Econometric Theory, 12, 1996, 793–813. Printed in the United States of America.

CONDITIONAL QUANTILE ESTIMATION AND INFERENCE FOR ARCH MODELS

ROGER KOENKER
University of Illinois

QUANSHUI ZHAO
The City University of Hong Kong

Quantile regression methods are suggested for a class of ARCH models. Because conditional quantiles are readily interpretable in semiparametric ARCH models and are inherently easier to estimate robustly than population moments, they offer some advantages over more familiar methods based on Gaussian likelihoods. Related inference methods, including the construction of prediction intervals, are also briefly discussed.

1. INTRODUCTION

The autoregressive conditional heteroskedasticity (ARCH) model and its many variants have been widely studied and applied throughout econometrics since their introduction by Engle (1982). Gaussian innovations and Gaussian maximum likelihood still predominate this literature despite considerable evidence contrary to the Gaussian hypothesis. Work by Bollerslev (1987), Nelson (1991), and others has sought to address this discrepancy by employing likelihoods based on the Student's *t*-distribution and other alternatives. It is clear that misspecification of the form of the conditional distribution used to define the likelihood can create serious problems for parameter estimation and conditional prediction intervals. This motivation has led us to investigate methods that are not so sensitive to the normality assumption. Recent theoretical work on adaptive estimation of ARCH-type models by Linton (1994) and Drost, Klaassen, and Werker (1994) provides an alternative approach to these problems. See Granger, White, and Kamstra (1989) for an alternative approach to interval forecasting based on ARCH models. Recent reviews of the large ARCH literature include those by Bollerslev, Chou, and Kroner (1992), Bera and Higgins (1993), and Bollerslev, Engle, and Nelson (1994).

We express our appreciation to Joel Horowitz and the referees who handled the paper for their extremely helpful comments. This research was partially supported by NSF grant SBR 93-20555. Address correspondence to: Roger Koenker, Department of Economics, University of Illinois, Champaign, IL 61820, USA.

 0266-4666/96 $9.00 + .10

Because quantiles are readily interpretable in semiparametric location-scale models and are inherently easier to estimate robustly than population moments, we have tried to adapt the quantile regression ideas of Koenker and Bassett (1978) to the ARCH setting. Quantile regression ideas extend naturally to many time-series models. Koul and Saleh (1992) treated quantile regression methods for autoregressive processes, whereas Koul and Mukherjee (1994) studied quantile regression in long-memory models. Portnoy (1991) considered the large sample theory of quantile regression under very general dependence conditions. Here, employing some earlier results from Koenker and Zhao (1994) on heteroskedastic linear models, we consider quantile regression methods for a particular class of ARCH models.

Rather than modeling conditional variance, as is typically done in the ARCH literature, we will focus on models for conditional scale. ARCH models for conditional scale (standard deviation) have been proposed for the Gaussian context by Taylor (1986), Schwert (1989), and Nelson and Foster (1994), among others. Once outside the normal model, scale provides a more natural dispersion concept than variance, as noted by Bickel and Lehmann (1976), and offers substantial advantages from the robustness viewpoint (see, e.g., Bickel, 1978; Carroll and Ruppert, 1988; Newey and Powell, 1987).

Consider a stochastic process $\{y_t\}$ generated by the autoregressive process,

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + e_t, \tag{1.1}$$

whose error term satisfies

$$e_t = (\gamma_0 + \gamma_1 |e_{t-1}| + \cdots + \gamma_q |e_{t-q}|) u_t \tag{1.2}$$

with $\gamma_0 > 0$, $(\gamma_1, \ldots, \gamma_q)' \in R_+^q$, and $\{u_t\}$ are independent and identically distributed (i.i.d.) random variables with mean zero and finite variance. This is an ARCH model of the general type introduced by Engle (1982). However, the conditional variance structure of this model is rather different from Engle's original model, where the error term ϵ_t satisfies

$$\epsilon_t = (\gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \cdots + \gamma_q \epsilon_{t-q}^2)^{1/2} v_t \tag{1.3}$$

with v_t i.i.d. $N(0,1)$. Nevertheless, equations (1.2) and (1.3) have close connections. For sequences of local alternatives to the hypothesis of conditional homoskedasticity, the two models coalesce.

The conditional quantile functions of y_t in the model of (1.1) and (1.2) given information, $\mathcal{F}_{t-1}$, to time $t-1$ are immediately seen to be

$$Q_t(\tau | \mathcal{F}_{t-1}) = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \left(\gamma_0 + \sum_{j=1}^{q} \gamma_j |e_{t-j}| \right) F^{-1}(\tau),$$

where F denotes the common distribution function of the innovations $\{u_t\}$. Quantile regression offers a direct approach to the estimation of the parameters of these conditional quantile functions as well as a related inference apparatus.

In the next section of the paper, we will begin by considering a purely heteroskedastic model without any location-shift AR component. A simple linear quantile regression estimator will be shown to have a natural (Bahadur) asymptotic linear representation under rather mild regularity conditions. Weighted quantile regression estimators of the type introduced in Koenker and Zhao (1994) are then shown to improve upon the efficiency of the unweighted initial estimator. Section 4 is devoted to two-step estimators of the full model of (1.1) and (1.2) in which an initial $\sqrt{n}$-consistent estimator of the α's enables us to apply the methods of the previous sections *to residuals* to estimate the remaining ARCH parameters.

Section 5 briefly discusses some extensions to nonlinear quantile regression estimation of related ARCH models. Inference is taken up in Section 6, where we also discuss methods of constructing prediction intervals. Proofs of all propositions in the text are relegated to the Appendix.

2. THE PURELY CONDITIONALLY HETEROSKEDASTIC LINEAR MODEL

In this section, we will suppress the location shift represented in the model of (1.1) and (1.2) by the autoregressive component and consider the model

$$y_t = \sigma_t u_t = (\gamma_0 + \gamma_1 |y_{t-1}| + \cdots + \gamma_q |y_{t-q}|) u_t, \tag{2.1}$$

where $0 < \gamma_0 < \infty$, $\gamma_1, \ldots, \gamma_q \geq 0$, and $\{u_t\}$ are i.i.d. random variables with distribution function F and density function f.

Let $y_{-(q-1)}, \ldots, y_{-1}, y_0, y_1, y_2, \ldots, y_n$ be observations from model (2.1), and denote

$$Z_t = (1, |y_{t-1}|, \ldots, |y_{t-q}|)', \qquad t = 1, \ldots, n.$$

As in Koenker and Bassett (1978), we define the τth regression quantile estimator,

$$\hat{\gamma}(\tau) = \operatorname{argmin}_\gamma \sum_1^n \rho_\tau (y_t - Z_t'\gamma) \tag{2.2}$$

for $0 < \tau < 1$, and $\rho_\tau(u) = u(\tau - I(u < 0))$. We shall see that $\hat{\gamma}(\tau)$ estimates $\gamma F^{-1}(\tau)$, which we will denote by $\gamma(\tau)$. Computation of the regression quantiles by standard linear programming techniques is extremely efficient. See Koenker and d'Orey (1987) for a detailed discussion of algorithmic aspects. It may be noted that the problem of constraining the estimated ARCH parameters to be nonnegative is relatively straightforward in the linear programming formulation of the quantile regression problem; however, we will defer discussion of this important practical issue. For a discussion of this computational issue in the Gaussian case, see Geweke (1989).

The following lemma establishes sufficient conditions for the stationarity and ergodicity of the process $\{y_t\}$.

LEMMA 2.1. *In model* (2.1), *suppose that* $\mu_r = (E|u_t|^r)^{1/r} < \infty$ *for some* $1 \leq r < \infty$ *and the polynomial*

$$\phi(z) = z^q - \mu_r(\gamma_1 z^{q-1} + \cdots + \gamma_{q-1} z + \gamma_q) \tag{2.3}$$

has all its roots inside the unit circle. Then, the process $\{y_t\}$ *is strictly stationary and ergodic and* $E|y_t|^r < \infty$.

The asymptotic behavior of the quantile regression estimator, $\hat{\gamma}(\tau)$ is described in the following result.

