} D_n^{-1} x_t \end{array}\right],$$

$$\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) \Rightarrow N\left(0, \Omega^{-1} M \Omega^{-1}\right)$$

where

$$M = \lim_n \left[\frac{1}{n} \sum_{t=m+1}^{n} M_t \left(\mathrm{E}\Psi_t \Psi_t^\top \right) M_t^\top \right] = \mathrm{E}\left[M_t \Xi M_t^\top \right]$$

where

$$\begin{aligned}
\Xi &= \mathrm{E}\Psi_t\Psi_t^\top \\
&= \mathrm{E}\begin{bmatrix}
\frac{\psi_\tau^2(u_{t\tau})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)^2} & \frac{\psi_\tau(u_{t\tau})\psi_{\tau_1}(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)} & \cdots & \frac{\psi_\tau(u_{t\tau})\psi_{\tau_K}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\
\frac{\psi_\tau(u_{t\tau})\psi_{\tau_1}(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)} & \frac{\psi_{\tau_1}^2(u_{t\tau_1})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)^2} & \cdots & \frac{\psi_{\tau_1}(u_{t\tau_1})\psi_{\tau_K}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\
\cdots & \cdots & \ddots & \cdots \\
\frac{\psi_\tau(u_{t\tau})\psi_{\tau_K}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \frac{\psi_{\tau_1}(u_{t\tau_1})\psi_{\tau_1}(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \cdots & \frac{\psi_{\tau_K}^2(u_{t\tau_K})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)^2}
\end{bmatrix} \\
&= \begin{bmatrix}
\frac{\tau(1-\tau)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)^2} & \frac{\tau\wedge\tau_1-\tau\tau_1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)} & \cdots & \frac{\tau\wedge\tau_K-\tau\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\
\frac{\tau\wedge\tau_1-\tau\tau_1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)} & \frac{\tau_1(1-\tau_1)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)^2} & \cdots & \frac{\tau_1\wedge\tau_K-\tau_1\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} \\
\cdots & \cdots & \ddots & \cdots \\
\frac{\tau\wedge\tau_K-\tau\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \frac{\tau_1\wedge\tau_K-\tau_1\tau_K}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_1)\right)f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)} & \cdots & \frac{\tau_K(1-\tau_K)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau_K)\right)^2}
\end{bmatrix},
\end{aligned}$$

then

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) \\
&= -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1} \times \\
&\quad \left\{ \frac{1}{\sqrt{n}} \sum_t z_t \psi_\tau(u_{t\tau}) + f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right) \Gamma L_{m/K} \left[G^\top A_n G \right]^{-1} G^\top A_n \left[\frac{1}{\sqrt{n}} \sum_{t=m+1}^{n} \Upsilon_{Kt} \otimes \left[D_n^{-1} x_t \right] \right] \right\} \\
&\quad + o_p(1) \\
&= -\Omega^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_t z_t \frac{\psi_\tau(u_{t\tau})}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)} + \left[\Gamma L_{m/K} \left[G^\top A_n G \right]^{-1} G^\top A_n \left[\frac{1}{\sqrt{n}} \sum_{t=m+1}^{n} \Upsilon_{Kt} \otimes \left[D_n^{-1} x_t \right] \right] \right] \right\} \\
&\quad + o_p(1) \\
&= -\Omega^{-1} \frac{1}{\sqrt{n}} \sum_t M_t \Psi_t + o_p(1),
\end{aligned}$$

thus

$$\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) \Rightarrow N\left(0, \Omega^{-1} M \Omega^{-1}\right).$$

A.6. **Proof of Corollary 6.** By Theorem 4,

$$\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) = -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}\Omega^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_t z_t \psi_\tau(u_{t\tau}) + \Gamma_2(\theta, a^*) \sqrt{n}(\widetilde{a} - a^*) \right\}.$$

Setting $\widetilde{a}_0 = 1$, we obtain $\widetilde{Q}_{\varepsilon_t}(\tau) = \widetilde{\alpha}_0(\tau)$ and solving the equations $\widetilde{\alpha}_j(\tau) = \widetilde{a}_j \widetilde{Q}_{\varepsilon_t}(\tau)$, for $j = 1, 2, \cdots, m$, gives

$$\widetilde{a}_0 = 1, \widetilde{a}_1 = \frac{\widetilde{\alpha}_1(\tau)}{\widetilde{\alpha}_0(\tau)}, \cdots, \widetilde{a}_m = \frac{\widetilde{\alpha}_m(\tau)}{\widetilde{\alpha}_0(\tau)}$$

then

$$\begin{aligned}
\widetilde{a}_j &= \frac{\widetilde{\alpha}_j(\tau)}{\widetilde{\alpha}_0(\tau)} \\
&= \{\alpha_j(\tau) + [\widetilde{\alpha}_j(\tau) - \alpha_j(\tau)]\} \left\{ \frac{1}{\alpha_0(\tau)} - \frac{\widetilde{\alpha}_0(\tau) - \alpha_0(\tau)}{\alpha_0(\tau)^2} + \frac{(\widetilde{\alpha}_0(\tau) - \alpha_0(\tau))^2}{\alpha_0(\tau)^2 \widetilde{\alpha}_0(\tau)} \right\} \\
&= \frac{\alpha_j(\tau)}{\alpha_0(\tau)} + \frac{[\widetilde{\alpha}_j(\tau) - \alpha_j(\tau)]}{\alpha_0(\tau)} - \frac{\alpha_j(\tau)\left[\widetilde{\alpha}_0(\tau) - \alpha_0(\tau)\right]}{\alpha_0(\tau)^2} + O_p\left(\frac{m}{n}\right) \\
&= a_j + \frac{[\widetilde{\alpha}_j(\tau) - \alpha_j(\tau)]}{\alpha_0(\tau)} - \frac{\alpha_j(\tau)\left[\widetilde{\alpha}_0(\tau) - \alpha_0(\tau)\right]}{\alpha_0(\tau)^2} + O_p\left(\frac{m}{n}\right)
\end{aligned}$$

Thus, noting that $\widetilde{a}_0 = 1$, we have,

$$\begin{aligned}
& \Gamma_2(\theta(\tau), a^*)(\widetilde{a} - a^*) \\
=& -f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right) \sum_{j=1}^m \mathrm{E}\left[\frac{1}{\sigma_t} \sum_{k=1}^p \theta_k z_t \frac{\partial \sigma_{t-k}(a)}{\partial a_j}\right] (\widetilde{a}_j - a_j) \\
=& -f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right) \sum_{j=1}^m \mathrm{E}\left[\frac{z_t}{\sigma_t} \frac{1}{\alpha_0(\tau)} \left(\sum_{k=1}^p \theta_k \frac{\partial \sigma_{t-k}(a)}{\partial a_j}\right)\right] [\widetilde{\alpha}_j(\tau) - \alpha_j(\tau)] \\
& + f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right) \sum_{j=1}^m \mathrm{E}\left[\frac{z_t}{\sigma_t} \frac{\alpha_j(\tau)}{\alpha_0(\tau)^2} \left(\sum_{k=1}^p \theta_k \frac{\partial \sigma_{t-k}(a)}{\partial a_j}\right)\right] [\widetilde{\alpha}_0(\tau) - \alpha_0(\tau)] + O_p\left(\frac{m^2}{n}\right) \\
=& \frac{1}{\sqrt{n}} f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right) R^\top \left[\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)} D_n^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=m+1}^n x_t \psi_\tau(u_{t\tau})\right)\right] + o_p(\frac{1}{\sqrt{n}}) \\
=& R^\top D_n^{-1} \left(\frac{1}{n} \sum_t x_t \psi_\tau(u_{t\tau})\right) + o_p(1),
\end{aligned}$$

where $R^\top$ is a $(p+q+1) \times (m+1)$ matrix defined as

$$R^\top = \left(-\sum_{j=1}^m \frac{\alpha_j(\tau) r_j}{\alpha_0(\tau)^2}, \frac{r_1}{\alpha_0(\tau)}, \cdots, \frac{r_m}{\alpha_0(\tau)}\right),$$