THEOREM 2.1. *In model* (2.1), *suppose that the density, f, is bounded and continuous, that* $f(F^{-1}(\tau)) > 0$ *for any* $0 < \tau < 1$, *and that there exists a* $\delta > 0$ *such that* $E|y_t|^{2+\delta} < \infty$. *Then,* $\hat{\gamma}(\tau)$ *has the following Bahadur representation*:

$$\sqrt{n}(\hat{\gamma}(\tau) - \gamma(\tau)) = \frac{D_1^{-1}}{f(F^{-1}(\tau))} n^{-1/2} \sum_1^n Z_t \psi_\tau(u_t - F^{-1}(\tau)) + o_p(1).$$

Furthermore,

$$\sqrt{n}(\hat{\gamma}(\tau) - \gamma(\tau)) \to^{\mathcal{D}} N\left(0, \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))} D_1^{-1} D_0 D_1^{-1}\right),$$

where $\psi_\tau(u) = \tau - I(u < 0)$ *and* $D_r = EZ_1 Z_1'/\sigma_1^r$.

It may appear odd to those accustomed to maximum likelihood estimation that $\hat{\gamma}(\tau)$ converges not to γ but to $\gamma(\tau) = \gamma F^{-1}(\tau)$; however, the latter is the natural parameter describing the conditional quantile functions for this model. Indeed, the parameter γ is not even identified until we impose some normalization on the vector γ or the scale of F. A natural resolution of this ambiguity would be to set $\gamma_0 = 1$, in which case the vector γ may be estimated as

$$\tilde{\gamma} = (\hat{\gamma}(1-\tau) - \hat{\gamma}(\tau))/(\hat{\gamma}_0(1-\tau) - \hat{\gamma}_0(\tau)).$$

A remarkable feature of Theorem 1 is the mildness of the moment conditions imposed on the conditional density. However, the conditions are natural in view of the boundedness of the function $\psi_\tau(\cdot)$. The preceding lemma may be used to reexpress the moment condition in terms of the innovation moments. We do not view this relaxation of the moment conditions for asymptotic normality as an esoteric refinement of the theory applying only in the realm of a pathological "world without moments." Rather, it is indicative of the performance of the estimator and related inference procedures, relative to the Gaussian MLE, whenever the innovation density has longer-than-Gaussian tails. One may conjecture that even the $(2+\delta)$-moment condition used here could be weakened were we willing to norm $\hat{\gamma}(\tau) - \gamma(\tau)$ by a data-dependent factor rather than $\sqrt{n}$ (see, e.g., Pollard, 1991).

3. WEIGHTED QUANTILE REGRESSION

For linear models with heteroskedastic errors, weighted least-squares estimators with estimated weights can often be shown to inherit the same asymptotic behavior as estimators based on the unknown, optimal weights (see, e.g., Carroll and Ruppert, 1988). Similar results are obtained by Koenker and Zhao (1994) for quantile regression estimates of the linear heteroskedastic model when the design matrix is nonrandom. The Huber (1967) form of the covariance matrix in Theorem 2.1 involving the expression $D_1^{-1}D_0D_1^{-1}$ suggests that similar gains are likely to be available in the present context.

To explore this, we will consider the weighted quantile regression estimator,

$$\tilde{\gamma}(\tau) = \operatorname{argmin}_{\gamma} \sum_{1}^{n} \hat{\sigma}_t^{-1}\rho_\tau(y_t - Z_t'\gamma), \tag{3.1}$$

where $\hat{\sigma}_t$ denotes an appropriate estimator of the conditional scale parameter σ_t. It is apparent that we need only estimate σ_t "up to scale" and we will require that

$$\hat{\sigma}_t = \xi\sigma_t + n^{-1/2}Z_t'\delta_n \tag{3.2}$$

for some scalar $\xi > 0$ and sequence $\delta_n = \mathcal{O}_p(1)$. Only relative scale matters in (3.1), so the presence of ξ in (3.2) causes no harm. Quantile regression offers a wide variety of simple methods for constructing $\hat{\sigma}_t$ to satisfy (3.2). The simplest expedient is to choose for some $\tau \in (\frac{1}{2}, 1)$

$$\hat{\sigma}_t = Z_t'\hat{\gamma}(\tau),$$

which by Theorem 2.1 can be immediately seen to satisfy (3.2). Only slightly more complicated would be to use an interquantile range estimator of the form

$$\hat{\sigma}_t = Z_t'(\hat{\gamma}(1-\tau) - \hat{\gamma}(\tau)).$$

In these cases, the factor ξ in (3.2) takes the values $F^{-1}(\tau)$ and $F^{-1}(1-\tau) - F^{-1}(\tau)$, respectively. More sophisticated L-estimators of scale are considered in Koenker and Zhao (1994).

Corresponding to Theorem 2.1, we have the following result for the weighted quantile regression estimator.

THEOREM 3.1. *Under the conditions of Theorem* 1, *if* $\hat{\sigma}_t$ *satisfies condition* (3.2), *then* $\tilde{\gamma}(\tau)$ *has the Bahadur representation*

$$\sqrt{n}(\tilde{\gamma}(\tau) - \gamma(\tau)) = \frac{D_2^{-1}}{f(F^{-1}(\tau))}\, n^{-1/2}\sum_{1}^{n}\frac{Z_t}{\sigma_t}\,\psi_\tau(u_t - F^{-1}(\tau)) + o_p(1);$$

therefore,

$$\sqrt{n}(\tilde{\gamma}(\tau) - \gamma(\tau)) \xrightarrow{\mathcal{D}} N\left(0, \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\, D_2^{-1}\right).$$

Comparing this result with Theorem 2.1, it is easy to see that the weighted estimator is more efficient than its unweighted counterpart. For any two positive definite matrices A and B, we say $A \leq B$, if $B - A$ is nonnegative definite. In this case, $D_2^{-1} \leq D_1^{-1} D_0 D_1^{-1}$. To see this, let

$$D = \begin{pmatrix} D_2 & D_1 \\ D_1 & D_0 \end{pmatrix} = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \begin{pmatrix} \sigma_t^{-2} & \sigma_t^{-1} \\ \sigma_t^{-1} & 1 \end{pmatrix} \otimes [Z_i Z_i'] \quad \text{a.s.},$$

where $\otimes$ denotes the Kronecker product. D is a nonnegative definite matrix because the matrices

$$\begin{pmatrix} \sigma_t^{-2} & \sigma_t^{-1} \\ \sigma_t^{-1} & 1 \end{pmatrix} \quad \text{and} \quad Z_t Z_t'$$

are nonnegative definite; so is the Kronecker product. If D_2 is positive definite, then there exists orthogonal matrix P, such that

$$P'DP = \begin{pmatrix} D_2 & 0 \\ 0 & D_0 - D_1 D_2^{-1} D_1 \end{pmatrix},$$

so $D_0 - D_1 D_2^{-1} D_1$ is nonnegative definite. Hence, $D_1^{-1} D_0 D_1^{-1} - D_2^{-1}$ is nonnegative definite if D_1 is nonsingular. This result assures that the weighted RQ estimator is asymptotically more efficient than the unweighted one, with asymptotic equivalence when there is no ARCH effect.

4. ESTIMATES BASED ON RESIDUALS

Modeling heteroskedasticity is often based on residuals with the effect of a location component removed. For example, in Engle's (1982) ARCH paper, the estimation is carried out in two steps: the first step involves estimating the autoregressive parameters by least squares and computing the residuals, and the second involves estimating the ARCH parameters by regressing the squared residuals on the lagged squared residuals. In this section, we study the asymptotic behavior of quantile regression estimators when analogous two-step procedures are employed using residuals.

Consider the model of (1.1) and (1.2), which we will write as

$$y_t = X_t'\alpha + (Z_t'\gamma)u_t \tag{4.1}$$

with $X_t = (1, y_{t-1}, \ldots, y_{t-p})'$ and $Z_t = (1, |e_{t-1}|, \ldots, |e_{t-q}|)'$. Suppose we have an $\sqrt{n}$-consistent estimator, $\hat{\alpha}_n$, of the AR parameter, $\alpha = (\alpha_0, \ldots, \alpha_p)'$, that is,

$$\hat{\alpha}_n = \alpha + n^{-1/2}\delta_n, \tag{4.2}$$

where $\|\delta_n\| = \mathcal{O}_p(1)$. We will see that when the innovation distribution F is symmetric, such estimators are readily available. In asymmetric cases, the situation is somewhat more challenging, and we defer discussion of these cases to Section 5.

Given $\hat{\alpha}_n$, we may write $\hat{e}_t = y_t - X_t'\hat{\alpha}_n$ and we can estimate the ARCH parameter by

$$\check{\gamma}(\tau,\hat{\alpha}_n) = \text{argmin}_\gamma \sum \rho_\tau(\hat{e}_t - \hat{Z}_t'\gamma),$$

where $\hat{Z}_t$ replaces the $|e_{t-i}|$'s in Z_t by their corresponding residuals. As we highlight in the next theorem, the impact of the initial estimator, $\hat{\alpha}_n$, may persist in the asymptotic behavior of $\check{\gamma}(\tau,\hat{\alpha}_n)$, although we will see that it vanishes in the symmetric-innovations model.

As earlier, we will assume that $\gamma_0 > 0$, $\gamma_i \geq 0$, $i = 1,\ldots,q$, and that the $\{u_t\}$ sequence is i.i.d. with distribution function F. The stationarity result of Lemma 2.1 is extended easily to this model.

LEMMA 4.1. *In model* (4.1), *suppose that the roots of the polynomial* $\phi(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p$ *lie outside the unit circle and the conditions of Lemma* 2.1 *hold*; *then*, $\{y_t\}$ *is strictly stationary and ergodic and* $E|y_t|^r < \infty$.

The asymptotic theory of the two-step estimator is given by the following result.

THEOREM 4.1. *Under the conditions of Theorem* 2.1, *and assuming that* $\hat{\alpha}_n$ *satisfies* (4.2),

$$\sqrt{n}(\check{\gamma}(\tau,\hat{\alpha}_n) - \gamma(\tau)) = \frac{D_1^{-1}}{f(F^{-1}(\tau))} n^{-1/2} \sum_1^n Z_t \psi_\tau(u_t - F^{-1}(\tau)) + D_1^{-1}G_1\sqrt{n}(\hat{\alpha}_n - \alpha) + o_p(1)$$

with $G_1 = EZ_1(X_1 - B_1\gamma(\tau))'/\sigma_1$ *and* $B_t = (0,\text{sgn}(e_{t-1})X_{t-1},\ldots,\text{sgn}(e_{t-q})X_{t-q})$. *Furthermore, if f is symmetric about zero, and α_0 is restricted to be zero, then* $G_1 = 0$ *and*

$$\sqrt{n}(\check{\gamma}(\tau,\hat{\alpha}_n) - \gamma(\tau)) \to^{\mathfrak{D}} N\left(0, \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))} D_1^{-1}D_0D_1^{-1}\right).$$

Finally, if f is symmetric about zero and α_0 is unrestricted, then for $0 \leq \tau \leq \frac{1}{2}$

$$\sqrt{n}((\check{\gamma}(1-\tau,\hat{\alpha}_n) - \check{\gamma}(\tau,\hat{\alpha}_n))/2 - \gamma(1-\tau)) \to^{\mathfrak{D}} N\left(0, \frac{\tau(\frac{1}{2}-\tau)}{f^2(F^{-1}(\tau))} D_1^{-1}D_0D_1^{-1}\right).$$

Symmetry is clearly crucial in assuring that the contribution of the initial estimator, $\hat{\alpha}_n$, vanishes in the Bahadur representation of $\check{\gamma}(\tau,\hat{\alpha}_n)$. Symmetry is also critical in assuring that simple $\sqrt{n}$-consistent estimators, $\hat{\alpha}_n$, exist. Under symmetric innovations, it is easy to show that the least-squares estimator of α is $\sqrt{n}$-consistent provided our moment conditions are satisfied and that the l_1-estimator is $\sqrt{n}$-consistent provided the innovation density is continuous and strictly positive at the median.

As in Section 3, we can construct weighted quantile regression estimators that improve upon the unweighted estimator $\check{\gamma}(\tau,\hat{\alpha}_n)$.

THEOREM 4.2. *Under the conditions of Theorem* 4.1 *and assuming that* $\hat{\sigma}_t$ *is* $\sqrt{n}$*-consistent up to scale, that is, satisfies condition* 3.2, *the estimator,*

$$\tilde{\gamma}(\tau,\hat{\alpha}_n) = \operatorname{argmin}_\gamma \sum_1^n \hat{\sigma}_t^{-1}\rho_\tau(\hat{e}_t - \hat{Z}_t'\gamma),$$

has the Bahadur expansion,

$$\sqrt{n}(\tilde{\gamma}(\tau,\hat{\alpha}_n) - \gamma(\tau)) = \frac{D_2^{-1}}{f(F^{-1}(\tau))}\, n^{-1/2}\sum_1^n \sigma_t^{-1}Z_t\psi_\tau(u_t - F^{-1}(\tau))$$
$$+ D_2^{-1}G_2\sqrt{n}(\hat{\alpha}_n - \alpha) + o_p(1),$$

where $G_2 = EZ_1(X_1 - B_1\gamma(\tau))'/\sigma_1^2$. *Furthermore, if f is symmetric about zero and* α_0 *is restricted to be zero, then* $G_2 = 0$ *and*

$$\sqrt{n}(\tilde{\gamma}(\tau,\hat{\alpha}_n) - \gamma(\tau)) \to^{\mathfrak{D}} N\left(0, \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\, D_2^{-1}\right).$$

Finally, if f is symmetric about zero and α_0 *is unrestricted, then for* $\tau \in (0,\frac{1}{2})$

$$\sqrt{n}((\tilde{\gamma}(1-\tau,\hat{\alpha}_n) - \tilde{\gamma}(\tau,\hat{\alpha}_n))/2 - \gamma(1-\tau)) \to^{\mathfrak{D}} N\left(0, \frac{\tau(\frac{1}{2}-\tau)}{f^2(F^{-1}(\tau))}\, D_2^{-1}\right).$$

5. NONLINEAR QUANTILE REGRESSION MODELS

Symmetry of the innovations distribution plays a critical role in the discussion of the previous section. In asymmetric models, we can, in principle, find a τ_0 such that $F^{-1}(\tau_0) = 0$ and thereby construct an $\sqrt{n}$-consistent estimator of α based on $\hat{\alpha}_n(\tau_0)$. Even so, we saw that the preliminary, $\hat{\alpha}_n(\tau_0)$, inflated the asymptotic variance of the second step estimator of the ARCH parameters unless the innovation distribution is symmetric. In this section, we briefly describe some nonlinear quantile regression methods for jointly estimating the parameters (α,γ) and mention some extensions to GARCH-type models. Although the models and methods described here are computationally more demanding than the linear quantile regression methods already described, we believe that they represent a potentially rewarding topic of future research.

To illustrate the nonlinear approach, consider the simple special case of our basic model in (1.1) and (1.2),

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + e_t$$

with $e_t = \sigma_t u_t$, $\sigma_t = \gamma_0 + \gamma_1 |e_{t-1}|$, $|\alpha_1| < 1$, $\gamma_0 > 0$, $0 \le \gamma_1 < 1$, and the innovations $\{u_t\}$ i.i.d. with cumulative distribution function F. The conditional quantiles of y_t may be expressed as

$$Q_t(\tau|\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 y_{t-1} + (\gamma_0 + \gamma_1 |y_{t-1} - \alpha_0 - \alpha_1 y_{t-2}|)F^{-1}(\tau).$$

Consider the nonlinear (in parameters) quantile regression estimator

$$\hat{\theta}_n(\tau) = \text{argmin}_\theta \textstyle\sum_t \rho_\tau(y_t - \xi_t(\theta)),$$

where

$$\xi_t(\theta) = \theta_0 + \theta_1 y_{t-1} + \theta_3 |y_{t-1} - \theta_2 - \theta_1 y_{t-2}|.$$

It is straightforward to show that the expectation of the objective function is uniquely minimized with $\theta_0(\tau) = (\alpha_0 + \gamma_0 F^{-1}(\tau), \alpha_1, \alpha_0, \gamma_1 F^{-1}(\tau))$ and consequently that $\hat{\theta}_n(\tau)$ converges to $\theta_0(\tau)$, almost surely. See Jurečková and Prochazka (1994) for details on the asymptotics of nonlinear quantile regression. Note that no symmetry assumption is required here, nor do we need to provide an $\sqrt{n}$-consistent initial estimator of the α parameters.