and

$$r_j = \sum_{k=1}^{p} \theta_k \mathrm{E}\left(\frac{\partial \sigma_{t-k}(a)}{\partial a_j} \frac{z_t}{\sigma_t} \right) = \sum_{k=1}^{p} \theta_k \mathrm{E}\left(\begin{bmatrix} |u_{t-k-j}| / \sigma_t \\ \sigma_{t-1} |u_{t-k-j}| / \sigma_t \\ \cdots \\ \sigma_{t-p} |u_{t-k-j}| / \sigma_t \\ |u_{t-1}| |u_{t-k-j}| / \sigma_t \\ \cdots \\ |u_{t-q}| |u_{t-k-j}| / \sigma_t \end{bmatrix} \right).$$

Consequently, the two-step estimator of $\theta(\tau)^\top = (\beta_0(\tau), \beta_1(\tau), \cdots, \beta_p(\tau), \gamma_1(\tau), \cdots, \gamma_q(\tau))$ has the Bahadur representation:

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\theta}(\tau) - \theta(\tau)\right) \\
&= \Gamma_1(\theta(\tau), \alpha_0)^{-1} \left\{ \sqrt{n} G_n(\theta(\tau), a^*) + \Gamma_{20} \sqrt{n}(\widetilde{a} - a^*) \right\} + o_p(1) \\
&= -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)} \Omega^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_t z_t \psi_\tau(u_{t\tau}) + \Gamma_{20} \sqrt{n}(\widetilde{a} - a^*) \right\} + o_p(1) \\
&= -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)} \Omega^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_t z_t \psi_\tau(u_{t\tau}) + R^\top D_n^{-1} \left(\frac{1}{\sqrt{n}} \sum_t x_t \psi_\tau(u_{t\tau}) \right) \right\} + o_p(1) \\
&= -\frac{1}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)} \Omega^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_t \left[z_t + R^\top D_n^{-1} x_t \right] \psi_\tau(u_{t\tau}) \right\} + o_p(1) \\
&\Rightarrow N\left(0, \frac{\tau(1-\tau)}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)^2} \Omega^{-1} \underline{M} \Omega^{-1} \right)
\end{aligned}$$

where

$$\begin{aligned}
\underline{M} &= \lim \frac{1}{n} \sum_t \left[z_t + R^\top D_n^{-1} x_t \right] \left[z_t^\top + x_t^\top D_n^{-1} R \right] \\
&= \lim \frac{1}{n} \sum_t \left[z_t z_t^\top + R^\top D_n^{-1} x_t z_t^\top + z_t x_t^\top D_n^{-1} R + R^\top D_n^{-1} x_t x_t^\top D_n^{-1} R \right] \\
&= \underline{M}_1 + \underline{M}_2 + \underline{M}_3
\end{aligned}$$

and $\underline{M}_1 = \mathrm{E}\left[z_t z_t^\top\right]$, and

$$\underline{M}_2 = \lim \frac{1}{n} \sum_t \left[R^\top D_n^{-1} x_t z_t^\top + z_t x_t^\top D_n^{-1} R \right], \underline{M}_3 = \lim \frac{1}{n} \sum_t R^\top D_n^{-1} x_t x_t^\top D_n^{-1} R.$$

■

References

Aitchison, J., and J. Brown (1957): *The Lognormal Distribution*. Cambridge U. Press.

Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999): "Coherent measures of risk," *Mathematical Finance*, 9, 203–228.

Bai, Z., and Y. Wu (1994): "Limiting behavior of M-estimators of regression coefficients in high dimensional linear models, I. Scale-dependent case," *Journal of Multivariate Analysis*, 51, 211–239.

Bassett, G., R. Koenker, and G. Kordas (2004): "Pessimistic Portfolio Allocation and Choquet Expected Utility," *Journal of Financial Econometrics*, 4, 477–492.

Bercu, B., and A. Touati (2008): "Exponential inequalities for self-normalized martingales with applications," *Annals of Applied Probability*, forthcoming.