The nonlinear quantile regression estimator may be extended to higher order models in an obvious way. More importantly, from a practical vantage point, it is straightforward to impose constraints on the coefficients of higher order models in an effort to provide more parsimonious parameterizations. Indeed, the nonlinear quantile regression approach is easily extended to GARCH-type models that have proven extremely popular in econometric applications.

6. INFERENCE

In this section, we turn our attention to inference about ARCH effects. We will adopt an approach that is closely related to the quantile regression tests for heteroskedasticity suggested in Koenker and Bassett (1982). In future work, we hope to explore the rank-based approach to inference of Gutenbrunner, Jurečková, Koenker, and Portnoy (1994) in the ARCH context.

Consider the hypothesis

$$H_0 : R\gamma = 0,$$

where R denotes an $m \times q + 1$-dimensional matrix whose first column, corresponding to the intercept of the ARCH model, γ_0, consists of zero, because we know that γ_0 must be positive. In the purely heteroskedastic model of Section 2, we may consider the sequence of local alternatives,

$$H_n : R\gamma = \zeta/\sqrt{n}.$$

From Theorem 2.1, we have by contiguity

$$\sqrt{n}R\hat{\gamma}(\tau) \to^{\mathcal{D}} N(\zeta F^{-1}(\tau), \omega^2 R D_1^{-1} D_0 D_1^{-1} R'),$$

where $\omega^2 = \tau(1-\tau)/f^2(F^{-1}(\tau))$. Therefore, we have the following result.

THEOREM 5.1. *Under H_n and the conditions of Theorem* 2.1,

$$T_n = n\omega^{-2}\hat{\gamma}'(\tau)R'(RD_1^{-1}D_0D_1^{-1}R')^{-1}R\hat{\gamma}(\tau) \to^{\mathfrak{D}} \chi_m^2(\delta),$$

where δ denotes the noncentrality parameter $\omega^{-2}(F^{-1}(\tau))^2\zeta'(RD_1^{-1}D_0D_1^{-1}R')^{-1}\zeta$. Under H_n and the conditions of Theorem 3.1,

$$T_n = n\omega^{-2}\tilde{\gamma}'(\tau)R'(RD_2^{-1}R')^{-1}R\tilde{\gamma}(\tau) \to^{\mathfrak{D}} \chi_m^2(\delta),$$

where δ denotes the noncentrality parameter $\omega^{-2}(F^{-1}(\tau))^2\zeta'(RD_2^{-1}R')^{-1}\zeta$.

We will not dwell here on the problem of estimating the nuisance parameter ω^2, which appears in the definition of T_n, but the reader may consult, for example, Koenker and Bassett (1982) or Koenker (1994). As noted, for example, in Bollerslev et al. (1994), there are additional problems when the null hypothesis, H_0, places γ on the boundary of the positive orthant. In this case, it would be useful to explore extensions of the work by Wolak (1987) and others on inference under inequality constraints. In the more general model of Section 4, the same considerations yield a test based on estimating the ARCH parameters in the quantile regression of $\hat{e}_t$ on $\hat{Z}_t$, in the notation used there. Finally, as noted in Koenker and Bassett (1982), heteroskedasticity tests based on several distinct quantiles may be constructed based on the joint asymptotic normality of corresponding estimators $\hat{\gamma}(\tau_1),\dots,\hat{\gamma}(\tau_m)$. Again, we will defer a detailed discussion of this approach to subsequent work.

Another important application of ARCH models is out-of-sample prediction. Quantile regression offers a natural approach to the construction of prediction intervals for ARCH-type models, as noted, for example, by Granger, White, and Kamstra (1989). Of course, if the parameters of the model were known exactly, the conditional quantile function itself could be used. The interval

$$[Q_{n+s}(\alpha/2), Q_{n+s}(1-\alpha/2)]$$

would provide an exact $1-\alpha$ level interval for an s-step-ahead forecast. The methods of Portnoy and Zhou (1994) suggest the modification $[\hat{Q}_{n+s}(\alpha/2 - h_n), \hat{Q}_{n+s}(1-\alpha/2+h_n)]$, where $h_n \to 0$ to account for parameter uncertainty.

An important practical aspect of the forecast interval problem involves computing $\hat{Q}_{n+s}(\cdot)$. This is straightforward in the one-step ahead case but more problematic for $s > 1$. Geweke (1989) discussed a Bayesian approach to this problem via Monte Carlo integration and importance sampling. A similar approach seems reasonable for the quantile regression problem.

Let $\hat{Q}_t(\tau|\mathcal{F}_{t-1})$ denote the conditional quantile function of y_t given the information up to time $t-1$. A draw from the one-step-ahead forecast distribution is given by

$$\hat{y}_{n+1} = \hat{Q}_t(U|\mathcal{F}_{t-1}), \tag{6.1}$$

where U is a uniformly distributed random variable on $[0,1]$. Applying (6.1) recursively we can compute a sample path of forecasts $(\hat{y}_{n+1}, \hat{y}_{n+2}, \ldots, \hat{y}_{n+s})'$, and repeatedly applying this procedure and computing the $\alpha/2 - h_n$ and $1 - \alpha/2 + h_n$ quantiles of the empirical distribution of the forecasts could then be used to construct the final prediction intervals. It may appear that the use of (6.1) is computationally prohibitive because it appears to require a quantile regression estimate for each possible realization of $U \in (0,1)$. However, as we have emphasized in previous work (e.g, Koenker and d'Orey, 1987), the entire function $\hat{Q}(\tau | \Upsilon)$ is easily computed by standard parametric linear programming techniques, yielding a piecewise constant function on a known grid that is then readily evaluated by the forecasting simulation.

REFERENCES

Bera, A.K. & M.L. Higgins (1993) On ARCH models: Properties, estimation, and testing, *Journal of Economic Survey* 7, 305–366.

Bickel, P.J. (1975) One-step Huber estimates in linear models. *Journal of the American Statistical Association* 70, 428–433.

Bickel, P.J. (1978) Using residuals robustly. *Annals of Statistics* 6, 266–291.

Bickel, P.J. & E. Lehmann (1976) Descriptive statistics for nonparametric models. III: Dispersion. *Annals of Statistics* 4, 1139–1158.

Bollerslev, T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542–547.

Bollerslev, T., R. Chou, K. Kroner (1992) ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 50, 5–59.

Bollerslev, T., R.F. Engle, & D. Nelson (1994) ARCH models. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4. Amsterdam: North-Holland.

Brown, B.M. (1971) Martingale central limit theorem. *Annals of Mathematical Statistics* 1, 59–66.

Carroll, R.J. & D. Ruppert (1988) *Transformation and Weighting in Regression.* London: Chapman and Hall.

Drost, F.C., C.A.J. Klaassen, & B.J.M. Werker (1994) Adaptiveness in time series models. In P. Mandl & M. Hušková (eds.), *Asymptotic Statistics: Proceedings of the 5th Prague Symposium.* New York: Springer-Verlag.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1007.

Geweke, J. (1989) Exact predictive densities for linear models with ARCH disturbances. *Journal of Econometrics* 40, 63–86.

Granger, C.W.J., H. White, & M. Kamstra (1989) Interval forecasting: An analysis based upon ARCH-quantile estimators. *Journal of Econometrics* 40, 87–96.

Gutenbrunner, C., J. Jurečková, R. Koenker, & S.L. Portnoy (1994) Tests of linear hypotheses based on regression rank scores. *Journal of Nonparametric Statistics* 2, 307–331.

Hájek, Y. (1969) *Nonparametric Statistics.* Holden-Day.

Huber, P.J. (1967) Behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium* 1, 221–233.

Jurečková, J. (1977) Asymptotic relations of M-estimates and R-estimates in linear regression models. *Annals of Statistics* 5, 464–472.

Jurečková, J. & B. Prochazka (1994) Regression quantiles and trimmed least squares estimators in the nonlinear regression model. *Journal of Nonparametric Statistics* 3, 201–222.

Koenker, R. (1994) Confidence intervals for regression quantiles. In P. Mandl and M. Hušková

(eds.), *Asymptotic Statistics: Proceedings of the 5th Prague Symposium.* New York: Springer-Verlag.
Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–50.
Koenker, R. & G. Bassett (1982) Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43–61.
Koenker, R. & V. d'Orey (1987) Computing regression quantiles. *Applied Statistics* 36, 383–393.
Koenker, R. & Q. Zhao (1994) L-estimation for linear heteroscedastic models. *Journal of Nonparametric Statistics* 3, 223–235.
Koul, H.L. & Mukherjee (1994) Regression quantiles and related processes under long range dependence. *Journal of Multivariate Statistics* 51, 318–337.
Koul, H.L. & E. Saleh (1992) Autoregression Quantiles and Related Rank-Scores Process. Technical report RM-527, Michigan State University.
Linton, O. (1994) Adaptive estimation of ARCH models. *Econometric Theory* 9, 539–569.
Nelson, D. (1991) Conditional heteroscedasticity in asset returns. *Econometrica* 59, 347–370.
Nelson, D. & D. Foster (1994) Asymptotic filtering theory for univariate ARCH models. *Econometrica* 62, 1–42.
Newey, W. & J. Powell (1987) Asymmetric least squares estimation and testing. *Econometrica* 55, 819–847.
Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.
Portnoy, S.L. (1991) Asymptotic behavior of regression quantiles in non-stationary, dependent cases. *Journal of Multivariate Analysis* 38, 100–113.
Portnoy, S.L. & Q. Zhou (1994) A Direct Approach to Construct Confidence Intervals and Confidence Bands for Regression Quantiles. Preprint.
Ruppert, D. & R.J. Carroll (1980) Trimmed least square estimation in the linear model. *Journal of the American Statistical Association* 75, 828–838.
Schwert, G.W. (1989) Why does stock market volatility change over time? *Journal of Finance* 44, 1115–1154.
Stout, W.F. (1974) *Almost Sure Convergence.* New York: Wiley.
Taylor, S. (1986) *Modeling Financial Time Series.* New York: Wiley.
Wolak, F. (1987) An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association* 82, 782–793.

APPENDIX

Proof of Lemma 2.1. Denote $\|x\|_r = (E|x|^r)^{1/r}$. Then,

$$\begin{aligned}\|y_t\|_r &= \mu_r(E(\gamma_0 + \gamma_1|y_{t-1}| + \cdots + \gamma_q|y_{t-q}|)^r)^{1/r}\\ &\le \mu_r(\gamma_0 + \gamma_1\|y_{t-1}\|_r + \cdots + \gamma_q\|y_{t-q}\|_r).\end{aligned}$$

Thus, for $\xi_t = (\|y_{t-1}\|_r, \dots, \|y_{t-q}\|_r)'$,

$$\xi_t \le A\xi_{t-1} + b, \tag{A.1}$$

where

$$A = \begin{bmatrix} \mu_r\gamma_1 & \mu_r\gamma_2 & \cdots & \mu_r\gamma_q \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdot \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} \mu_r\gamma_0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}.$$

The characteristic polynomial of the matrix A is $\phi(z)$ defined by (2.3), so all the eigenvalues of matrix A fall inside the unit circle. Consequently,

$$\xi_t \le (I - A)^{-1} b < \infty,$$

and, in particular, $\|y_t\|_r < \infty$.

To show ergodicity, let $Z_t = (|y_{t-1}|, \ldots, |y_{t-q}|)'$ and $v_t = |u_t|$; then,

$$Z_t = A_{t-1} Z_{t-1} + b_{t-1},$$

where

$$\mathbf{A}_t = \begin{bmatrix} \gamma_1 v_t & \cdots & \gamma_{q-1} v_t & \gamma_q v_t \\ 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \qquad \mathbf{b}_t = \begin{bmatrix} \gamma_0 v_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

It follows that

$$Z_t = \sum_{k=1}^{\infty} \left(\prod_{i=1}^{k-1} \mathbf{A}_{t-i} \right) \mathbf{b}_{t-k},$$

which is the limit of monotone increasing measurable functions of $\{v_t\}$ and, hence, is measurable.

Now, let $\{X_t : t = \ldots, -1, 0, 1, \ldots\}$ be a k-variate i.i.d. process and $h : (R^k)^\infty \to R^k$ be a measurable function; then, $Y_t = h(X_t, X_{t-1}, \ldots)$ is strictly stationary and ergodic (see Stout, 1974, Lemma 3.4.8 and Theorem 3.5.8). It follows that the process $\{(|y_t|, \mathrm{sgn}(u_t))\}$ is strictly stationary and ergodic, as is $\{y_t\} = \{|y_t| \mathrm{sgn}(u_t)\}$. ∎

Before proving Theorem 2.1, we will introduce a sequence of lemmas that will be needed. We will assume throughout what follows that $u_1, u_2, \ldots$ are i.i.d. random variables with df F and density function f.

LEMMA A.1. *Let $g_1, g_2, \ldots$ be random variables such that $E|g_t|^{2+\delta} \le S < \infty$ for some $\delta > 0$. Then,*

$$n^{-1/2} \max_{t \le n} |g_t| = o_p(1).$$

Proof. For any $\epsilon > 0$,

$$P\left(n^{-1/2} \max_{t \le n} |g_t| \ge \epsilon\right) \le \sum_{1}^{n} P(|g_t| \ge \epsilon n^{1/2}) \le nE|g_1|^{2+\delta} / (\epsilon n^{1/2})^{2+\delta} \to 0.$$ ∎

LEMMA A.2. *Suppose* $\{g_t\}$ *satisfies the conditions of Lemma* 2.1 *and* $\{H_t\}$ *is a sequence of random p-vectors such that* $E\|H_t\|^{2+\delta} \le S < \infty$, *with* g_t *and* H_t *independent of* $(u_t, u_{t-1}, \ldots)$ *and*

$$n^{-1}\sum_1^n g_t H_t' \to^P G,$$

a nonrandom, nondegenerate matrix. Then,

$$V(\Delta) = n^{-1/2}\sum_1^n g_t \psi_\tau(u_t - F^{-1}(\tau) - n^{-1/2}H_t'\Delta) \tag{A.2}$$

satisfies

$$\sup_{\|\Delta\| \le M} |V(\Delta) - V(0) + f(F^{-1}(\tau)G\Delta| = o_p(1) \tag{A.3}$$

for fixed M, $0 < M < \infty$.