Berkowitz, J., P. Christoffersen, and D. Pelletier (2009): "Evaluating Value-at-Risk models with desk-level data," *Management Science*, forthcoming.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Boos, D. (1984): "Using Extreme Value Theory to Estimate Large Percentiles," *Technometrics*, 26, 33–39.

Boudoukh, J., M. Richardson, and R. F. Whitelaw (1998): "The best of both worlds," *Risk*, 11, 64–67.

Campbell, S. (2005): "A review of Backtesting Procedures," Working paper, Federal Reserve Board, DC.

Chen, X. (2008): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, Amsterdam. Elsevier.

Coroneo, L., and D. Veredas (2008): "Intradaily seasonality of return distribution: A quantile regression approach and intradaily VaR," preprint.

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

Giot, P., and S. Laurent (2003): "Value-at-risk for long and short trading positions," *Journal of Applied Econometrics*, 18, 641–663.

Gourieroux, C., and J. Jasiak (2008): "Dynamic Quantile Models," *Journal of Econometrics*, 147, 198–205.

He, X., and Q. Shao (2000): "On Parameters of Increasing Dimensions," *Journal of Multivariate Analysis*, 73, 120–135.

Huber, P. (1973): "Robust Regression: Asymptotics, conjectures and Monte Carlo," *Annals of Statistics*, 1, 799–821.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–49.

Koenker, R., and Z. Xiao (2006): "Quantile Autoregression," *Journal of American Statistical Association*, 101, 980–1006.

Koenker, R., and Q. Zhao (1996): "Conditional quantile estimation and inference for ARCH models," *Econometric Theory*, 12, 793–813.

Kuan, C.-M., J.-H. Yeh, and Y.-C. Hsu (2009): "Assessing value at risk with CARE: Conditional Autoregressive Expectile models," *Journal of Econometrics*, forthcoming.

Kupiec, P. (1995): "Techniques for verifying the accuracy of risk management models," *Journal of Derivatives*, 2, 73–84.

Mammen, E. (1989): "Asymptotics with Increasing Dimension for Robust Regression with Applications to the Bootstrap," *The Annals of Statistics*, 17, 382–400.

Manganelli, S. (2002): "Codes for the paper CAViaR: Conditional Autoregressive Value at Risk by Regression Quantile," European Central Bank.

McNeil, A. (1998): "Calculating quantile risk measures for financial time series using extreme value theory," ETH E-Collection: `http://e-collection.ethbib.ethz.ch/cgi-bin/show.pl?type=bericht&nr=85`.

Neftci, S. (2000): "Value at risk calculations, extreme events, and tail estimation," *Journal of Derivatives*, 7, 23–37.

Nelson, D. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Pan, J., and D. Duffie (1997): "An Overview of Value at Risk," *Journal of Derivatives*, pp. 7–49.

Portnoy, S. (1985): "Asymptotic behavior of M estimators of p regression parameters when p/n is large," *Annal of Statistics*, 13, 1403–1417.

RiskMetrics Group, A. (1996): *RiskMetrics: Technical Document.* J.P. Morgan and Reuters, New York.

Rossi, D., and A. Harvey (2009): "Quantiles, Expectile and Splines," *Journal of Econometrics*, forthcoming.

Taylor, J. W. (2008a): "Estimating Value at Risk and Expected Shortfall Using Expectiles," *Journal of Financial Econometrics*, 6, 231–252.

——— (2008b): "Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall," *Journal of Financial Econometrics*, 6, 382–406.
Taylor, S. (1986): *Modelling Financial Time Series*. Wiley, New York.
Wang, T. (2000): "A class of dynamic risk measures," Working Paper, University of British Columbia.
Welsh, A. (1989): "On M-processes and M-estimation," *Annals of Statistics*, 17, 337–361.
Wu, G., and Z. Xiao (2002): "An Analysis of Risk Measures," *Journal of Risk*, 4, 53–75.

Department of Economics, Boston College, Chestnut Hill, MA 02467; xiaoz@bc.edu.

Department of Economics, University of Illinois, Champaign, Il, 61820; rkoenker@uiuc.edu.