Proof. The proof is an adaptation of the methods of Ruppert and Carroll (1980) and Bickel (1975). Let $\xi_t(\Delta) = \psi_\tau(u_t - F^{-1}(\tau) - n^{-1/2}H_t'\Delta)$, and define

$$\eta_t(\Delta) = \xi_t(\Delta) - \xi_t(0) - E_{t-1}(\xi_t(\Delta) - \xi_t(0)),$$

where E_{t-1} denotes the conditional expectation with respect to the σ-field generated by $\mathcal{F}_{t-1} = \sigma\{g_t, H_t, u_{t-1}, u_{t-2}, \ldots\}$. We wish to show

$$T(\Delta) = n^{-1/2}\sum_1^n g_t \xi_t(\Delta) = o_p(1). \tag{A.4}$$

Because

$$|\xi_t(\Delta)| \le I(-n^{-1/2}\|H_t\|M < u_t - F^{-1}(\tau) < n^{-1/2}\|H_t\|M) + F(F^{-1}(\tau) + n^{-1/2}\|H_t\|M) - F(F^{-1}(\tau) - n^{-1/2}\|H_t\|M),$$

it follows that

$$E_{t-1}|\xi_t(\Delta)|^2 \le 4\big(F\big(F^{-1}(\tau) + n^{-1/2}\max_{t\le n}\|H_t\|M\big) - F\big(F^{-1}(\tau) - n^{-1/2}\max_{t\le n}\|H_t\|M\big)\big) = \kappa_n,$$

where $0 \le \kappa_n \le 4$, and $\kappa_n \to 0$ in probability, using Lemma A.1 and the continuity of F. Thus,

$$\begin{aligned} P(|T(\Delta)| \ge \epsilon) &\le \frac{1}{n\epsilon^2} E\sum_1^n g_t^2 E_{t-1}|\xi_t(\Delta)|^2 \\ &\le \frac{1}{n\epsilon^2}\sum_1^n E|g_1|^2\kappa_n \le \frac{1}{\epsilon^2}(S_\delta)^{2/(2+\delta)}(E\kappa_n^{(2+\delta)/\delta})^{\delta/(2+\delta)}, \end{aligned} \tag{A.5}$$

which converges to zero, because $E\kappa_n^{(2+\delta)/\delta} \to 0$ by Lebesgue's dominated convergence theorem.

Next, using Bickel's (1975) chaining approach, we will show that

$$\sup_{\|\Delta\| \le M} |T(\Delta)| = o_p(1). \tag{A.6}$$

Decompose $\{\|\Delta\| \le M\}$ into cubes based on the grid $(j_1\theta M, \ldots, j_k\theta M)$, where $j_i = 0, \pm 1, \ldots, \pm[1/\theta] + 1$, and $\theta > 0$ is a fixed (small) number. Denote $P(\Delta)$ the lower vertex of the cube that contains Δ. Then,

$$\sup_{\|\Delta\| \le M} |T(\Delta)| \le \sup_{\|\Delta\| \le M} |T(P(\Delta))| + \sup_{\|\Delta\| \le M} |T(P(\Delta)) - T(\Delta)|. \tag{A.7}$$

The first term on the right-hand side is $o_p(1)$ because it is the maximum of finite number of $o_p(1)$ terms. For the second term, because ξ_t is monotone in Δ,

$$\begin{aligned}|\eta_t(P(\Delta)) - \eta_t(\Delta)| &\le \eta_t(P(\Delta)) - \eta_t(P(\Delta) + 1\theta M)\\ &\quad + 2E_{t-1}(\xi_t(P(\Delta)) - \xi_t(P(\Delta) + 1\theta M))\\ &\le \eta_t(P(\Delta)) - \eta_t(P(\Delta) + 1\theta M) + 2\theta M n^{-1/2}\|f\|\,\|H_t\|.\end{aligned}$$

The second term on the right-hand side of (A.7) can thus be divided into

$$\sup_{\|\Delta\| \le M} |T(P(\Delta) - T(P(\Delta) + 1\theta M))| + 2\theta M\|f\| n^{-1}\sum_1^n |g_t|\,\|H_t\|,$$

where the first term is $o_p(1)$ and the second term is bounded by an arbitrarily small number θ. Thus, (A.6) follows.

Finally,

$$\begin{aligned}&\sup_{\|\Delta\| \le M}\left|n^{-1/2}\sum_1^n g_t E_{t-1}\xi_t(\Delta) + f(F^{-1}(\tau))G\Delta\right|\\ &= \sup_{\|\Delta\| \le M} n^{-1}\left|\sum_1^n g_t H_t'\Delta f(F^{-1}(\tau) - \theta^* n^{-1/2}H_t'\Delta) - f(F^{-1}(\tau))G\Delta\right|\\ &\le n^{-1}\sum_1^n g_t\|H_t\| o_p(1) + M\|f\|\left\|G - n^{-1}\sum_1^n g_t H_t'\right\| = o_p(1).\end{aligned} \tag{A.8}$$

The lemma follows by combining (A.6) and (A.8). ■

LEMMA A.3. *Suppose g_t, H_t, Z_t are random vectors, independent of $u_t, u_{t+1}, \ldots$ and $E|g_t|^{2+\delta}, E\|H_t\|^{2+\delta}, E\|Z_t\|^{2+\delta} \le S_\delta < \infty$ for some $\delta > 0$. Define*

$$U(\Delta) = n^{-1}\sum_1^n \frac{g_t Z_t'}{1 + n^{-1/2}Z_t\delta_n}\psi_\tau(u_t - F^{-1}(\tau) - n^{-1}H_t'\Delta),$$

where $\|\delta_n\| = \mathcal{O}_p(1)$. Then,

$$\sup_{\|\Delta\| \le M} \|U(\Delta)\| = o_p(1) \tag{A.9}$$

for any fixed $M > 0$.

Proof. Because it is known by Lemma A.1 that $n^{-1/2}\max_{t\le n}\|Z_t\| = o_p(1)$, Taylor's expansion yields

$$\frac{1}{1 + n^{-1/2}Z_t\delta_n} = 1 - \frac{n^{-1/2}Z_t\delta_n}{(1 + \theta^* n^{-1/2}Z_t\delta_n)^2},$$

where $0 \le \theta^* \le 1$. Thus,

$$
\begin{aligned}
\sup_{\|\Delta\|\le M} \|U(\Delta)\|
&\le \sup_{\|\Delta\|\le M} \left\| n^{-1} \sum_1^n g_t Z_t' \psi_\tau(u_t - F^{-1}(\tau) - n^{-1} H_t' \Delta) \right\| \\
&\quad + n^{-3/2} \sum_1^n |g_t| \, \|Z_t\|^2 \cdot \mathcal{O}_p(1) \\
&\le \left\| n^{-1} \sum_1^n g_t Z_t' \psi_\tau(u_t - F^{-1}(\tau)) \right\| \\
&\quad + \sup_{\|\Delta\|\le M} \left\| n^{-1} \sum_1^n g_t Z_t' (\psi_\tau(u_t - F^{-1}(\tau) - n^{-1} H_t' \Delta) - \psi_\tau(u_t - F^{-1}(\tau)) \right\| \\
&\quad + n^{-1} \sum_1^n |g_t| \, \|Z_t\| \cdot n^{-1/2} \max_{t\le n} \|Z_t\| \cdot \mathcal{O}_p(1) \\
&\le \left\| n^{-1} \sum_1^n g_t Z_t' \psi_\tau(u_t - F^{-1}(\tau)) \right\| \\
&\quad + n^{-1} \sum_1^n |g_t| \, \|Z_t\| I(|u_t - F^{-1}(\tau)| \le n^{-1/2} \|H_t\| M) + o_p(1). \qquad \textbf{(A.10)}
\end{aligned}
$$

In the preceding inequalities, we have used the results obtained in Lemma A.1 that $n^{-1/2} \max_{t\le n} \|Z_t\| = o_p(1)$ implied by the finite $2 + \delta$th moment.

Now, $\{g_t Z_t' \psi_\tau(u_t - F^{-1}(\tau))\}$ is a martingale difference sequence, which has $1 + \delta/2$ moment. Therefore,

$$
\sum_1^\infty E \| g_t Z_t' \psi_\tau(u_t - F^{-1}(\tau)) \|^{1+\delta/2} / t^{1+\delta/2} < \infty.
$$

So by Theorem 3.3.1 in Stout (1974, p. 137),

$$
n^{-1} \sum_1^n g_t Z_t' \psi_\tau(u_t - F^{-1}(\tau)) \to^P 0,
$$

which implies that the first term on the right-hand side of (A.10) is $o_p(1)$. For the second term, note that

$$
\begin{aligned}
E&\left(n^{-1} \sum_1^n |g_t| \, \|Z_t\| I(|u_t - F^{-1}(\tau)| \le n^{-1/2} \|H_t\| M) \right) \\
&\le n^{-1} \sum_1^n E |g_t| \, \|Z_t\| \eta_n \\
&\le n^{-1} \sum_1^n (E|g_t|^{2+\delta} E\|Z_t\|^{2+\delta})^{1/(2+\delta)} \cdot (E\eta_n^{(2+\delta)/\delta})^{\delta/(2+\delta)} \qquad \textbf{(A.11)}
\end{aligned}
$$

with

$$
\eta_n = F\big(F^{-1}(\tau) + n^{-1} \max_{t\le n} \|H_t\| M\big) - F\big(F^{-1}(\tau) - n^{-1} \max_{t\le n} \|H_t\| M\big),
$$

which is bounded by 1 and converges to 0 in probability. Again by dominated convergence, $E\eta_n^{(2+\delta)/\delta} \to 0$. Thus, the second term on the right-hand side of (A.10) is also $o_p(1)$ and (A.12) follows. ∎

LEMMA A.4. *Let $V_n(\Delta)$ be a vector function that satisfies*

(i) $-\Delta' V_n(\lambda\Delta) \geq -\Delta' V_n(\Delta)$, $\lambda \geq 1$,
(ii) $\sup_{\|\Delta\| \leq M} \| V_n(\Delta) + f(F^{-1}(\tau))D\Delta - A_n \| = o_p(1)$,

where $\|A_n\| = \mathcal{O}_p(1)$, $0 < M < \infty$, $f(F^{-1}(\tau)) > 0$, and D is a positive-definite matrix. Suppose that Δ_n is a vector such that $\|V_n(\Delta_n)\| = o_p(1)$. Then, $\|\Delta_n\| = \mathcal{O}_p(1)$ and

$$\Delta_n = \frac{D^{-1}}{f(F^{-1}(\tau))} A_n + o_p(1). \tag{A.12}$$

Proof. The argument is based on Jurečková (1977) and Koenker and Zhao (1994). For any $\epsilon > 0$, and $\eta > 0$,

$$\begin{aligned}
&P\Big(\inf_{\|\Delta\|=M} [-\Delta' V_n(\Delta)] < \eta M\Big) \\
&\quad \leq P\Big(\inf_{\|\Delta\|=M} [-\Delta' V_n(\Delta)] < \eta M, \inf_{\|\Delta\|=M} [-\Delta'(-f(F^{-1}(\tau))D\Delta + A_n)] \geq 2\eta M\Big) \\
&\qquad + P\Big(\inf_{\|\Delta\|=M} [-\Delta'(-f(F^{-1}(\tau))D\Delta + A_n)] \leq 2\eta M\Big) \\
&\quad \leq P\Big(\sup_{\|\Delta\|=M} \|V_n(\Delta) + f(F^{-1}(\tau))D\Delta - A_n\| \geq \eta\Big) \\
&\qquad + P(\|A_n\| \geq f(F^{-1}(\tau))\lambda_1(D)M - 2\eta),
\end{aligned}$$

where $\lambda_1(D)$ represents the minimum eigenvalue of D. By assumption (ii) and the preceding inequality, it is obvious that $M > 0$, $\eta > 0$, and $N_0 > 0$ can be chosen in such a way that

$$P\Big(\inf_{\|\Delta\|=M} [-\Delta' V_n(\Delta)] < \eta M\Big) < \epsilon, \tag{A.13}$$

for $n \geq N_0$.

Now for any Δ, $\|\Delta\| \geq M$, denote $\Delta = \lambda\Delta_1$, where $\lambda \geq 1$ and $\|\Delta_1\| = M$. From assumption (i), we have

$$\|V_n(\Delta)\| \geq [-\Delta_1 V_n(\lambda\Delta_1)]/M \geq [-\Delta_1 V_n(\Delta_1)]/M.$$

Thus,

$$P\Big(\inf_{\|\Delta\|\geq M} \|V_n(\Delta)\| < \eta\Big) \leq P\Big(\inf_{\|\Delta\|=M} [-\Delta' V_n(\Delta)] < \eta M\Big) < \epsilon. \tag{A.14}$$

Finally, for sufficiently large n,

$$\begin{aligned}
P(\|\Delta_n\| \geq M) &\leq P(\|\Delta_n\| \geq M, \|V_n(\Delta_n)\| < \eta) + P(\|V_n(\Delta_n)\| \geq \eta) \\
&\leq P\Big(\inf_{\|\Delta\|\geq M} \|V_n(\Delta)\| < \eta\Big) + \epsilon \leq 2\epsilon,
\end{aligned}$$

because $\|V_n(\Delta_n)\| = o_p(1)$. Thus, we have established that $\|\Delta_n\| = \mathcal{O}_p(1)$. From this result and assumption (ii), it follows that

$$V_n(\Delta_n) + f(F^{-1}(\tau))D\Delta_n - A_n = o_p(1).$$

Thus, (A.12) follows. ■

LEMMA A.5. *Let $\hat{\beta}$ be the minimizer of the function*

$$\sum_1^n w_t \rho_\tau(y_t - x_t' b),$$

where $w_t > 0$, x_t are $\mathfrak{F}_{t-1}$-measurable, and $y_t | \mathfrak{F}_{t-1}$ has a positive density function. Then,

$$\left\| n^{-1/2} \sum_1^n w_t x_t \psi_\tau(y_t - x_t' \hat{\beta}) \right\| \le \dim(x_1) n^{-1/2} \max_{t \le n} \| w_t x_t \|.$$

Proof. The proof follows from Ruppert and Carroll (1980). ■

Proof of Theorem 2.1. Let

$$V(\Delta) = n^{-1/2} \sum_1^n Z_t \psi_\tau(u_t - F^{-1} - n^{-1/2} \sigma_t^{-1} Z_t' \Delta),$$

and set $\hat{\Delta}_n = \sqrt{n}(\hat{\gamma}(\tau) - \gamma(\tau))$ and $\gamma(\tau) = \gamma F^{-1}(\tau)$. Then,

$$V(\hat{\Delta}_n) = n^{-1/2} \sum_1^n Z_t \psi_\tau(y_t - Z_t' \hat{\gamma}(\tau)).$$

By Lemmas A.1 and A.5,

$$\| V(\hat{\Delta}_n) \| \le (q+1) n^{-1/2} \max_{t \le n} \| Z_t \| = o_p(1). \tag{A.15}$$

Next, by Lemma 2.1, Z_t is strictly stationary and ergodic and has a $(2 + \delta)$ moment. Thus,

$$n^{-1} \sum_1^n \frac{Z_t Z_t'}{\sigma_t} \to E \frac{Z_1 Z_1'}{\sigma_1} = D_1,$$

where D_1 is positive definite. Further by Lemma A.2, we have

$$\sup_{\|\Delta\| \le M} \| V(\Delta) - V(0) + f(F^{-1}(\tau)) D_1 \Delta \| = o_p(1),$$

with $V(0) \to^{\mathfrak{D}} N(0, \tau(1-\tau) D_0)$ by Brown's (1971) martingale central limit theorem.

Finally, because $\psi_\tau(u)$ is an increasing function in u, it is obvious that the function

$$-\Delta' V(\lambda \Delta) = n^{-1/2} \sum_1^n (-Z_t' \Delta) \psi_\tau(u_t - F^{-1}(\tau) - n^{-1/2} \sigma_t^{-1} Z_t' \Delta \lambda)$$

is an increasing function of λ, so the conditions of Lemma A.4 hold. Combining this with (A.15), one gets $\| \hat{\Delta}_n \| = \mathcal{O}_p(1)$ and

$$\hat{\Delta}_n = \frac{D_1^{-1}}{f(F^{-1}(\tau))} V(0) + o_p(1),$$

and the result follows by Slutsky's theorem. ■

Proof of Theorem 3.1. By (3.2), we have

$$\hat{\sigma}_t = \xi(\sigma_t + n^{-1/2} Z_t' \delta_n),$$

where $\|\delta_n\| = \mathcal{O}_p(1)$. Thus,

$$\sum_1^n \frac{1}{\hat{\sigma}_t} \rho_\tau(y_t - Z_t'b) = \frac{1}{\xi} \sum_1^n \frac{\sigma_t^{-1}}{1 + n^{-1/2}\sigma_t^{-1} Z_t' \delta_n} \rho_\tau(y_t - Z_t'b).$$

Denote

$$U(\Delta) = n^{-1/2} \sum_1^n \frac{\sigma_t^{-1} Z_t}{1 + n^{-1/2}\sigma_t^{-1} Z_t' \delta_n} \psi_\tau(u_t - F^{-1}(\tau) - n^{-1/2}\sigma_t^{-1} Z_t' \Delta)$$

and

$$V(\Delta) = n^{-1/2} \sum_1^n \sigma_t^{-1} Z_t \psi_\tau(u_t - F^{-1}(\tau) - n^{-1/2}\sigma_t^{-1} Z_t' \Delta).$$

It follows from Lemma A.3 that

$$\sup_{\|\Delta\| \le M} \|U(\Delta) - V(\Delta)\| = o_p(1).$$

Applying Lemma A.2 and the preceding result, we have

$$\sup_{\|\Delta\| \le M} \|U(\Delta) - V(0) + f(F^{-1}(\tau)) D_2 \Delta\| = o_p(1),$$

where $D_2 = EZ_1 Z_1'/\sigma_1^2$.

Let $\hat{\Delta}_n = \sqrt{n}(\tilde{\gamma}(\tau) - \gamma(\tau))$. Like the proof of Theorem 2.1, one may obtain $\|U(\hat{\Delta})\| = o_p(1)$ from Lemma A.5. Further by Lemma A.4, one gets

$$\hat{\Delta}_n = \frac{D_2^{-1}}{f(F^{-1}(\tau))} V(0) + o_p(1),$$

where $V(0) \to N(0, \tau(1-\tau)D_2)$ again by Brown's martingale central limit theorem. Applying Slutsky's theorem completes the proof. ■

Proof of Theorem 4.1. For the model

$$y_t = X_t'\alpha + (Z_t'\gamma)u_t.$$

If $\hat{\alpha}_n$ is an $\sqrt{n}$-consistent estimator of α, that is, $\hat{\alpha}_n = \alpha + n^{-1/2}\delta_n$, $\|\delta_n\| = \mathcal{O}_p(1)$, then

$$\hat{e}_t = y_t - X_t'\hat{\alpha}_n = e_t - n^{-1/2} X_t' \delta_n$$

and

$$|\hat{e}_t| = |e_t| - n^{-1/2} X_t' \operatorname{sgn}(e_t) \delta_n.$$

If we denote

$$B_t = (0, X_{t-1} \operatorname{sgn}(e_{t-1}), \ldots, X_{t-q} \operatorname{sgn}(e_{t-q})),$$

then

$$\hat{Z}_t' = Z_t' - n^{-1/2} \delta_n' B_t.$$

And denoting $\Delta_1 = \sqrt{n}(b - \gamma(\tau))$, we may write

$$\hat{e}_t - \hat{Z}_t'b = e_t - n^{-1/2}X_t'\delta_n - Z_t'b + n^{-1/2}\delta_n'B_t\gamma(\tau) + o_p(n^{-1/2})$$
$$= \sigma_t(u_t - F^{-1}(\tau) - n^{-1/2}\sigma_t^{-1}[Z_t'\Delta_1 + X_t'\delta_n - \gamma'(\tau)B_t'\delta_n]) + o_p(n^{-1/2}).$$

Let $H_t' = \sigma_t^{-1}(Z_t', X_t' - \gamma'(\tau)B_t')$, and define

$$V(\Delta) = n^{-1/2}\sum_1^n Z_t\psi_\tau(u_t - F^{-1}(\tau) - n^{-1/2}H_t'\Delta).$$

Then, by Lemma A.2, we have

$$\sup_{\|\Delta\|\le M} \|V(\Delta) - V(0) + f(F^{-1}(\tau))G\Delta\| = o_p(1),$$

where $G = \lim n^{-1}\sum_1^n Z_tH_t' = (D_1, G_1)$ and $G_1 = EZ_1(X_1 - B_1\gamma(\tau))'/\sigma_1$.

Because $\|\delta_n\| = \mathcal{O}_p(1)$, with $\Delta = (\Delta_1, \Delta_2)$ replaced by (Δ_1, δ_n), we have

$$\sup_{\|\Delta_1\|\le M} \|V(\Delta_1, \delta_n) - V(0) + f(F^{-1}(\tau))(D_1\Delta_1 + G_1\delta_n)\| = o_p(1).$$

Furthermore, in the definition of $V(\Delta)$, let Z_t be replaced by $\hat{Z}_t$, and hence denote it $\hat{V}(\Delta)$. Lemma A.3 assures that $\sup_{\|\Delta_1\|\le M}\|V(\Delta) - \hat{V}(\Delta)\| = o_p(1)$. Using the same argument as that in the proof of Theorem 2.1, we have $\|\hat{V}(\hat{\Delta}_1, \delta_n)\| = o_p(1)$; hence, by Lemma A.4

$$\hat{\Delta}_1 = \frac{D_1^{-1}}{f(F^{-1}(\tau))}V(0) - D_1^{-1}G_1\delta_n + o_p(1).$$

Finally, suppose f is symmetric and $f(0) > 0$. Then,

$$\mathrm{sgn}(u_1), \ldots, \mathrm{sgn}(u_n), |u_1|, \ldots, |u_n|$$

are mutually independent (cf. Hájek, 1969, Theorem 19A). Because Z_t's are functions of $\{|u_t|\}$, it follows that the series $\{Z_t\}$ and $\{\mathrm{sgn}(u_t)\}$ are mutually independent. Thus,

$$E(Z_1/\sigma_1)(X_t\,\mathrm{sgn}(e_t)\gamma(\tau))' = E\big(E(\mathrm{sgn}(u_t)\,|\,Z_1, \mathcal{F}_{t-1})(Z_1/\sigma_1)(X_t\gamma(\tau))'\big) = 0,$$

because $\mathrm{sgn}(u_t)$ is independent of Z_1 and $\mathcal{F}_{t-1}$ and $E\,\mathrm{sgn}(u_t) = 0$. It follows that $EZ_1(B_1\gamma(\tau))'/\sigma_1 = 0$.

It remains to show $EZ_1X_1'/\sigma_1 = 0$. When $\alpha_0 = 0$, so $X_1 = (y_0, y_{-1}, \ldots, y_{1-p})'$, and we have

$$y_t = \sum_{j=0}^\infty a_j\sigma_{t-j}|u_{t-j}|\mathrm{sgn}(u_{t-j}),$$

where the a_j's are nonrandom and satisfy $\sum_j |a_j| < \infty$. Thus,

$$E(Z_1/\sigma_1)y_t = \sum_{j=0}^\infty a_jEZ_1\sigma_{t-j}|u_{t-j}|\mathrm{sgn}(u_{t-j})/\sigma_1 = 0$$

and therefore $G_1 = 0$. When α_0 is not restricted to zero, the intercept effect of $\hat{\alpha}_n$ does not vanish from the Bahadur representation even under symmetry. However, in this case consideration of the interquantile range estimator of γ shows

$$(\check{\gamma}(1 - \tau, \hat{\alpha}_n) - \check{\gamma}(\tau, \hat{\alpha}_n))/2 \to (\gamma(1 - \tau) - \gamma(\tau))/2 = \gamma(1 - \tau)$$

and that the contribution of the preliminary estimator cancels, while the scalar factor in the asymptotic covariance matrix is replaced by

$$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \begin{pmatrix} \tau(1-\tau) & \tau^2 \\ \tau^2 & \tau(1-\tau) \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \Big/ f^2(F^{-1}(\tau)) = \frac{\tau(\frac{1}{2}-\tau)}{f^2(F^{-1}(\tau))}. \quad \blacksquare$$

Proof of Theorem 4.2. The proof follows by combining the arguments for Theorems 3.1 and 4.1. ■

Proof of Theorem 6.1. The proof is immediate from Theorems 2.1 and 3.1. ■